

Modeling of Polish Intonation for Statistical-Parametric Speech Synthesis

Tomasz Kuczmarski

May 6, 2021

Adam Mickiewicz University



Faculty of Modern Languages and Literature
Institute of Ethnolinguistics

Modeling of Polish Intonation for Statistical-Parametric Speech Synthesis

Tomasz Kuczmariski

Supervisor Prof. zw. Dr hab. Inż. Grażyna Demenko

Tomasz Kuczmarski

Modeling of Polish Intonation for Statistical-Parametric Speech Synthesis

May 6, 2021

Supervisor: Prof. zw. Dr hab. Inż. Grażyna Demenko

Adam Mickiewicz University

Institute of Ethnolinguistics

Faculty of Modern Languages and Literature

al. Niepodległości 4

61-874 Poznań

The work presented in this thesis was partially carried out within the scope of the research grant Harmonia no. UMO-2014/14/M/HS2/00631 received from the National Science Centre.

Badania przedstawione w pracy zostały częściowo zrealizowane w ramach grantu badawczego Harmonia nr UMO-2014/14/M/HS2/00631 przyznanego przez Narodowe Centrum Nauki.

Dla Mamy

Contents

1	Introduction	1
1.1	Motivation	3
1.2	Objective	5
1.3	Methodology	6
1.4	Contribution	8
1.5	Outline	9
2	Background	13
2.1	Speech Synthesis	15
2.2	Deep Neural Models of Speech	21
2.2.1	Convolutional Neural Networks	23
2.2.2	Neurobiological Foundations	25
2.2.3	Deep Temporal Convolutional Neural Model of Speech	31
2.2.4	Deep Neural Network Explainability	35
2.3	Intonation Modeling	41
2.3.1	European School	42
2.3.2	American School	44
2.3.3	Research in Polish Intonation	46
3	Methodology	51
3.1	Aim of the Current Work	51
3.2	Motivation	51
3.2.1	Unificationist Approach to Intonation Modeling	51
3.2.2	Interfacing Phonetics and Phonology	54
3.2.3	Explainable Deep Neural Network-Based Model of Intonation	55
3.3	Deep Temporal Convolutional Neural Network as a Scientific Model .	58
3.4	Methods	61
3.4.1	Dataset	61
3.4.2	Feature Extraction	65
3.4.3	Model Implementation	73
3.4.4	Model Training	75

3.4.5	Inference	77
3.4.6	Feature Relevance Analysis	78
3.4.7	Neural Source-Filter Resynthesis	78
3.4.8	Perceptual Evaluation	81
4	Results	87
4.1	Objective Evaluation	93
4.2	Subjective Evaluation	93
4.3	Feature Relevance	99
5	Discussion and Conclusions	117
6	Future Plans and Challenges	133
	Bibliography	139
A	Listing of Features	169
B	Listing of Feature Groups	207

Introduction

” *Language disguises thought.*

— **Ludwig Wittgenstein**

Tractatus Logico-Philosophicus

This dissertation should begin with a definition of its research subject. However, the term intonation is easily understood intuitively but becomes rather problematic when one tries to formulate a proper scientific definition. The problem arises from the very nature of intonation itself. Elusive and indefinable, a steady flow of change in perceived pitch, neither does it fit entirely into the standard symbolic and categorical framework of linguistics, nor into the experimental and physicalist phonetics. At the same time, intonation determines the meanings of utterances and carries a whole range of other information critical to successful speech communication. The meaning conveyed can be both categorical and gradual. It does not seem to be tied to any specific segment of the carrier utterance, hence the popular name – suprasegmentals. The fact that a single intonation contour might express a whole range of meanings on all possible levels of language and beyond, makes it that much more challenging to study. Syntactic and discourse information are mixed with cues to the internal states of the speaker, such as their emotion, health or even sex and social background. All this information is passed through this single channel, encoded in the perceived height of the fundamental frequency of speech. Despite the complexity, humans extract all that information effortlessly. Human brain and cochlea have evolved for millions of years to do just that, interpret the internal mental states of other members of the population, even from very brief exchanges of vocalizations. For the early humans (as for most other vertebrates) this ability to momentarily assess sounds produced by others often meant life or death (or which way to run). When humans started developing language communication, intonation also started adjusting and taking on new roles and passing on new information through conventional categorical signals. On the other hand, it also never really ceased to fulfil some of its primary functions; like to convey the emotional state of the speaker, for example, through the (much less intentional) amount of fundamental frequency variation. Even the most influential proponents of excluding intonation from the language altogether, deeming it part of the performance (or *parole*),

eventually admit that at least some of its parts might be of interest to linguistics (therefore, moving it closer to the actual *langue* in the saussurean sense) [48]. Some other lines of thought, like cognitive linguistics, move intonation to a more central position by assigning meanings to various intonation signs and treating those as metaphors¹. These conceptual metaphors themselves are also often argued to be embodied in the humans themselves and resulting from the human composure and the environment in which they are immersed in [189].

Moreover, intonation is usually considered from just a single perspective within the traditional dualistic view of language, which by itself, is underlying many linguistic paradigms, not excluding some of the most contemporary ones. Intonation is either studied as an acoustic phenomenon by phonetics or as a symbolic grammar by phonology. This dichotomous approach and the epistemological gap it created has become one of the main obstacles in building comprehensive scientific theories and models of the phenomenon. The missing interface between the physicalist measurements of phonetics and the discrete grammars of cognitive categories formulated within phonology becomes very evident in the context of speech technology where all theory of language is put to a pragmatic functional test. The only models that have proven to be useful for application in real-world speech processing systems are the probabilistic models delivered by the rapidly developing field of Artificial Intelligence (or more specifically Machine Learning). This comes in hand with what some of the contemporary linguists suggest; that the mapping from mental constructs, whatever they are in terms of intonation, to the physical realization is of probabilistic nature and that it is embodied in the human brain itself.

The current work leans towards a unificationist and a non-reductive physicalist (emergentist) approach. This is in line with some of the “new developments in physics, biology, psychology, and crossdisciplinary fields such as cognitive science, artificial life, and the study of non-linear dynamical systems [that] have focused strongly on the high level *collective behaviour* of complex systems, which is often said to be truly emergent [...]” [89]. In this work, the author also assumes that intonation is crucial to language communication and although some parts of it might exist as part of the paralanguage it is merely the lack of proper methodologies in linguistics (and communicology) that render intonation problematic and in some cases even result in banishing it from the field altogether. Intonation being an aspect of language spanning all of its possible levels from phonetics, through phonology, syntax, semantics and discourse, should be studied accordingly, within frameworks

¹A downward movement is often a metaphor for the definite and an upward movement for the unknown.

that provide a unified view, focusing not only on the physical (phonetic) or psychological (phonological, grammatical) aspects of the problem in question, but also (or mainly) on the mappings between the realization and the mental construct. This work also assumes that these mappings are of probabilistic nature and that they are embodied in the human brain where this subjective experience emerges from a combination of numerous processes.

1.1 Motivation

A number of intonation models that fit the above description already exist in the field of speech technology, which extensively employs probabilistic methods in the form of various Machine Learning (ML) models. These models, especially recently, have proven to yield very good results both for the analysis and synthesis of speech and all of its components. This fact provides some positive evidence for the soundness and effectiveness of the probabilistic approach. Some of the best models to date are based on Deep Neural Networks (DNNs). Interestingly, a number of these state-of-the-art deep neural architectures are at least loosely founded on the design principles of the human brain. Examples include the Long Short-Term Memory (LSTM) networks and attention-based models that, as their names suggest, use some simplistic approximations of the human short-term memory and attention. Deep Convolutional Neural Networks (CNNs), where probably the first to be founded on the principles of information processing in the human brain [105]. CNNs use building blocks that approximate the actual computations that specific layers of cells in the visual cortex conduct when processing visual data. These architectures have outperformed humans in a recent image classification challenge, setting a new state-of-the-art. The first notable application of CNNs in the domain of speech modeling (aside from Automatic Speech Recognition) was the Wavenet synthesizer [239]. It demonstrated a surprising naturalness of synthesized speech and was even able to spontaneously reproduce a range of non-speech sounds like filled pauses, labial clicks and breathing sounds. Although the choice of Wavenet's architecture, the Deep Temporal Convolutional Neural Network, was based on purely practical premises (the solution was adopted from the visual domain [312]), some recent neurobiological studies show that the general nature of processing information in all of human's sensory cortices might be based on the same principles, which might in turn provide some neurobiological support for the unintentional architecture of that model.

The main problem with such pragmatic models lies in their evident lack of explanatory and exploratory power. Most of these models still act as black-boxes and provide only mere reflections of the quantitative nature of the modeled phenomena. In this way, they do not present much value to science².

However, this situation has recently started changing. The huge implementational potential of DNNs in a number of different domains caused a strong demand for methods that are able to explain the outputs of these model. A model trained to recognize pictures of polar bears might be cheating by using the dominating color of the background, as polar bears usually appear on snowy pictures. Similarly, a model designed to detect subclinical breast cancer from screening mammography could learn to assign positive diagnosis due to some bias in the training data, and a deep learning model trained on historical recruitment data of some huge global corporation in which 70% of the employees had been white men, could, and if not constrained surely would, consider the sex and race a significant feature for future predictions of candidate fit. Such cases provided the necessary stimulation for the development of AI explainability algorithms. The first step in that direction was made with the visualization of the internal filters of image classifying CNNs and the outputs of the model's intermediate layers. The domain of visual data was ideal for early experimentation as the output was easily interpretable by a naked eye. The first results were very surprising as they revealed the ability of the CNNs to extract latent features from images in a similar manner a human brain would be expected to. From simple detection of edges, through more complex shapes and patterns, towards meaningful parts of the classified objects, like bird feathers and dogs ears, noses, human faces and even whole complex situations. The huge potential of these methods stimulated further work in that direction and the development of many new explainability algorithms. These algorithms immediately started being adopted for scientific experimentation, especially in computational neuroscience and medicine. Examples in domains other than that of visual data are still quite sparse. The field of language and speech still remains relatively unexplored, with only a number of studies in machine translation [265] and a probably only a single study regarding speech recognition [29].

AI explainability methods could provide a foundation for a methodology to be applied for studying these (if not all) concepts of language that resist being framed within highly categorical, symbolic grammars and other similar systems. This opportunity seems especially appealing in case of intonation which up to this day remains relatively unstudied through systematic experimental studies. Depending

²Especially in its traditional positivist understanding.

on the choice of the information that comprise the model input, whether these are phonological categories or some semantic features, these methods might present a completely new look on how certain information are relevant and add up to give rise to the final output of systems based on the same principles that underlie the human brain³.

This study, except for its undeniable relevance for explaining intonation, would benefit the overall speech synthesis technology as it could potentially help determine what kind of information is required for effective prediction of intonation contours for any given utterance, which are optional and which are completely redundant. This would allow better design of datasets used for training speech synthesis models and could help significantly reduce the time, complexity and effort of the training itself as less features might be used instead of the traditionally sparse feature vector that contains all of the possible information that could be extracted or engineered. Despite the rapid progress being made in technology, especially in the recent decades, the problem of the human-machine communication through speech is far from being solved, with intonation processing being one of the main problem areas, both in synthesis and analysis.

The undeniable potential of the deep convolutional neural architectures, their neurobiological foundations, the unprecedented look into the internal workings of these networks that the explainability algorithms offer in connection with the timeliness and significance of the problem they could help solve when combined comprise the main motivation behind the current work. Also to the best of the author's knowledge, no similar attempt to apply these methods in a scientific study of speech production has been made so far.

1.2 Objective

The current work, therefore, aims to:

- Build a robust biologically-inspired neural model of the probabilistic mapping between discrete low-level linguistic features of an utterance and its intonation contours (F_0 values).
- Build a state-of-the-art neural source-filter resynthesis framework for Polish read speech.

³Or rather some very idealistic generalizations of very small parts of the human brain, and that is if we are at least remotely close to understanding how the brain actually works.

- Deploy the intonation model within the resynthesis framework and measure the perceived naturalness of the output intonation contours, and to
- operationalize the results of these measurements as an indicator of the model's robustness.
- Develop a method to explain the relevance of the individual input linguistic features for the produced intonation contour.
- Analyze how specific linguistic features contribute to the F_0 contours of an utterance.

The objectives and steps outlined above are designed to test the following hypotheses:

Hypothesis 1. *The continuous F_0 contours of an utterance emerge from its discrete linguistic features through a series of successive probabilistic mappings into intermediate latent representations.*

Hypothesis 2. *The biologically-inspired Deep Temporal Convolutional Network can be an effective model of these mappings and hence of Polish neutral read speech intonation in the context of statistical-parametric speech synthesis.*

Hypothesis 3. *The set of shallow linguistic features used in this thesis provides information which is sufficient for synthesis of natural sounding intonation in the context of statistical-parametric speech synthesis.*

Hypothesis 4 (contributory methodological). *A Deep Temporal Convolutional Network can become an explanatory scientific model of mappings between linguistics features and the intonation of an utterance.*

1.3 Methodology

In order to test the above hypotheses, this work uses a mixture of quantitative inductive methods. A neural model of intonation is built and evaluated through a series of psychoacoustic experiments, whose results are necessarily operationalized as an indicator of the performance and adequacy of the constructed model. Then a method for revealing the relations between the input categories and the output F_0 values based on the constructed model and with the use of explainability algorithms

is proposed. This method is later applied to a set of test samples isolated from the original speech dataset. In this way, extensive results are produced reflecting the positive and negative evidence for the predicted F_0 values provided by individual linguistic features of an utterance and its parts. The results are aggregated within various feature groups to provide a number of different perspectives of the gathered data.

For this purpose, this work implements a Deep Temporal Convolutional Neural Network as a model of intonation in neutral Polish read speech. The limitations imposed on the domain, although very idealistic, were introduced to necessarily constrain the number of processes that might influence the realization of intonation contours, and as a consequence also help reduce the necessary complexity of the model to an achievable level. A speech corpus originally built for the purpose of building a Polish BOSS unit selection synthesizer [69], and designed with special emphasis on suprasegmental coverage of the Polish language was used as the training, test and validation data set for the current study.

As the first step, an inventory of 1297 various quantitative, qualitative and positional features were extracted. The corpus consisting of a total of 1908 variable length utterances was split in an 8:1:1 ratio into training, validation and test sets. Using a Python/Keras programming stack the model architecture along with a training and evaluation framework were implemented and deployed onto a special computational cloud infrastructure where a number of experiments were run.

The resulting model was evaluated both objectively through calculation of various Mean Squared Error metrics, as well as in an extensive perceptual evaluation study conducted with a specially designed web application and with the use of a neural source-filter speech resynthesizer, that was also specially trained for this purpose.

The model was finally applied to infer the F_0 values for all samples in the holdout test set. Additionally, for each of the predictions, feature relevance analysis was performed with help of a specially adjusted implementation of the Layer-wise Relevance Propagation algorithm. The individual results were aggregated using a number of calculation methods both for individual features and using a number of high-level feature groupings as an attempt to capture more general trends in the data. Feature relevance rankings were calculated both for individual features and for feature groups using various abstraction levels.

1.4 Contribution

The current work contributes to the study of intonation in a number of ways. First and foremost, the traditional dualistic approach is criticized and a unificationist approach based on a physicalist (emergentist) view of the phenomenon is proposed in Section 3.2.1 instead. The epistemological gap between phonetics and phonology is addressed by placing the main focus, not on the acoustic measurements of the fundamental frequency, or the development of formal grammars of intonation but on the mappings between these two levels.

As a result, a fully functional pragmatic intonation model ready to be applied in a statistical-parametric speech synthesis system is built and evaluated in the context of a state-of-the-art Neural Source-Filter synthesizer [319], which was also implemented for the purpose of this study and is the first implementation of this synthesis method for the Polish language. The intonation model was developed along with a special method for generating prediction explanations in a form of input feature relevance *w.r.t.* model's output [17, 264, 234]. This explanation method, to the best of author's knowledge, has never been applied in a scientific study of intonation or language production, especially as part of a scientific explanatory model.

The source code for model training, inference and evaluation and for running the explainability algorithms is made openly available in a public repository at https://github.com/mrslacklines/intonation_synthesis along with the necessary documentation for replicating the results. The repository includes tools that allow building and running it on any platform, including the Amazon Web Services computational cloud, with a single command, without the necessity to manually install any additional libraries.

The repository also contains a full set of results in the form of data plots and comma-separated data files. These include relevance-analysis results for individual files and test set aggregate rankings, both for individual features, as well as for a number of different abstraction levels and feature groupings. The results also include all of the resynthesized and natural speech samples used in the perceptual evaluation of the intonation model. The results can be obtained at the code repository in the results folder⁴.

In the process of developing the model a modest contribution was also made to the code of the currently most developed Python library for Deep Neural Network

⁴https://github.com/mrslacklines/intonation_synthesis/tree/master/intonation_synthesis/results

explainability – Innvestigate [6]. A *pull request* was staged to the official repository with a number of changes that allow to use the library with the most recent versions of the industry standard tensor and Deep Learning libraries – Tensorflow [1] and Keras [47, 123]. The contributed code changes can be viewed in the official repository at <https://github.com/albermax/innvestigate/pull/229>.

A fully functional working application was also developed for the purpose of conducting perceptual evaluation of the model. The source code is published in another public repository at https://github.com/mrslacklines/listening_experiments. With slight modification the code can be easily reused to perform professional ABX discrimination and mean opinion score (MOS) experiments. The evaluation experiment application is deployed onto a professional high-availability web infrastructure and can still be accessed and taken part in at <http://fonetyka.cudaniewidy.org/experiment>.

Last but not the least, this work includes an exhausting argumentation for the adoption of Deep Neural Networks, and more specifically the Deep Temporal Convolutional Neural Network in connection with the Layer-wise Relevance Propagation explainability method as scientific models. This argumentation, further supported with the results of the current study, provide an important methodological contribution to the proposed approach to the study of intonation.

1.5 Outline

The current chapter, provides only a very short introduction to the general problem of intonation. It touches upon the potential incompatibility of the traditional toolbox of linguistics and the actual nature of intonation as its controversial subject. Section 1.1 looks at this problem as a source of inspiration for choosing Convolutional Neural Networks and Artificial Intelligence Explainability as the modeling framework for the current work. The actual scientific objectives are listed and translated into four main research hypotheses, which this work will attempt to test, in Section 1.2. The methodology selected for this purpose is also briefly outlined in Section (1.3). The last Section of this chapter (Section 1.4) lists the main contributions of this work.

The next chapter (Chapter 2), aims at providing a more detailed view of the general problem of linguistic inquiry into intonation modeling. It starts with a brief summary of the history from the perspective of the many different methods used to approach this specific problem, and tries to emphasize the impracticability of the dualistic

tradition. In Section 2.1, the concept of speech synthesis is introduced. It gives another short historical summary aiming to demonstrate the importance of speech synthesis in building scientific models of speech production. The main milestones and different technologies are listed as a path that led to the current state-of-the-art, recently set by the Deep Convolutional Neural Network-based models. Next, an extensive summary of this modeling method is provided in Section 2.2. Starting with the original implementation of this idea in the visual domain, this chapter gives an in-depth description of the neurobiological inspiration and design principles of these networks in Section 2.2.2. It then lists results of recent neurobiological studies suggesting that similar biological analogies may be helpful in the domain of speech. In Section 2.2.4, the high potential of these networks is further explored through an introduction to AI Explainability, and especially to the current leading-edge method – the Layer-wise Relevance Propagation algorithm. The next part of this chapter (Section 2.3), finally takes a closer look at intonation as a research problem. Because many exhausting reviews of this topic can be found, this work gives only a short overview of the different schools (in Section 2.3.1 and 2.3.2) of thought and how they formed in the relatively short history of this scientific domain. A separate Section 2.3.3 was devoted to a summary of research considering the intonation of Polish, which is the language of interest of this work.

Chapter 3 introduces the reasoning behind the current methodology. The current framework is supported from a few different viewpoints in Sections 3.2.1, 3.2.2 and 3.2.3. The controversy around adopting a neural network as a linguistic model is addressed from many perspectives, based on the various notions of a scientific model itself in Section 3.3. Additionally, Layer-wise Relevance Propagation explainability algorithm is proposed there as means for addressing the apparent lack of explanatory and exploratory power of such models. Section 3.4 of this chapter, provides a detailed description of the specific methods used in this work. It starts with the specification of the initial dataset in Section 3.4.1. Then, in Section 3.4.2, the specific set of linguistic features is described along with methods used for extracting them from the original data. Next, the technicalities of model implementation, training and testing are reported. Finally, the methods used for perceptual evaluation of the resulting model are presented. These include the adoption of Neural Source-Filter vocoder (in Section 3.4.7) as a resynthesizer for generating stimuli for the perceptual experiment, whose design is described in Section 3.4.8.

The results are presented in Chapter 4. The chapter starts with an overview of the general quality of F_0 predictions as compared to the ground truth values. The results of objective (Section 4.1) and subjective (Section 4.2) evaluation are covered

next. Section 4.3 provides a comprehensive look at the results of feature relevance analysis, both for individual predictions, as well as from a number of more general perspectives and abstraction levels.

Chapter 5 provides a discussion of the advantages and disadvantages of the current approach. It identifies the necessary compromises and the effects they had on the final results, but also the accuracy of the method in general. It also includes an attempt to analyze and interpret the results.

The main body of the dissertation ends with Chapter 6, where a number of ideas for future work, given the high potential of the current method, is presented.

Two appendices are additionally included at the end of this dissertation. Appendix A contains a listing and description of all linguistic features used for intonation modeling in this work. These are included in the form of a raw machine code along with the regular expression-like matching masks that were used as one of the steps to extract them. Appendix B, in turn, contains raw Python code and description of feature groups that were used for the analysis of feature relevance for the F_0 contour output by the model. These are important additions to the current work and the reader might want to refer to that data at various points. Because of their length, however, they were moved to the back matter in order to avoid distraction from the main argument in the text.

Background

“ *If I have seen further it is by standing on the shoulders of Giants.*

— Isaac Newton

1675

At the 6th International Congress of Phonetic Sciences (ICPhS) in Prague in 1967, Dennis Fry in his keynote speech [102, 260] argued that it is the study of prosody features in speech that should become one of the main present-day tasks of phonetic sciences. His exact words were:

“We do not have enough direct measurements and observations in the area of prosodic features and a good proportion of the data we have are not particularly well organised. To take as an example tone and intonation and their relation to fundamental frequency, we need a more systematic approach to observations in this area and in particular a much sharper awareness of the different functions of affective and grammatical intonation” [102, 260].

This appeal must have been indeed heard by the scientific community as the number of prosodic research started growing exponentially from that moment. Language pedagogy was probably the first domain of language study that openly recognized the essential role of prosody in speech communication. This first wave of prosody research was purely descriptivist and Soviet researchers were undeniably leading it. They were later joined by a number of researchers from other parts of the world. First those interested in the tonal languages, that by their very own nature require a deeper inquiry into the tonal patterns and their relation to meaning [43, 3, 130, 42]. Numerous descriptions of English intonation were soon to follow [242, 36, 35, 157, 268, 172, 15, 93, 61, 327, 127]. This, in turn, opened the door for research into other European languages [98, 97, 259].

At the same time, a number of linguists were making first attempts at describing intonation from a more theoretic viewpoint, in relation to pragmatics [165, 216, 120, 64, 326] or syntax [30]. In the early 40s of the 20th century, when in Europe

linguistic study of prosody was mostly seen as superfluous and insignificant, American structuralists built the first fully-fledged linguistic (phonological) models of intonation [323, 33, 246, 282]

With the development of electric devices that provided tools for sophisticated measurements and experimentation, prosody research started gaining momentum and recognition as a proper science. As Rossi [260] points out, it was mainly speech synthesis that drove prosody research. It served both as a model of, as Fry [102] described it, “how the features which appear in the data are used by the people who employ the particular language”, and as an immediate goal – a system that can automatically produce human like speech¹. Fry’s programmatic contribution was further supported by Jones [162]. This approach started dominating Europe as it was corresponding with the, then popular, russelian empiricism and the realism of William James [273]. It did not assume the superiority of experimental phonetics, but rather saw phonological models as theories that needed validation through experimentation. Work by Denes [79], Uldall [309], Hadding-Koch [126], Delattre et al. [68], Lieberman and Michaels [206], Isacenko and Schädlich [151], Mettas [222], Cohen, Hart, et al. [54] and Öhman [238] constitute the theoretical background of present prosody research where models are validated mainly with the use of speech technology and listening experiments. This new instrumentalist approach, however, was still widely seen as non-linguistic, especially among the traditional grammarians. Although there were many notable voices for the inclusion of phonetics as part of linguistics [101], the dualistic gap became cemented in the domain of speech research for years to come and to this day it is a significant burden carried even by some of the contemporary research. As described by Cutler and Ladd [63], intonation research at that time was either concerned with measuring and describing the concrete physical acoustic shapes or with building theoretical models and grammars at some abstract level of representation. The origin of these epistemological differences lies in the definition of the research subject itself. Intonation was either seen as a system of some abstract cognitive units [323, 246, 245], or as their concrete physical realization through the contours of fundamental frequency.

However, as the fast growing domain of speech technology has proven, linking one to the other is at the heart of the problem of understanding intonation. An implementation of a system that can translate an abstract phonological or cognitive description of intonation into an acoustic signal (or the other way around) allows for experimental evaluation of the linguistic theory behind the model. This is, in

¹Speech synthesis is further introduced in Section 2.1.

fact, compliant with how D. Fry [101] thought a truly scientific approach to the study of intonation should look like:

- i “A science must deal with data that obstruct thought: these data appear in the form of empirical facts.”
- ii “A science does not speak of observables in terms of objects, but rather in terms of relations existing between those objects: science seeks to uncover a structure, a Reality, beneath empirical facts, using a model whose results must feed the theory, which need revision in order to send back new questions to the model.”

He notes further that if we want “to discover the regularities of symbolic representation and to identify the structuring system underlying pre-systematic phenomena”, intonational research should focus on studying “a noise that manifests mental representations” and accepting that the “independence [of these two levels] is the source of variability and the lack of a one-to-one correlation” [260]. The scope of a comprehensive theory of language remains enormous, ranging from philosophy, through cognitive science to information theory, mathematics and experimental physics [296]. Efforts at integrating some of these fields within some single shared metatheory result in building stronger and more adequate models of language and intonation. Even Chomsky [48], as cited by Tatham [296], saw that a model characterising language and speech production would eventually consist of descriptions of how thought could be mapped to sound. Such mappings at various levels of language, including the prosodic level, constitute the core problem for contemporary linguistics as well as speech and language technology.

2.1 Speech Synthesis

The deep human need for understanding the mystery of speech and testing that understanding through implementation of the principles into various kinds of inanimate calculatory systems has a history far longer than that of linguistics, speech technology and artificial intelligence altogether. The first mentions of such devices start with the mythical brazen heads². Many historic and mythical figures like Gerbert of Aurillac, Boethius, Faust, Arnaldus de Villa Nova, Enrique de Villena, Virgil, Rob Grosseteste, Saint Albertus Magnus and even the Norse God Odin were believed

²A legendary automaton – a male head made of brass or bronze, variously mechanical or magical. It was said to be able to correctly answer any questions, although sometimes restricted to simple “yes” or “no” answers

to had been in possession of a such a mechanical talking head [201]. The first well-documented scientific examples of modeling speech production were, however, much simpler and date back to only 250 years ago.

The flute-like resonator models of the human vocal tract constructed by Kratzenstein [152, 95, 267, 201] could produce sounds imitating the five basic vowels (/a/, /e/, /i/, /o/, and /u/) and provided some initial insight into their articulatory origin. A more complete and, more importantly, fully dynamic model was built soon after by Wolfgang von Kempelen in 1791. His bellows-based machine could produce combinations of sounds and the author used it in a number of articulatory experiments. As reported by Lemmetty [201] in his review of the history of speech synthesis:

“[...] the essential parts of the machine were a pressure chamber for the lungs, a vibrating reed to act as vocal cords, and a leather tube for the vocal tract action. By manipulating the shape of the leather tube he could produce different vowel sounds. Consonants were simulated by four separate constricted passages and controlled by the fingers. For plosive sounds he also employed a model of a vocal tract that included a hinged tongue and movable lips. His studies led to the theory that the vocal tract, a cavity between the vocal cords and the lips, is the main site of acoustic articulation. Before von Kempelen’s demonstrations the larynx was generally considered as a center of speech production.”

This is a great example of how a simple mechanistic model can help support a theory of the much more complicated process of speech production or even help shift the dominating paradigm. The success of Kemepelen’s machine triggered the development of a number of improved versions, including the works of Charles Wheatstone and Alexander Graham Bell [152, 267].

The developments in electrical devices at the beginning of the 20th century brought completely new possibilities to the domain of speech synthesis. Mechanical source modulation and resonance control were first replaced by electrical circuits in 1922 by Stewart [176] and Wagner [152] and the idea was improved upon by Obata and Teshima [267]. Although these devices could only synthesize simple vowels, they demonstrated the importance of the first three formants in vowel production and perception. The idea of an electrical circuit-based model of speech production, as implemented in the notable VODER, designates a turning point for speech synthesis research. It demonstrated that it is possible to produce fully intelligible speech with completely artificial means. The device had to be played like a musical instrument by a specially trained operator who simultaneously controlled the type of source

excitation, fundamental frequency and the configurations of ten bandpass filters comprising the spectral filter. Apart from stimulating further research into speech synthesis, the source-filter model became a standard, which is used to this day even in the most advanced parametric systems [304, 166, 333, 239]. The idea of parametric synthesis was further explored by Lawrence [192], Fant and Martony [92], Carlson, R. [41], Holmes W. [144] and Barber S. [22], amongst others. In the process, many important findings were made, especially regarding the acoustic phonetic features of the signal and its relation to the naturalness and intelligibility of the produced speech. At the same time, assessing the validity of the implemented models of speech production operationalized through synthesizer's performance and measured with perceptual testing became a standard [142, 143, 177]. The development of speech parameterization techniques also resulted in the development of a number of important algorithms, like Linear Predictive Coding (LPC) [267] which became a core method in a number of domains including modern telecommunication.

Electric speech synthesizers based on a completely different idea were also developed around the same time [258, 176, 201]. They implemented a model of speech articulation through sets of recorded control signals. The work by Teranishi and Umeda [300] is one of the successful implementations of this technique and also the first ever full end-to-end Text-To-Speech (TTS) system. It included a sophisticated text analysis module implementing complicated heuristics [201, 177]. That work set an important milestone, as speech synthesis was now ready for commercial applications and this single fact became the main driving factor behind the development of the many systems to follow. The quality of speech started improving fast from this point. The development of systems such as the MITalk and its successor Klattalk [176, 7] started a fast paced arms race which gave birth to such industry standards as Festival [297, 31] or BOSS [175]. Concatenative systems that rely on reorganization of previously recorded speech samples, despite their apparent simplicity, also helped validate a number of notions. For example diphone synthesis demonstrated the need for explicit modeling of intersegmental transitions rather than static states only and unit selection algorithms demonstrate the influence of context on the realization of a speech segment.

With numerous practical applications and as a new and very attractive research topic, TTS has revealed a significant problem in the field of speech research - the impracticability of the dualistic gap between the phonological and phonetic views of the intonation. Intonation in itself was one of the biggest challenges for speech synthesis research and the current phonological models were found to translate rather poorly to the naturalness of fundamental frequency contours, especially in

implementations that were based on sets of discrete rules or those that employed databases of speech recordings as inventories of acoustic exemplars for constructing new unseen utterances. This was the moment that speech synthesis research faced a challenge of bridging the phonetic-phonological gap in order to move forward with improving the quality and naturalness of synthetic speech, which at that time was still rather poor. In traditional TTS synthesis it was a matter of finding an adequate phonological representation of input intonation and devising a method for mapping these to the output contours of fundamental frequency. It soon became clear that such a straightforward approach is far from ideal. As Hirschberg [136] notes:

“Research on prosody, as on many linguistic phenomena which rely upon context for their interpretation, is more a matter of finding likelihoods — not simple mappings from syntax or semantics or even from an underlying meaning representation to a clear set of prosodic features.”

Although the formal tools for implementing models based on such probabilistic mappings were readily available at that time, it was only with the recent advent of high performance computing that they had a chance to become a real alternative for the simple rule- or data-based models [201].

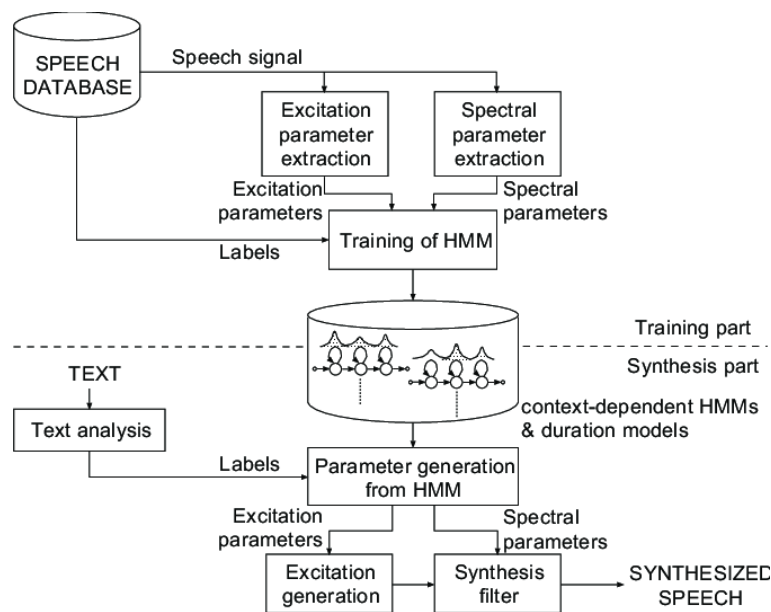


Fig. 2.1.: An overview of the basic HMM-based speech synthesis system. Adopted from the HTS Slides [121] (released under the Creative Commons Attribution 3.0 license)

The last couple of decades became dominated by various probabilistic and statistical models including Hidden Markov Models [198, 252, 336, 183] and their numerous derivatives such as the HSMM, GMM [336, 167] and more recently the Deep Neural

Networks [333, 334]. Based on purely statistical methods they turned out to be much more effective at the task of modeling the complex phenomenon of speech. They have proven to produce much more natural and smooth intonation as compared with systems built based on the more traditional data-based methods. The Hidden Markov Model-Based Speech Synthesis System (HTS) set a new standard for the domain for years to come [328]. A Hidden Markov Model (HMM) is a collection of states connected by transitions with two sets of probabilities in each: a transition probability which provides the probability for taking this transition, and an output probability density function (PDF) which defines the conditional probability of emitting each output symbol from a finite alphabet, given that the transition is taken [198, 201]. Thanks to its statistical but also deeply parametric nature, and hence a small footprint, this synthesis method allowed the use of *quintphones* as the base segment and included an unprecedented number of quantitative and positional information about speech segments at various levels of the utterance unavailable to concatenative systems. Similar ideas can also be found in some of the last unit selection systems that aimed at selecting the best fitting segments for a given context from their vast databases of segmented speech recordings. The size of the databases used in such systems was, however, the single biggest limiting factor and the eventual cause of abandoning of this technology. With the source openly available and an active online community of users³ the base idea in HTS was soon improved even further. The progress was made mainly in the formulation of the probabilistic models underlying the whole system and the machine learning pipeline in general [334]. An overview of the HTS speech synthesis method is presented in Figure 2.1. “In the training part, spectrum and excitation parameters are extracted from speech database and modeled by context dependent HMMs. In the synthesis part, context dependent HMMs are concatenated according to the text to be synthesized. Then spectrum and excitation parameters are generated from the HMMs by using a speech parameter generation algorithm. Finally, the excitation generation module and synthesis filter module synthesize speech waveform using the generated excitation and spectrum parameters” [304].

For many years, the quality of HMM-based systems was unequalled. It was only the original authors of HTS that pushed the bar even further and set a new state of the art by substituting the central model with a Deep Neural Network (DNN) [333]. This concept is also not a new one. Neural networks were used successfully for modeling speech before [267] and DNNs have been around since the 80s. However, it was the combination of the growing availability of cheap and fast large-scale computations, dedicated processing chips, developments in the neural architectures and the fact

³<http://hts.sp.nitech.ac.jp>

that it was implemented in a version of an already popular system, that all made up the recent success of the method.

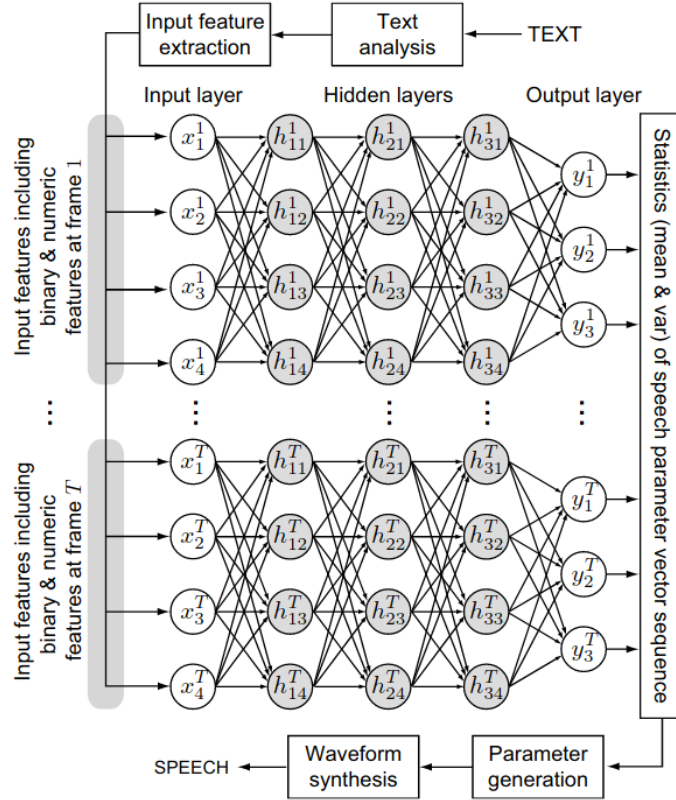


Fig. 2.2.: General DNN-based speech synthesis system. Adopted from Zen et al. [333].

The DNNs were first implemented as an alternative to HMMs. The latter were known to produce very natural sounding speech but were also suffering from a rather low quality of the audio signal itself. This was attributed partly to the method being based on vocoding, partly to the accuracy of acoustic models, as well as some issues connected with over-smoothing of the parameters [335]. The DNNs were largely introduced to solve the accuracy problem [333]. Figure 2.2 illustrates the original architecture of the DNN-based speech synthesis framework. The input text is first converted to a sequence of features $\{x_n^t\}$, where x_n^t denotes the n -th input feature at time frame t . The input features are comprised of binary answers to questions about linguistic contexts, including both questions about the identity of the linguistic unit and the relative and absolute position and length of the current linguistic unit. The input features are mapped to output features $\{y_m^t\}$ by a trained DNN using forward propagation, where y_m^t denotes the m -th output feature at frame t . The output features include spectral and excitation parameters and their first and second time

derivatives – velocity and acceleration. Similarly to the HMM-based synthesizer, speech parameters are then generated by setting the predicted output features from the DNN as mean vectors and pre-computed variances of output features from all training data as covariance matrices, the speech parameter generation algorithm can generate smooth trajectories of speech parameter features which satisfy both the statistic and dynamic features. Finally, a waveform synthesis module outputs a synthesized waveform given the speech parameters [303, 333].

Initial evaluation has shown that the quality and naturalness of speech generated with the DNN-based system had greatly surpassed that of the traditional HMM-based one. Importantly, as one of the main motivations for the implementation of DNNs in place of HMMs, the authors cite the fact that “the human speech production system is believed to have layered hierarchical structures in transforming the information from the linguistic level to the waveform level” [333]. The success of this work started a yet another revolution in the field with a number of systems being released every year based on new deep neural architectures and producing results that sometimes exceed even the most optimistic anticipations of the authors themselves.

2.2 Deep Neural Models of Speech

The last few decades had been an undeniable renaissance of deep learning in general. The first deep architecture that was successfully employed for modeling speech was the deep perceptron, a network consisting of a number of fully interconnected deep layers. Its authors motivated their architectural decision by the fact that “multiple hidden layers can represent some functions more efficiently than those with one hidden layer” [333].

The classical feed-forward multi-layered perceptrons (MLPs) were quickly replaced by the more sophisticated Recurrent Neural Networks (RNNs) [262] and especially the Long Short-Term Memory networks (LSTMs) [141].

“Recurrent networks [...] have an internal state that can represent context information. [...] [they] keep information about past inputs for an amount of time that is not fixed a priori, but rather depends on its weights and on the input data.” “A recurrent network whose inputs are not fixed but rather constitute an input sequence can be used to transform an input sequence into an output sequence while taking into account contextual information in a flexible way” [26].

These were the advantages that made the recurrent architecture ideal for modeling speech and language along many other problems connected with modeling of time series. However, RNNs have been found to frequently suffer from the vanishing gradient problem. LSTMs introduced a refined architecture that solved this issue improving the performance even further.

“Hence standard RNNs fail to learn in the presence of time lags greater than 5–10 discrete time steps between relevant input events and target signals. The vanishing error problem casts doubt on whether standard RNNs can indeed exhibit significant practical advantages over time window-based feedforward networks. A recent model, Long Short-Term Memory (LSTM), is not affected by this problem. LSTM can learn to bridge minimal time lags in excess of 1000 discrete time steps by enforcing constant error flow through *constant error carousels* (CECs) within special units, called cells” [109].

The technical idea behind these networks is best expressed through the following analogy by Graves and Schmidhuber [117]. “The Long Short Term Memory architecture was motivated by an analysis of error flow in existing RNNs which found that long time lags were inaccessible to existing architectures, because back-propagated error either blows up or decays exponentially.” As illustrated in Figure 2.3: “an LSTM layer consists of a set of recurrently connected blocks, known as memory blocks. These blocks can be thought of as a differentiable version of the memory chips in a digital computer. Each one contains one or more recurrently connected memory cells and three multiplicative units – the input, output and forget gates – that provide continuous analogues of write, read and reset operations for the cells. [...] The net can only interact with the cells via the gates.”

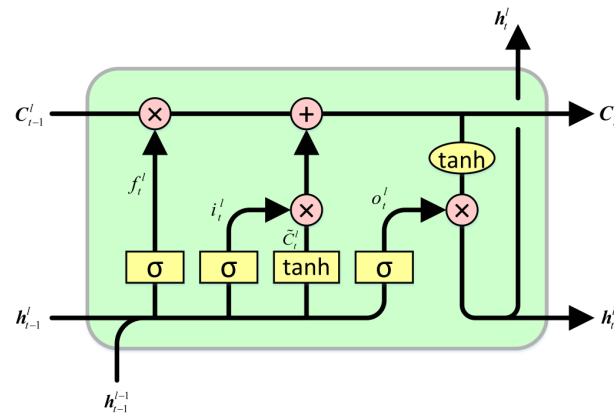


Fig. 2.3.: Long Short Term Memory gate scheme. Adopted from Hochreiter and Schmidhuber [141].

In many domains connected with processing sequential data, LSTMs along and their multiple different flavors like the Gated Recurrent Unit (GRU) networks [46] became the *de facto* standard for many years. It was only the recent successful adaptation of Deep Convolutional Neural Networks to the temporal data domain that allowed to redefine the state of the art once again.

2.2.1 Convolutional Neural Networks

As with most Deep Neural Network architectures, Convolutional Neural Networks (CNNs) are not a new concept [105, 52, 180]. They have been around for decades and except for a narrow range of applications remained dormant until the recent rapid developments in computing power and availability of large training datasets. The key to grasping the idea behind a CNN is understanding the convolution operation itself. It is best visualized in the domain of image processing where a small matrix - the filter kernel - is used for adding each consecutive pixel of the image to the neighbouring ones, weighted by the values in that kernel as shown in Figure 2.4.

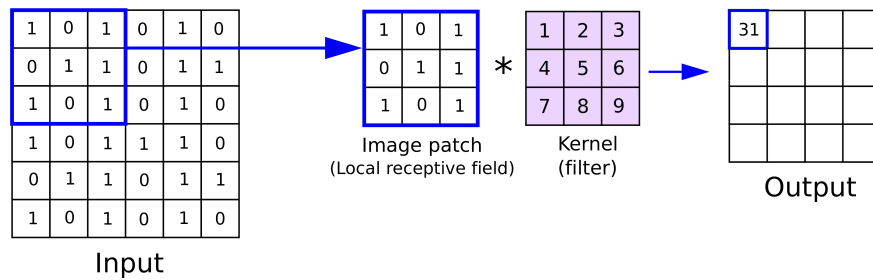


Fig. 2.4.: Example of a 2-dimensional matrix convolution.

This very simple operation allows for performing a range of useful filtering tasks, such as blurring, enhancing, edge detection, correlation filtering, etc. Some of the possible effects are demonstrated in Figure 2.5.

Convolutional Neural Networks exercise this simple idea by making the kernel itself a set of trainable hyperparameters of the network. As shown in Figure 2.6, on the example of LeNet-5 architecture [196], CNNs usually contain many consecutive convolutional layers, each of which contains a number of such convolutional filters, followed by a few fully connected layers. Each convolutional layer may be followed by some additional pooling or dropout layers [52, 180].

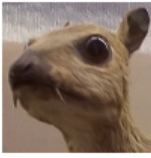
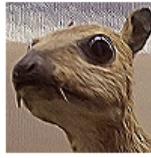
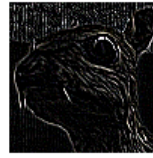
<i>Original</i>	<i>Gaussian Blur</i>	<i>Sharpen</i>	<i>Edge Detection</i>
$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	$\frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$
			

Fig. 2.5.: Examples of convoluting and image with different convolution kernels. Adopted from the Wikipedia [169].

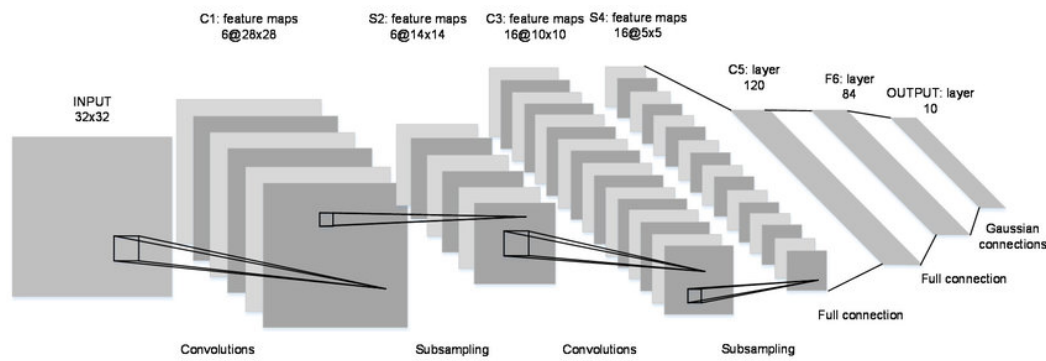


Fig. 2.6.: Convolutional Neural Network (LeNet-5). Adopted from LeCun et al. [196].

The pooling (also: sub- or down-sampling) operation reduces the dimensions of the data by combining the outputs of neuron clusters at one layer, also called feature maps, into a single neuron in the next layer computing a maximum (or an average) output value either globally or locally. This operation also helps simplify the complex and resource-consuming computations. Dropout layers simply forget some part of the data in order to help the network generalize and prevent overfitting [286], which is one the typical problems plaguing neural architectures with higher numbers of trainable parameters.

With such a setup, the network can discover and learn the most significant features of the image for a given problem and class. With the use of such consecutive convolutions and multiple processing channels, the network can detect very complex features. For example, it may learn to consecutively detect edges, shapes, patterns and eventually whole complex objects like legs and wings. This idea was a huge step forward as feature extraction now became a fully integrated automatic step

performed by the network itself instead of an arduous manual task preceding the actual training. Automatically extracted features are usually also much more relevant for a given problem and contain less inherent bias than a manually devised feature set. Image CNNs can also detect useful spatial relations in the data that would normally be inaccessible to the human eye.

2.2.2 Neurobiological Foundations

The first working description and implementation of a Convolutional Neural Network was the Neocognitron [105]. The original idea was based on one of the most important findings in neuroscience made by Hubel and Wiesel [146, 147], for which they have won the Noble Prize. In a famous experiment on cats they have shown that neural processing of information in the visual cortex occurs in two successive stages with the participation of simple and complex cells. These cells were defined based on the differing construction of their receptive fields. This discovery laid foundations for a new theory of how humans process visual information.

“The receptive field is a portion of sensory space that can elicit neuronal responses when stimulated. The sensory space can be defined in a single dimension (e.g. carbon chain length of an odorant), two dimensions (e.g. skin surface) or multiple dimensions (e.g. space, time and tuning properties of a visual receptive field). The neuronal response can be defined as firing rate (i.e. number of action potentials generated by a neuron) or include also subthreshold activity (i.e. depolarizations and hyperpolarizations in membrane potential that do not generate action potentials)” [8].

Although, there is an ongoing debate between the advocates of hierarchical, parallel and recurrent models of neural processing in the primary visual cortex, most of the underlying circuitry, and the idea that the receptive field structures become increasingly complex at successive stages of the visual pathway, remains the same [215]. According to Hubel and Wiesel [146, 147], simple cells have separate on- and off-subregions that can be mapped with small spots of light. Examples are presented in Figure 2.7 with red crosses representing on-subregions and blue triangles off-subregions. Hubel and Wiesel [146, 147] classified these cells based on the following criteria:

- i “they were subdivided into distinct excitatory and inhibitory regions”
- ii “there was summation within the separate excitatory and inhibitory parts”

iii “there was antagonism between excitatory and inhibitory regions”

iv “it was possible to predict responses to stationary or moving spots of various shapes from a map of the excitatory and inhibitory areas”

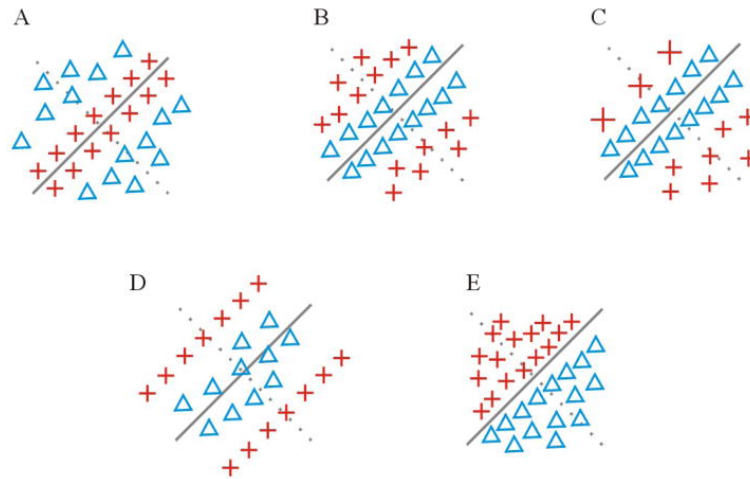


Fig. 2.7.: Simple receptive fields. Adopted from Hubel and Wiesel [147].

On the other hand, the *complex cell* is an umbrella term for many different cortical cells that are not simple, i.e.: any cortical neuron that does not have a simple receptive field. These were found to form a very diverse population. Three different examples of such cells are presented in Figure 2.8. Cell A generates on-off responses throughout the entire receptive field. Cell B responds exclusively to a black horizontal bar. Cell C has partially separated on- and off-regions but the receptive field cannot be mapped with small spots of light. The green icons on the left represent the complex receptive field with the stimuli (flashed bars or borders) overlaid [215].

According to most thalamocortical models (i.e. parallel and hierarchical), as shown in Figure 2.9, stimuli are processed in a series of visual areas. V1 neurons are most sensitive to low-level features, such as edges and lines [209, 146, 147]. In higher visual areas, like V4 and IT, receptive fields are larger, and neurons are sensitive to complex features, such as shapes and objects. Responses of high-level neurons are fully determined by the neural firing of lower-level neurons. For example, the neural firing to a square is determined by the neural firing for two vertical and two horizontal lines. Although the recently proposed recurrent model suggests that simple and complex cells originate from the same cortical circuit operating at different amplification gains, the main computational assumption that simple cells have segregated subregions within their receptive fields that respond to either the

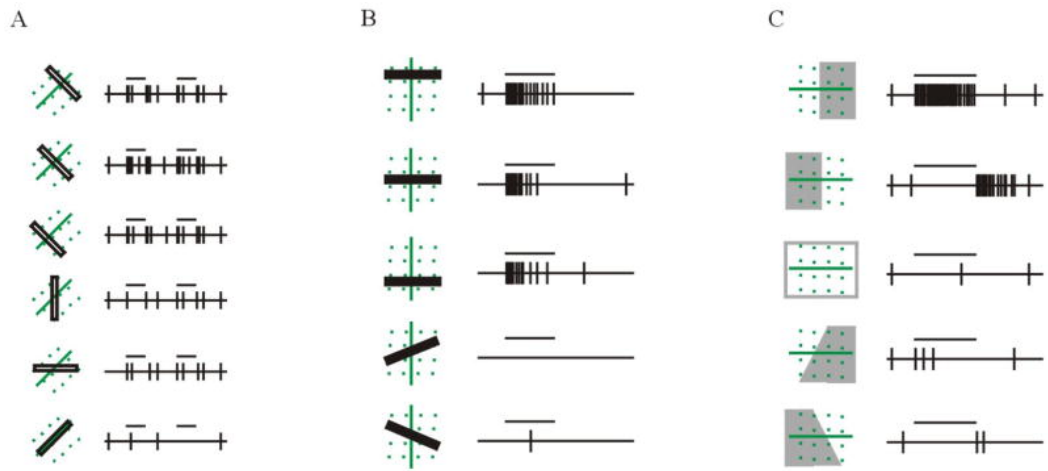


Fig. 2.8.: Three different types of complex receptive fields. Adopted from Hubel and Wiesel [147].

on- or offset of a light bar and by spatial summation within each of these regions and complex cells had on- and off-regions that were coextensive in space.

It can be observed that the simple and complex receptive fields and the layered processing of visual information altogether can be very effectively approximated with convolution and pooling operations implemented within deep multi-layered neural network architectures as seen in most of the state-of-the-art CNNs nowadays. Although the first implementation went by rather unnoticed, it was revisited several years later with a more practical approach by LeCun et al. [196], and it is this work that is now considered the starting point of Deep Convolutional Neural Network research.

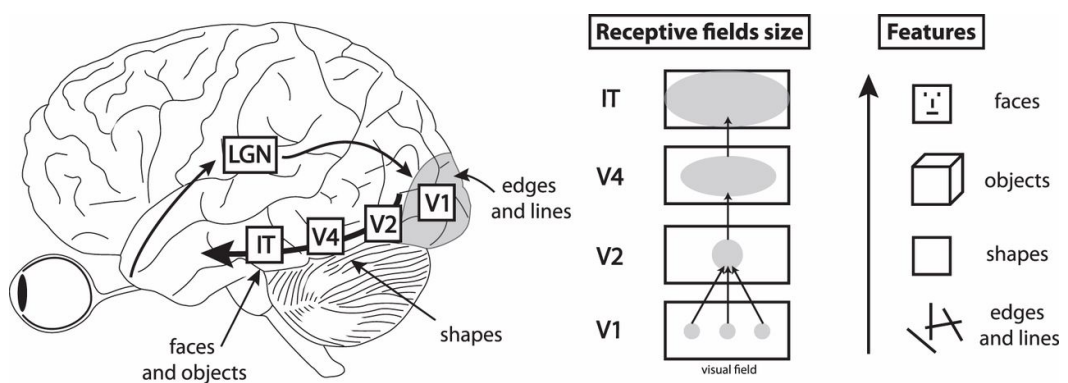


Fig. 2.9.: Hierarchical, feedforward visual processing in human brain. Adopted from Manassi et al. [208].

Although the mathematical formulations of these networks have existed for almost half of a century and first proof-of-concept implementations have been available for quite some time – their potential remained dormant because of their extremely high computational needs, at least for that time. As in most cases with deep learning, it was only the recent deep learning renaissance that brought the CNNs back into attention [135, 135, 179, 80, 280, 332, 44, 253]. They not only prove to outperform any other known technology in the domain of image classification and many other visual tasks but also comprise a perfect framework for computational neuroscience experimentation [207] because of their deep neurobiological foundations.

Speech and Intonation

As shown in recent electrocorticographic study on neurosurgical patients [295], intonational pitch is represented by a highly specialized and dedicated neural population in the human auditory cortex. The phonetically invariant representation of speaker-normalized relative intonation contours suggests that intonation is encoded as an isolated pitch contour, irrespective of any lexical information or phonetic content [305]. At the same time, these auditory cortical areas of the superior temporal lobe are important for the formation and maintenance of motor commands and auditory targets for speech production [168]. As reported by Guenther [122]:

“Using a combination of neural network modeling, neuroimaging, and auditory perturbation experiments, we have characterized the network of brain regions involved in auditory feedback control of segmental aspects of speech. This network involves auditory error cells in bilateral posterior superior temporal cortex which become active when the current auditory feedback mismatches the auditory target for the current speech sound. Projections from these auditory cortical areas to the right hemisphere ventral premotor areas, then on to primary motor cortex, transform perceived auditory errors into corrective movement commands for the speech articulators.”

This findings fit well within the theory of the phonetic-phonology interface central to some of the currently dominating models of intonation, such as the Autosegmental Metrical model [24, 124, 187, 245] as well as within the general model of speech production (model of speech production is presented symbolically in Figure 2.10 with boxes representing processing components and circles and ellipses representing knowledge stores). They all share a common general assumption of some finite lexicon of discrete building blocks for the linguistic component of intonational contours which constitute abstract targets for the eventual concrete phonetic implementation.

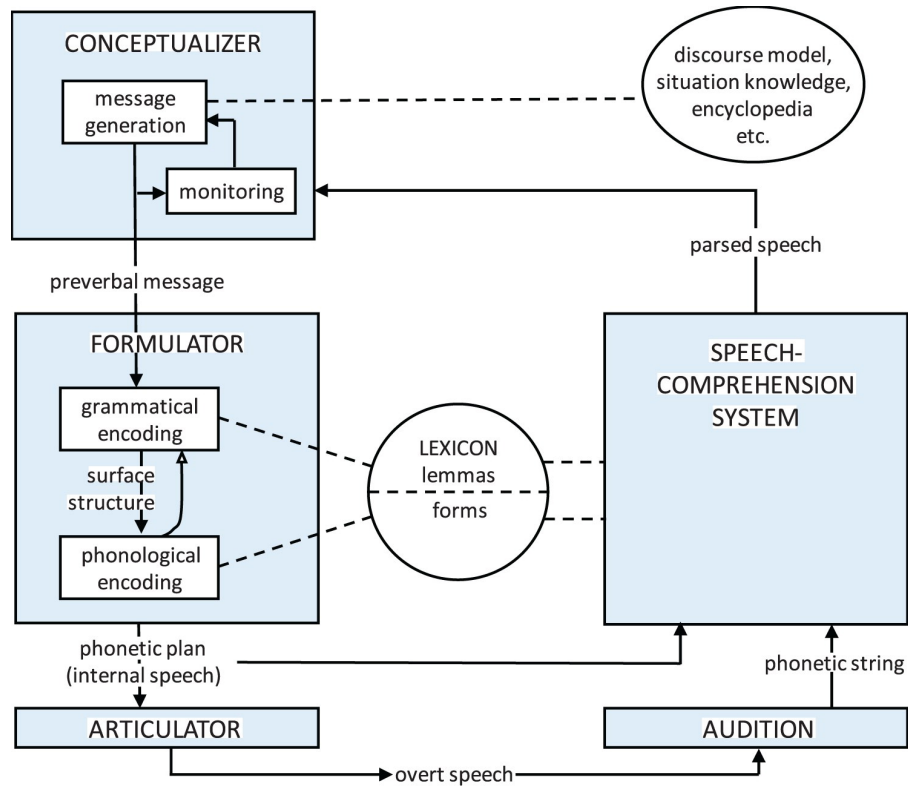


Fig. 2.10.: A blueprint for the speaker - model of speech production. Adopted from Levelt [202] and A. Meyer et al. [223].

The categorical nature of these intonation phonemes is further supported by numerous neuroscientific studies showing that the human brain separates the paralinguistic and linguistic components of intonation and processes them in terms of gradience or discrete categories through separate streams and even demonstrating varying activations onsets [91, 224, 225, 257, 324, 249].

Although the location of specific brain areas responsible for speech and intonation production and comprehension at different stages, from conceptual semantic and syntactic to the eventual articulatory, has been studied pretty well, it is the actual underlying computation and encoding of these linguistic intonational categories that pose a much bigger problem and are much more interesting from the perspective of linguistics. Most of the cognitive categories assumed by the currently dominating phonological models are usually rather ambiguous and necessarily biased by their author's intuitions. These categories are often plausible from the philosophical point of view and within a specific grammatical system within which they were devised but providing hard experimental evidence to support their existence often proves more challenging, even though triangulation is becoming a widely accepted

approach nowadays and more and more scientists employ experimental methods to phonology [237] or propose hybrid approaches such as the cognitive phonetics manifesto by Tatham [296]. As an example, it is still unclear whether the H and L tones, as proposed in the AM model, are identical with the actual mental categories of linguistic intonation. Even if they do exist in the mind, as might be suggested by psychological experimentation, they might as well reside at some completely different cognitive level. And even if they do exist as part of the phonetic-phonology interface it is still uncertain if they do not inhabit some intermediate level instead, such as (iii) in the following idea of abstraction levels by Shaumyan [272]:

- i constructs or phonemes, at the semiotic level, which are “free from any physical substance”,
- ii observation of sounds at the physical level,
- iii in between, a third level of relational physical elements, also called phonemoids and differentoids, which come into play as an interface: phonemoids and differentoids are defined by physical elements from the level of observation and “stand in the relation of differentiation to signs”.

The growing list of new experimental techniques in neuroscience allows a much deeper look into how the brain might implement these phonetic–phonology mappings, i.e. how the brain can infer the target intonational contour based on some features of the planned utterance, or how it might decompose and map the perceived intonational contour into cues to linguistic structure of the perceived utterance.

Numerous studies show that, similarly to the visual cortex, “speech comprehension involves hierarchical representations starting in primary auditory areas and moving laterally on the temporal lobe” [65]. Moreover, Tian et al. [301] have shown that analogues of visual simple and complex cells can also be found in the auditory cortex. This study used the original classification criterion as proposed by Hubel and Wiesel [147]:

“Simple cells were originally defined by the existence of segregated subregions within their RF that respond to either the on- or offset of a light bar and by spatial summation within each of these regions, whereas complex cells had on- and off-regions that were coextensive in space. [...] Here we report that response profiles of neurons in primary auditory cortex of monkeys show a similar distinction: one group of cells has segregated on- and off-subregions in frequency space; and another group shows on- and off-responses within largely overlapping response profiles.”

As they report, these findings support the existence of a common canonical processing algorithm within cortical columns. If both mechanisms seem to share the same hierarchical nature, with both convergent and divergent information flow in successive, nested processing layers [180, 196, 256, 270] it is reasonable to suspect that these hierarchically organized simple- and complex-type cells in the auditory cortex might be capable of extracting levels of gradually complex latent features from the input signal similarly to the CNN models of visual processing introduced above. The obvious differences are in the nature of the input signal itself. Although they are not as different as one might initially assume. Linearized neurocomputational models have demonstrated the existence of strong spectrotemporal and phonetic feature representations in superior temporal gyrus (STG) [65, 83, 148, 221] and motor cortex [45]. The Spectro-temporal receptive field and the 2-dimensional representation of the sound signal as inherited from the cochlea makes it much more similar to its visual counterpart. As defined by Shamma [271]:

“The spectrotemporal response field (STRF) of an auditory neuron is a time-frequency measure of the dynamic responses of an auditory neuron to impulsive energy delivered at various frequencies. As such, it gives simultaneously two types of information about the neuron. The first is its frequency tuning, or more specifically which frequencies excite the cell best and which inhibit it. The other is the nature of its temporal response, i.e., whether it is sustained in time or is rapidly adapting. This measure is linear and takes the stimulus spectrogram as its input and hence is often found to be useful in predicting responses of a neuron to unseen stimuli.”

2.2.3 Deep Temporal Convolutional Neural Model of Speech

Unprecedented performance of CNNs in visual tasks triggered a number of attempts to apply them to different problem domains. With first notable examples by H. Lee et al. [197] and Hau and K. Chen [129] CNNs started becoming popular also in the context of modeling speech, especially in Automatic Speech Recognition (ASR) systems [2, 145, 241, 203, 254].

In a 2016 paper [239], a Google-owned company Deepmind⁴ presented a fully autoregressive probabilistic end-to-end speech synthesizer with a CNN at its core – the WaveNet. The idea was based on Google’s PixelCNN [312] – an image generation network. WaveNet is using a technique called dilated causal convolutions. The causality is expressed through the joint probability of the modeled waveform $\vec{x} =$

⁴<https://deepmind.com>

$\{x_1, \dots, x_T\}$ being factorized as a product of conditional probabilities of all previous timesteps [239], i.e.:

$$p(\vec{x}) = \prod_{t=1}^T p(x_t | x_1, \dots, x_{t-1}) \quad (2.1)$$

Dilated convolutions are a concept borrowed from a wavelet transform algorithm called *à trous* (French for *with holes*). It was primarily designed to overcome the lack of translation-invariance of the discrete wavelet transform (DWT) which was achieved by removing the downsamplers and upsamplers in the DWT and upsampling the filter coefficients by a factor of $2^{(j-1)}$ in the j -th level of the algorithm [5, 275, 298, 299]. In WaveNet this concept was applied in order to increase the receptive fields of the network. Figure 2.11 illustrates how a stack of dilated causal convolutions with dilation factors $[1, 2, 4, 8]$ allow for a receptive field of 16 time steps for the generation of a single output.

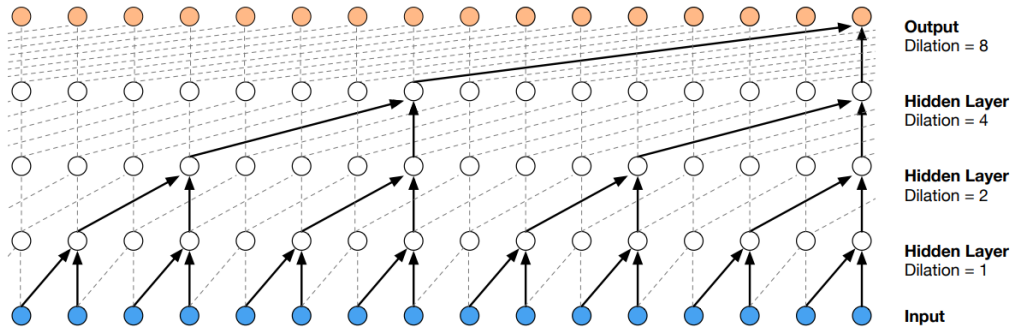


Fig. 2.11.: Dilated causal convolutions. Adopted from the original WaveNet paper [239].

Another key concept used in the PixelCNN network that was also applied to WaveNet was the introduction of gated convolutional layers, defined as:

$$\vec{z} = \tanh(W_{f,k} * \vec{x}) \odot \sigma(W_{g,k} * \vec{x}), \quad (2.2)$$

where $*$ denotes a convolution operator, $\odot$ denotes an element-wise multiplication operator, $\sigma(\cdot)$ is a sigmoid function, k is the layer index, f and g denote filter and gate, respectively, and W is a learnable convolution filter.

They have been introduced mainly to account for the single advantage that generative LSTMs had over the PixelCNN architecture, which allows every layer in the network to access the entire neighbourhood of previous pixels, while the region

of the neighbourhood available to pixelCNN grows linearly with the depth of the convolutional stack [312]. Gated activation units control what information will be propagated throughout the remaining layers and thus establish a kind of selective *long term memory*. These gated activation units are additionally implemented with residual [132, 239] and skip connections as demonstrated in Figure 2.12. These in turn help avoid the vanishing gradient problem which typically affects very deep networks and has been shown to improve the performance of modeling long sequences with complex spatial relations.

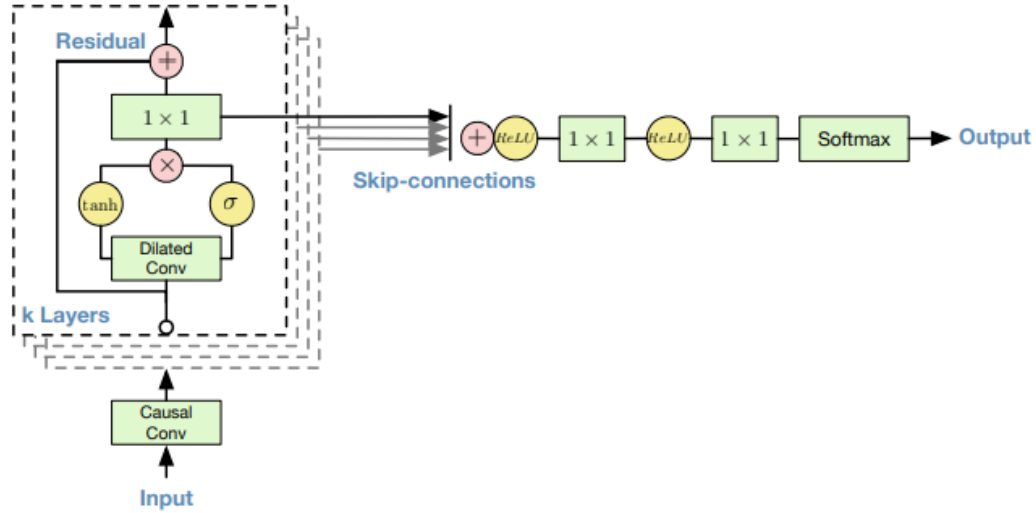


Fig. 2.12.: Residual and skip connections from a stack of k gated convolutional layers. Adopted from the original WaveNet paper [239].

The original WaveNet paper reports a Mean Opinion Score-based perceptual evaluation of the synthetic speech output by the network compared to Google’s other top speech synthesizers built on that same corpora. As shown in Figure 2.13, the network was achieving unprecedented results. Also the speech (and music) samples published online demonstrate previously unthinkable quality and naturalness, with previously unheard details such as breathing and mouth clicks in a number of online samples⁵.

Since then a number of speech synthesizers based upon the general idea of a Temporal Convolutional Network (TCN) [194, 20, 19] have been built, each improving the results attained by its predecessors even further. TCN became an umbrella term for the many different flavours of the same base idea.

⁵<https://deepmind.com/blog/article/wavenet-generative-model-raw-audio>

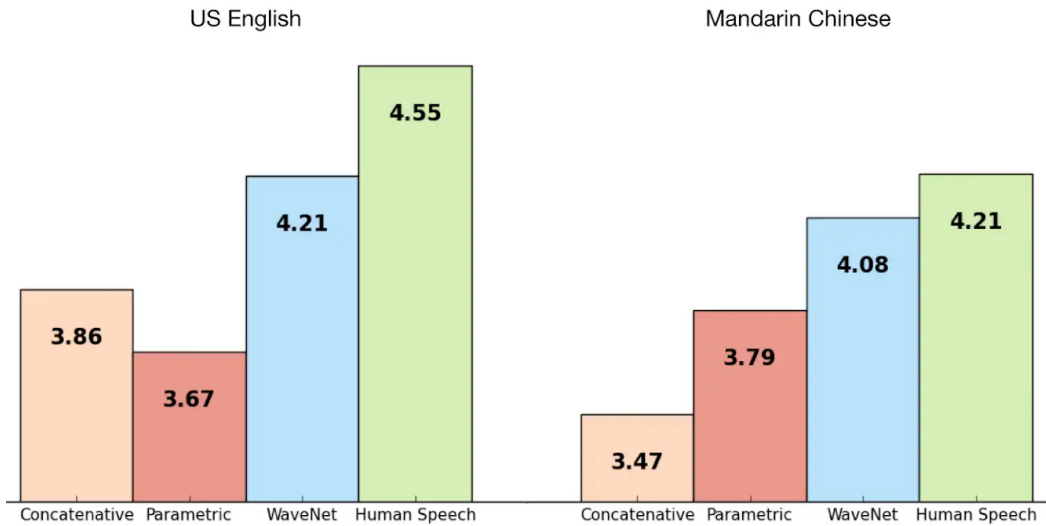


Fig. 2.13.: Google WaveNet evaluation results as compared with Google’s best concatenative and parametric systems. Adopted from Oord et al. [239].

Tacotron 1 & 2 [320, 274] introduced the first of the notable improvements. As illustrated in Figure 2.14, they implemented a sophisticated neural architecture that used the WaveNet as one of the central building blocks along with a number of other state-of-the-art methods like transforming the input text into character embeddings, attention mechanisms and additional LSTM blocks.

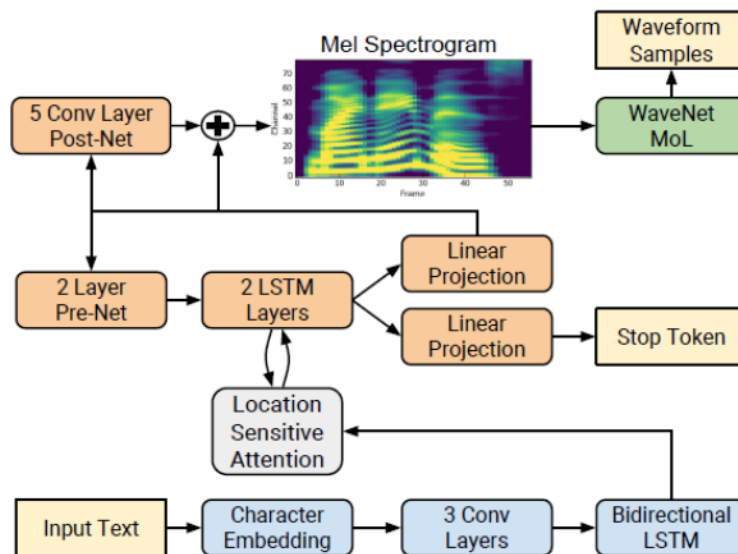


Fig. 2.14.: Tacotron 2 architecture. Adopted from Shen et al. [274] (©2018 IEEE).

The network was not only able to synthesize speech directly from text input, which demonstrated its outstanding ability to discover, extract and model complex relations,

but also raised the bar in terms of the quality of the synthesized speech even further⁶. The results were soon matched by a number of Baidu's⁷ DeepVoice synthesizers [14, 111, 248]. The DeepVoice systems demonstrated how well the underlying networks can mimic speaker characteristics by conditioning them on small samples of speech from a specific speaker while being trained on big multispeaker datasets beforehand. The improvements from that point were made mostly by applying optimizations to the training algorithms and using even bigger datasets for training.

This very recent leap forward in the quality of the TTS systems brought the TCNs into attention of other fields of science, also those studying human speech and language from a different perspective. Although the examples are still few, some noticeable work can be found in computational neuroscience. Angrick et al. [10, 11] have demonstrated good results in an attempt to directly synthesize continuously spoken speech from neural activity using electrocorticography in patients suffering from intractable epilepsy. Tamm et al. [294] have used a novel EEG-based method to recognize imagined vowels using a CNN and Cooney et al. [56] have done a similar study for the recognition of word-pairs.

The growing body of evidence suggesting that auditory processing of speech perception and production can resemble that in the visual cortex and the undeniable appeal of TCNs as models of neural computations occurring during these processes brought about a need to have a deeper look into the internal workings and representations of the networks, which to this point have been a traditional black box.

2.2.4 Deep Neural Network Explainability

Although the DNNs are popularly used as part of a scientific methodology in computational neuroscience and a number of heuristics have been used to infer the internal states of the network, they were long considered a typical black box. The visual CNNs provided an easy peek into how the network is processing the input through the highly interpretable convolutional filters. The learned parameters of the filters can be simply visualized in themselves as images or can be manually applied to any image and the output can be simply inspected by eye, revealing the visual effect of the operation. However, the convolutions in the deeper layers are sometimes less intuitive as they are applied on the output of the preceding layers' filters. Figure 2.15 illustrates the visualization of convolutional filters learned on a typical image classification dataset. Filters' kernels in lower layers are similar to the receptive

⁶https://rayhane-mamah.github.io/Tacotron-2_audio_samples/

⁷<http://research.baidu.com/Blog/index-view?id=91>

fields of simple cells in the human visual cortex, responding to simple lines and edges. The filters in higher layers are shown to respond to increasingly complex patterns and shapes with actual objects like dogs noses and human faces in the final layers of the network.



Fig. 2.15.: Learned convolutional filter visualization example. Adopted from Zeiler and Fergus [332].

In order to provide an even deeper insight into how parts of the input information are used by the network, a method based on the principle of backpropagation [262],

which is also the core method used in the training of the deep learning models, was proposed. It allowed to calculate feature saliency by computing the influence of a tensor *w.r.t.* another tensor on the final loss. The influence of the input on any given filter is this filter's saliency map [18, 279]. Since then, many different variations of this method including both improvements in the formal mechanism as well as the implementation details themselves were introduced.

Comparison of Sensitivity-based and Relevance-based Methods

We can relate to any of these methods as sensitivity analysis as they are all based on the effect of infinitesimal perturbations of the input values on the function value itself, e.g. $(\partial f / \partial x_p)^2$. The main drawback of this approach was very clearly explained by Samek et al. [264] with an analogy to a linear classifier, i.e. $f(x) = \sum_p x_p w_p$, for which sensitivity would be calculated as $R_p = w_p^2$ (or $R_p = |w_p|$ depending on the variant). Since it does not include the actual input activations (so the actual presence of some feature) and only the information about the inputs the classifier reacts to ($w_p > 0$, meaning which inputs if modified make the input data more or less belong to some class), it does not provide an optimal explanation. As shown in Figure 2.16, “regions consisting of pure background, e.g., the empty street, have large sensitivity, although these pixels are not really indicative for this image category. However, if we put motorbike-like structures at these particular locations, then this change would certainly increase the classification score.”

Another method called Layer-wise Relevance Propagation (LRP) was proposed [17, 264, 234] to address this issue. Figure 2.16 shows how the quality of the explanation provided by 3 different variants of this new method compares with the explanation provided by classical sensitivity-based methods in a perturbation study, where a number of image pixel values are ablated or randomized. The results are calculated as the Area Over Perturbation Curve (AOPC) relative to random output. For more information on this study please see the original paper by Samek et al. [264].

Details of the Layer-wise Relevance Propagation Method

The relevance in this framework can be defined as:

$$R_j = \sum_k \frac{a_j w_{jk}}{\sum_{0,j} a_j w_{jk}} R_k \quad (2.3)$$

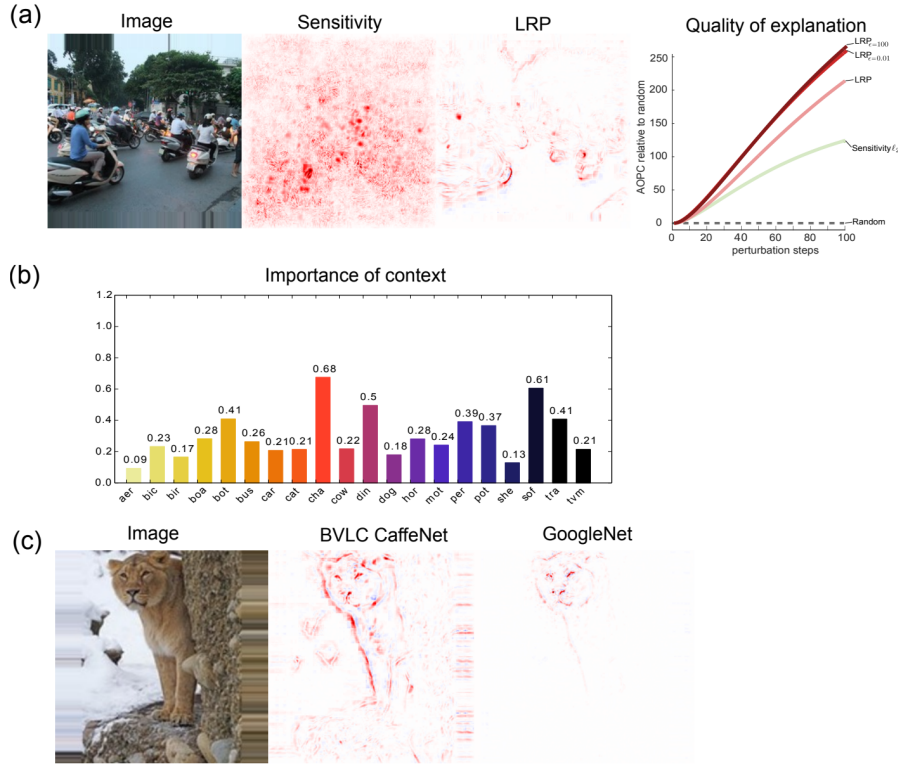


Fig. 2.16.: Results of sensitivity-based and relevance-based explainability methods. Based on Samek et al. [264]

With j and k being the indexes of neurons on two consecutive layers, a symbolizing the activation of a neuron, and w denoting the weights. This is the simplest form of LRP (LRP-0) and although many different flavors have been developed the main idea stays the same. The numerator in Equation 2.3 denotes the amount of influence of neuron j on neuron k in case of an active Rectified Linear Unit (ReLU) activation when it becomes linear. It is divided by the sum of contributions of all neurons of the preceding layer. This enforces the conservation principle of LRP which states that the magnitude of any output y is conserved through the backpropagation process and is equal to the sum of the relevance R of the input layer. This property holds for any consecutive layers j and k , and by transitivity for the input and output layer. The procedure of calculating relevance employs a method called Deep Taylor Decomposition [17, 233]. Figure 2.17 illustrates the computational flow of this method. A prediction for the class “cat” is obtained by forward-propagation of the pixel values, and is encoded by the output neuron x_f . The output neuron is assigned a relevance score $R_f = x_f$ representing the total evidence for the class “cat”. Relevance is then backpropagated from the top layer down to the input, where R_p denotes the pixel-wise relevance scores, that can be visualized as a heatmap.

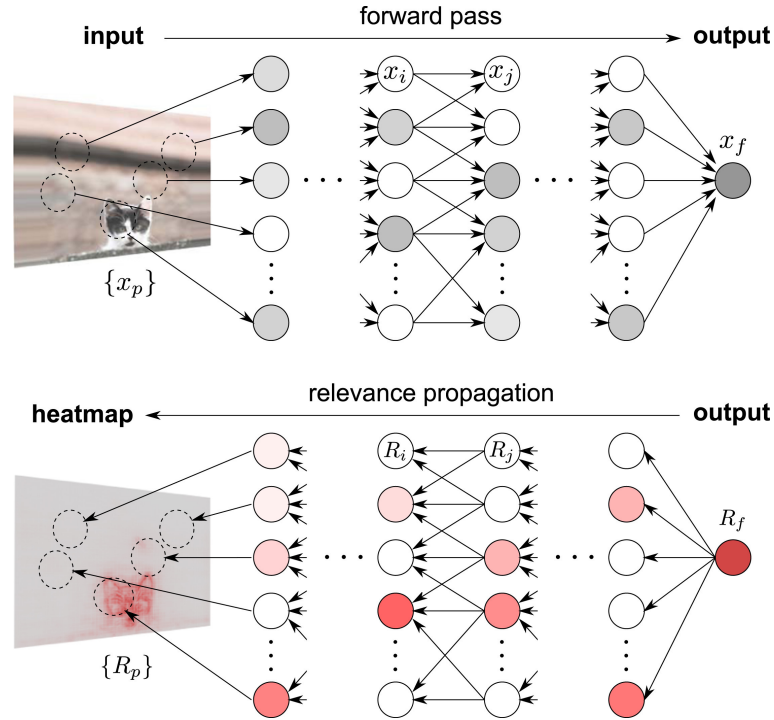


Fig. 2.17.: Computational flow of deep Taylor decomposition. Adopted from Montavon (2017) [233].

In case of ReLU activations, the first step consists of a typical forward pass with a small constant added to the whole expression to prevent later division errors. The weight can be also passed as an argument to an optional p function which differentiates all the different flavors of the LRP method. Next the relevance of the preceding layer is divided by the results of the forward-pass in order to apply the conservation principle. The third step implements a backward pass calculating the amount of relevance that trickles down to neuron j from the succeeding layer. Finally the relevance coming from above is multiplied with the activation of the neuron to calculate its own final relevance. In this way the component missing in the sensitivity based methods is introduced as we include both the activation of a neuron, and how much it contributes to the output (or the relevant neurons in the succeeding layer). In this way the method is able to produce much more relevant explanations as shown in Figure 2.16.

Explainability Methods for Temporal Data

Currently, a whole range of different methods for visualizing and explaining the internal states of the network are available, e.g.: SmoothGrad [281], Integrated-

Gradients [290], DeconvNet [332], Guided Backpropagation [284], PatternNet and PatternAttribution [171], the discussed LRP [17, 190, 234] and Deep Taylor Decomposition [233, 6], and Deconvolution Network [329]. Image classification became the go-to problem domain for the development, evaluation and implementation of these explainability algorithms, as the visual modality ensures easy manual inspection and interpretation of the results. LRP has been also successfully applied to some other problems and models as listed by Samek et al. [264], i.e. bag-of-words models [17], Fisher Vector and SVM classifiers [190], identification of relevant words in text documents [16], visualizing facial features related to age, happiness and attractiveness [13] and identifying relevant spatio-temporal EEG features in the context of brain-computer interfacing [289].

These methods can usually be extended for time series data too but the domain has remained relatively unexplored [276]. Some notable examples include the works of Kumar et al. [184] and Siddiqui et al. [277]. The latter work provided a fully functional implementation of a range of explainability methods aimed at the analysis of time-series models – the TSViz Python library. The examples covered in the original paper are mostly considered with time series forecasting problems, where a number of following outputs are predicted given a number of preceding samples, or with time series anomaly detection, which is mainly a classification problem.

“TSViz provides possibilities to explore and analyze the network from different dimensions at different levels of abstraction which includes identification of the parts of the input that were responsible for a particular prediction (including per filter saliency), importance of the different filters present in the network, notion of diversity present in the network through filter clustering, understanding of the main sources of variation learnt by the network through inverse optimization, and analysis of the network’s robustness against adversarial noise. As a sanity check for the computed influence values, we demonstrate our results on pruning of neural networks based on the computed influence information” [277].

Another notable example can be found in Lauritsen et al. [191], where the potential of explaining networks build for temporal data was demonstrated in a study, in which the authors predicted acute critical illness from electronic health records and then analyzed the cues to these predictions as extracted from the network itself using the explainability algorithms. This work employs another implementation of explainability methods from Alber et al. [6]⁸, which supports relevance calculation for time-series data out-of-the-box.

⁸<https://github.com/albermax/investigate>

The availability of such interpretability techniques will surely trigger many more studies focused on exploring processing patterns in biologically-inspired neural networks acting as models of human cognition for different modalities. Until now, to the best of the author's knowledge, no such study considering the perception or production of speech and intonation has been made.

2.3 Intonation Modeling

As defined by Féry [94], intonation is:

“The tonal structure of speech expressed by the melody produced by our larynx. It has a phonetic aspect, the fundamental frequency (F_0), and a grammatical (phonological) aspect.”

However, in the current work we will understand the melody in the above definition as only these features that pertain directly to the variations in the perceived pitch and not to the duration, intensity, voice quality or any other features, which will be here considered as part of the broader category of speech prosody. The term prosody is often used interchangeably with intonation but here we adopt the understanding proposed by Trubetzkoy [307], where it is defined as all rhythmic-melodic aspects of speech. So by prosody we will understand all suprasegmental features of speech including the intonation itself, duration, intensity and voice quality.

The second part of the definition given by Féry [94] is a good example for what Rossi [260] emphasized extensively – all definitions of intonation “are epistemological definitions, i.e., not a priori programmatic definitions, but a posteriori statements of a practice and methodology.”

These differences in attitudes, in turn, can usually be somehow traced back to the scientific tradition, or school, that influenced the scholars working on the problem. So in order to fully understand this extreme epistemological polarity it is necessary to understand the history of the western intonation research. Many fine reviews of past (and present) intonation models can be found in the literature, e.g. in works by Pierre and Martin [244], Di Cristo [82, 81], Gibbon [110], Bertinetto [28], Rossi et al. [261] and Rossi [260], Selkirk [269], Cruttenden [59], Ladd [186], Cutler et al. [62], Lacheret-Dujour [185] and Botinis et al. [38], and hence this work does not aim at providing yet another one. However, it is worth mentioning at least a few of the ideas that constitute the preliminaries to modern intonational studies, including this one.

Most of the current linguistic models of intonation are eventually phonological in nature [260] and deeply rooted in one of the main schools of phonology, American or European⁹. Moreover, they can be also divided into those that attempt to model only a single level of abstraction of the phenomenon, and those that try to approach it more holistically covering all of its aspects, from the abstract and cognitive to the ultimate concrete and physical. However, they are all mostly founded on that same assumption that intonation is a dualistic phenomenon that can be seen as a mapping from a system of discrete cognitive units to contours of fundamental frequency and the other way round. This perspective surprisingly puts little emphasis on modeling the mapping itself which, as already mentioned, seems to be at the very heart of the problem.

2.3.1 European School

In Europe, intonation was, and still is, studied mainly in relation to its function and meaning, following the main tenets of the Prague School of phonology and de Saussure's structuralism [67]:

- i Saussure argued for a distinction between *langue* (an idealized abstraction of language) and *parole* (language as actually used in daily life). He argued that a *sign* is composed of a *signified* (*signifié*, i.e. an abstract concept or idea) and a *signifier* (*signifiant*, i.e. the perceived sound or visual image).
- ii Because different languages have different words to refer to the same objects or concepts, there is no intrinsic reason why a specific signifier is used to express a given concept or idea. It is thus *arbitrary*.
- iii Signs gain their meaning from their relationships and contrasts with other signs. As he wrote, in language, there are only differences *without positive terms*.

De Saussure in his *Course in General Linguistics*, prescribes that linguists should focus not on the use of language (*parole*), but rather on that underlying system (*langue*). The Prague linguistic circle followed these ideas and consequently saw that the analysis of language should be a study of contrastive features. In phonology that was most evident in the consideration of minimal pairs – two different words

⁹The author was not able to find any comprehensive reviews or mentions of historical intonation research in other parts of the world. Although they might have existed as part of the rich philosophical traditions of China or India the current scene is dominated by the aforementioned lines of thought.

that contrast only in terms of a single speech sound, as the necessary requisite for distinguishing different phonemes in the inventory of speech sounds of a language. An intonational phoneme, a small meaningless unit that can be used to produce a large number of meaningful elements (morphemes), was therefore highly needed to fit in the programmatic assumptions [308]. Initially, some saw intonation as lacking the double articulation and simply considered it as an acoustic substance that links directly to meaning. The resulting models were pertaining directly to the intonational contour. Others attempted to devise systems of intonational phonemes and morphemes. The latter was also the more convenient angle as the substance was much more difficult to tackle than the content. It was always, however, in direct relation to their function.

The substance-based approach is also very noticeable in the Leningrad School of linguistics, based upon the work of Baudouin de Courtenay, where the prerequisite for a phoneme was the language speakers' awareness of it as a distinguishable category of sounds. This view required the phonological analysis to start from the actual acoustic substance. Some followers of this trend even argued that morphophonology does not require any further levels of abstraction below, except for the necessary concrete phonetic level. Prague School was focused more on content. Their mostly theoretical approach to intonation assumed its strong link to other linguistic constituents of a sentence as in the theme-rheme theory developed by Mathesius [216], Karcevskij [165], Groot [120], Daneš [64] and others. This idea was later revisited and improved upon by Ladd [186], as it gained additional momentum with Selkirk's [269] cognitive levels hierarchy and the related interpretability condition by Hirst, Di Cristo and Espesser [138, 260] stating that:

"Representations at all intermediate levels must be interpretable at both adjacent levels: the more abstract and the more concrete". Hence "functional representations which encode the information necessary for the syntactic and semantic interpretation of the prosody."

A number of superpositional models were also defined based mainly on the work of Öhman [238] and Fujisaki [103], like those by Möbius et al. [229] and Möbius [228], Mixdorff [227] and J. Van Santen et al. [316]. These models assume that sentence and word intonations can be modeled with separate components superimposed over some baseline and have strong articulatory and physiological foundations.

On the other hand, Gårding et al. [107] and Gårding [106], Botinis [37], Bannert [21], Grønnum [119] and Hjelmslev and Whitfield [140] argued that intonation should be first approached independently of other levels of language to account

for its concrete physical realization first and only then for the underlying content. This view was also shared by some followers of Pierrehumbert's model [245] and the proponents of the IPO model [292] whose theoretical foundations resemble the behavioral views of Bloomfield [34] in the American School.

Most contemporary European work on intonation in the European School is founded on some of the above principles, whether these are the works by descriptivists like Jassem, Crystal, Cruttenden, Fonagy, Kratochvil, Potapova, Svetozarova [137], or by intonational morphologists like Brazil, Gussenhoven, Couper-Kuhlen, Kohler, Martin or Rossi [260].

2.3.2 American School

The aforementioned behavioral model by Bloomfield, which treated words as stimuli and meaning as the speaker's reaction to that stimuli, gave rise to what we now refer to as the American School of Phonology as these assumptions had evident epistemological consequences. It suggested that the phonological study of language should be concerned only with the actual speech [34]. As Rossi [260] notes, Harris [128] claimed that the distributional features are sufficient for a definition of a phonological system. Phonetic features were widely used as in the work by [306]. Under this bottom-up concrete-to-abstract view, prosodic features were also seen as part of linguistic features and were usually included in the same manner as phonemes. Rossi [260] enlists the resulting principles that finally led Pierrehumbert [245] to publish her intonational model that, with slight improvements and additions, became a standard for all the current works in intonation. The principles include:

- i the requirement of treating intonational data as an autonomous level of analysis,
- ii the attempt "to deduce a system of phonological representation for intonation from observed features of F_0 contours",
- iii a relatively direct link between F_0 contours and the abstract phonological level, but an indirect link between the acoustic signal and functions,
- iv a method of discovery akin to the distributional model, and
- v the compositional conception of contour meaning similar to Harris's definition of meaning [128]

The work of Pierrehumbert is often referred to as the foundation of the Autosegmental-Metrical model [186]. The non-linear autosegmental principle, as proposed by Leben [195] and Goldsmith [113], stated that intonation is a level parallel to that of syllables, to which it is only loosely related and not determined by in terms of segmentation. That resembles the early suprasegmental assumptions by R. S. Wells [323], Trager and H. L. Smith [306] and others. The autosegmental theory, in turn, provided the necessary foundation for the Metrical Theory [205, 204] that used the fundamental frequency feature to express the relations between accents and phrases. Selkirk [269], in turn, combined it further with the increasingly popular generative phonology as it was the time when the chomskyan perspective on linguistics was rapidly gaining interest and provided yet another critical view of the traditional European structuralism [50]. But also the work of Pierrehumbert, with intonational level autonomy and phonetic analysis as a necessary part of modeling, contrasted the traditional programmatic assumptions of the European School. Not all of the European structuralists opposed these views, most notably the British School [60, 15] described intonation in terms of a single, unilinear representation, either as a set of holistic tunes or as linear successions of auditory categories, but also saw *phonetic* as referring to the auditory impression of a specific contour when analysed by a trained phonetician, and not to the acoustic realization as Pierrehumbert saw it. Her model is based on the direct analysis of the F_0 curves and has only two axiomatic assumptions:

- i the phonemes (H/L tones)
- ii the domain of selection of categories, that is the intonational phrase of the prosodic hierarchy proposed by Selkirk [269], and by Nespor and Vogel [251]

The resulting model does not only attempt to model intonation from bottom up, starting with the concrete F_0 contours and reaching abstract morphemes (the pitch accent) through an intermediate level of sequences of tones. She also assigns meaning to some of these morphemes. This reflects the earlier works by Bolinger [36, 35] and Trager and Smith [306, 282]. The intonational phrase is defined as:

$$x \geq 0 \text{ pitch accents} + 1 \text{ nuclear accent} + 1 \text{ phrase tone} + 1 \text{ boundary tone} \quad (2.4)$$

This work became very influential both within the United States [24, 136, 240, 250], and outside, where it provided the much needed rehabilitation of a field traditionally frowned upon by the practitioners of the European structuralism. As a consequence

a number of similar models were developed and include notable works by Mertens [220], Collier et al. [55], and Rossi et al. [261]. The differences are usually formal and lay mostly in the definition of the primitives inventories or the procedures of tone identification from the F_0 curves. Hirst and Di Cristo [138] created a model called MOMEL that includes an additional, third level of abstraction – a phonetic representation of the F_0 curves as a sequence of target points.

2.3.3 Research in Polish Intonation

Inquiry into the intonational system of the Polish language does not have a long history. First notable works include the handbook by Moneta [232] published in the 1720 and the following works by Nowaczyński [236], Golański [112], Elsner [88], Królikowski [181]. The purpose of prosodic study in that era was mainly to teach Polish rhetorics to Polish intelligentsia and to somehow "adjust" Polish for the needs of antique poetry and its rhythmic scheme, the dactylic hexameter. The first theoretical works should be however attributed to W. Mańczak [211, 212, 210] and M. Dłuska [85]. Both theories were formulated based on auditory judgments of their authors. There were notable differences between the key concepts and none of them specified what speech material was used in the analysis. Their work, however, helped reform the method of inquiry in the area. M. Steffen-Batogowa conducted the first systematic work on Polish intonational structure [288, 287]. A proper speech corpora was used in the study and the work included a detailed description of the methodology. It included both instrumental and perceptual analyses and the theory was based on the earlier work by Dłuska [84]. The study identified an inventory of 26 Polish intonemes. However, some key concepts used in that work, e.g. the phonological word, were still lacking acoustic definitions.

Demenko [70], on the other hand, based her work on the principles of the British tradition and the earlier works by Jassem [154, 156], which allowed to easily correlate some acoustic parameters with parts of the intonational phrase, defined as:

$$[\text{anacrusis}^*][\text{pre-nuclear intonation}^*[\text{nuclear intonation}]] \quad (2.5)$$

Here, anacrusis is a sequence of the lead-in unstressed syllables, the optional (marked with *) pre-nuclear intonation is composed of one or more pre-nuclear accents, and the ictal (nuclear) intonation consists of one and only one ictus – primary intonational accent and optional post-ictal accents (non-intonational).

Tab. 2.1.: Pre-ictal accent types

High	<i>H</i>
Low	<i>L</i>

Tab. 2.2.: Ictal accent types

Full falling	<i>HL</i>
Low falling	<i>ML</i>
High falling	<i>HM</i>
Extra low falling	<i>xL</i>
Low rising	<i>LM</i>
High rising	<i>MH</i>
Full rising	<i>LH</i>
Rising-falling	<i>LHL</i> or <i>MHL</i>
Even	<i>MM</i>

Demenko [70] listed the distinctive acoustic features of ictal (and pre-ictal) accents and identified that they are realized within the vowel. She also reported that vowels in nuclear syllables can show significant F_0 changes, as opposed to vowels in pre-ictal unaccented syllables. All of the possible pre-ictal and ictal accent types reported in that work are listed in Tables 2.1 and 2.2.

Examples of the 9 different Polish accents that were identified are illustrated in Figure 2.18. Top three examples of rising accents, followed by three examples of different falling accents in the middle can be characterized by the difference in the height of the final tone. According to Demenko [70], these accents can be realized both as a difference between the accented and post-accented syllables as well as an F_0 change on the accented syllable alone. The bottom row contains examples of level and rise-fall accents, which can be realized with an F_0 interval between the accented and post-accented syllables and a near-zero slope of fundamental frequency, or with a difference in duration between these syllables, and with a rise and fall realized on the accented vowel, respectively. Phrase boundary tones were identified as significant prosodic events realized through syllable duration, F_0 contours, intensity and pauses.

Additionally, the cited work describes some preliminary attempts at applying a number of physicalist models, including ANNs and the Fujisaki model [103], with the latter being reattempted several years later also by Demenko and Kuczmarski [74].

A number of studies concerned with modeling of Polish intonation were soon to follow. Francuzik et al. [99] and Jarmołowicz et al. [153] are two notable examples

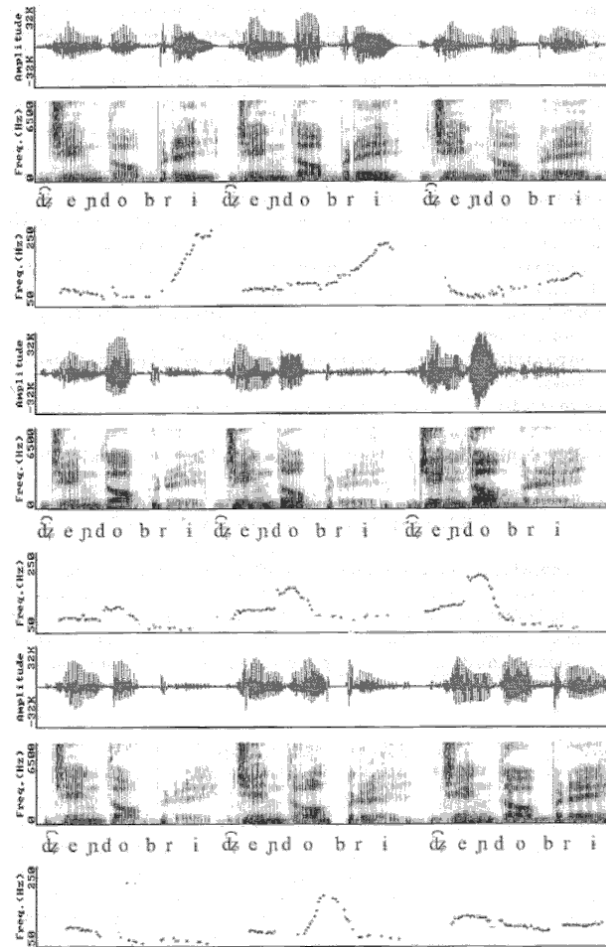


Fig. 2.18.: Acoustic realizations of the 9 different accents. Adopted from Demenko [70].

but a first attempt at formulating a full end-to-end bi-directional model of Polish intonation can be found in the work by Wagner [317], where the mappings between phonetic representations and surface phonological categories are founded on a specially developed F_0 contour stylization method and an inventory of pitch accents and boundary tones identified using statistical algorithms and a specially designed speech corpus. The model was specifically designed to be implemented in a speech synthesis system. The problem of intonation modeling in other domains of speech technology, i.e. Automatic Speech Recognition (ASR) and Computer-Aided Language Learning (CALL), was later addressed by Demenko et al. [75, 77]. The problem of intonation was also addressed in a number of works by Gonet, mainly from the perspective of glottodidactics and Polish-English comparative phonetics and phonology [114, 115, 116].

An extensive overview of the whole problem domain of Polish intonation processing in speech technology can be found in works by Demenko [71] and Wagner et al. [318].

Methodology

” *Why repeat the old errors, if there are so many new errors to commit?*

— Bertrand Russell

3.1 Aim of the Current Work

The main aim of the current work was to develop a model of how the continuous F_0 contours of an utterance emerge from the shallow discrete linguistic features of that utterance through a series of successive probabilistic mappings into intermediate latent representations, and to evaluate this model’s performance operationalized through the naturalness of its output in the context of a statistical parametric speech synthesizer. Although speech synthesis and speech technology in general are usually associated with practical applications, they also make up an important scientific tool that allows to study the language from many different perspectives. The model built within this work is aimed at providing a valuable contribution not only to the practical side of speech synthesis, but also to its scientific role in modeling speech production by including explanation methods and applying them to identify which of the many linguistic features used in this work demonstrate highest relevance for the synthesis of F_0 contours.

3.2 Motivation

3.2.1 Unificationist Approach to Intonation Modeling

More and more methods from other fields of science are adopted in linguistics to account for the interdisciplinary requirements of the modern research topics. The more theory is provided by the foundational fields of linguistics the more comprehensive models we build and the further we step beyond the traditional

frontiers of the discipline to accommodate for various processes that collectively constitute the act of language communication. Prosody research is a prominent example here, as its subject can be placed somewhere at the edge of language inquiry and does not easily obey the traditional frameworks of linguistics which are mostly designed to deal with grammars of easily separable segments.

Many of the meaningful components of intonation are expressed in terms of gradient instead of discrete categories. Although continuity can be easily handled with the tools of phonetics, it has always been problematic for phonology and has been traditionally assigned to the non- or para-linguistic level of speech communication. A number of phonological models have proven great for separating these components for the needs of traditional linguistic systems. As Wagner [317] points out:

“Phonetic models are regarded as quantitative. In phonetic models intonational features are described in terms of vectors of acoustic features or continuous parameters (e.g. duration, amplitude, slope, F_0 peak position) which interact with one another. In F_0 contour generation the values of the parameters are estimated from symbolic input by a regression model (e.g. Hunt and A. W. Black [149], Dusterhoff and A. W. Black [86], Mixdorff [227]). Depending on whether the model is sequential or superpositional the F_0 contour of an utterance results from interpolation between the estimated pitch targets (e.g. Momel [139], PaIntE [231], Tilt [297]) or superposition of the components of different temporal scopes (e.g. Fujisaki model [103] and its adaptations to different languages).”

Whereas:

“Phonological models are qualitative and sequential. In phonological models intonational tunes are considered as sequences of distinctive discrete tonal categories. As a result of detailed acoustic analyses an inventory of tonal categories and intonational grammar are defined which provide framework for transcription of intonation. As opposed to phonetic models which account for melodic aspects of intonation in the first place, phonological models represent the analytical approach: in the first place they account for functional aspects of intonation which are related to higher-level linguistic information. In F_0 contour generation the alignment and scaling of tonal targets is determined from rules devised by a human expert (Anderson et al. [9], Jilka [159], Jilka et al. [160]).”

However, “the whole variety of F_0 values available in the acoustics [...] [is reduced] to a mere binary opposition Low vs. High, and to some few additional, diacritic distinctions” [230, 23]. Moreover, more and more evidence, coming mainly from the neurobiological studies (as discussed in section 2.2.2), show that intonation

is processed by different parts of the human brain simultaneously, and that it is subjected to both continuous and categorical perception and undergoes both linguistic and non-linguistic processing at the same time. Many various multi-levelled heterogeneous cues combine to make up the final contour (impression) of the fundamental frequency. As an example, consider the biomechanics of the glottis and how it can impose natural non-cognitive limitations on the grammar of intonation units. It is not hard to imagine how a temporary physical state of the body, such as some violent exertion, for example, can affect the suprasegmental features of one's speech. Or how the brain's biophysical features affect the emergence of the cognitive constructs of tone and their perception. After all "We cannot think just anything — only what our embodied brains permit" [161]. Examples like these prove the dualistic perspective elusive at best. All these individual aspects can be addressed by their respective fields of science separately. Each of them can, and indeed does, provide important points of view of intonation on their own, but prosody research should benefit from methods that allow to study the phenomenon as a whole, and to allow patterns where two or more traditionally separate levels of language communication interact.

In the current work, the modeling framework is built under the aristotelian assumption that "The whole is greater than the sum of its parts" and that the synergy of these parts needs to be included in a model of intonation in order to allow scientific access to the very nature of the phenomenon. Leaning towards the lakoffian concept of embodied language, the author agrees that intonation cannot be understood without reference to the underlying "implementation details". The neurobiological part of these physical constraints of the body on mind can be expressed through models based on neural architectures that somehow resemble that of the human brain. Of course, even the most advanced artificial neural architectures are far from modeling even the simplest of brains. They can, however, still help us see general patterns. And with the accelerating advances in computing power and resources we will keep progressing towards more and more complete models. This of course converges with the problem of creating the Artificial General Intelligence. One of the prerequisites for AGI is the immersion in the environment as it is only through the embodied interaction with the environment that ideas and emotions can arise. Only with the knowledge of all external factors affecting the human body and mind can we attempt to model the complex internal states and their immediate expression through intonational features. This perspective, very appealing from the standpoint of the philosophy of language, is rather impractical in the current context as building such complex models is, of course, still virtually impossible.

Even a much more basic model that simply integrates all of the linguistic levels of intonation is still out of reach due to the numerous gaps in foundational research that still need filling, as well as the vast complexity of the integration task itself. However, each small step towards the unification of theories brings us closer to gaining a more complete perspective on the phenomenon.

3.2.2 Interfacing Phonetics and Phonology

The model built here concentrates mainly on addressing the aging gap between phonetic and phonological theories. This gap originates from the traditional dualistic perspective on language and linguistic methodology. As reviewed in the previous chapters, most of the current intonation models fall into this traditional dichotomous approach being either purely phonetic or phonological.

When addressed from the phonological perspective, intonation is often constrained to a discrete lexicon of cognitive units like tones or phrase breaks. This point of view necessarily ignores or oversimplifies a whole range of other aspects of intonation whose role and place in the eventual contour of the fundamental frequency is undeniable. Dealing with discrete minimal features, it neglects the gradient features completely, although they are as cognitively significant as their discrete counterparts. Theoretical phonological models are rarely validated through practical implementation despite the evident value of triangulation in theory building. This is probably mainly due to the fact that the physical correlate of the cognitive features and categories is very difficult to pinpoint and extract from the singular continuous flow of the fundamental frequency. Even if phonological research attempts to find such empirical evidence, it is more often a matter of creative interpretation that discounts all non-matching examples as anomalies related to the non-systematic nature of human language performance.

Phonetics, in turn, is focused on the quantitative study of these very real examples of intonation implementation in the speech signal and its fundamental frequency and the mechanics of their production. This approach, on the other hand, is exclusively descriptive and provides little explanation of the linguistic aspects of these physicalist measurements or how they relate to the meanings and ideas behind them. Phonetics basically ends where the dualistic mind begins. Although they both serve their purpose perfectly, they also both fail at explaining how these ideas are interpreted into sound, which is the core interest of linguistics with all of its subfields. However, being very problematic to approach with any of the currently available frameworks it is often treated as an unwanted child.

Since it is difficult to build a single cognitive or phonological system that translates to some easily separable characteristics or component of the fundamental frequency, and since many neurobiological studies provide evidence that the neural source of intonation is heterogenous, the author believes that intonation should be perceived not exclusively as a cognitive construct, a biological process or as its physical realization through fundamental frequency but as the relations between them, and that it is these relations that should be at the center of prosody research. This is also the main idea behind building the current model and developing methods that are able to reveal the complex structure of the possible underlying relations.

3.2.3 Explainable Deep Neural Network-Based Model of Intonation

A number of models, showing a tendency towards unification through the heterogeneity of considered features and emphasis on the mapping of phonological entities to phonetic realization, can be already found in speech technology. All machine learning-based models that learn to predict continuous sound output from discrete linguistic categories are in fact such models to some extent (see Section 2.1). The high quality of the output of such models, which are usually quite simple statistical predictors trained on large-scale domain-specific datasets, serves as a pragmatic evidence that models including a wide spectrum of features from many different levels and their sophisticated relations can successfully model the contours of the fundamental frequency.

However, until recently such models were of little value to science by itself as they did not provide any explanations and could not be used as tools to explore the phenomenon more profoundly than previously possible. Visual CNNs were the first to show outstanding potential for explainability. Popularized by the psychedelic images generated by Google's Deep Dream¹ the field of the so-called explainable AI is rapidly gaining momentum. Figure 3.1 shows an example of the effects of the Deep Dream algorithm, where a CNN was turned to enhance an input image, a photograph of the author in the Malaysian jungle, in such a way as to gradually elicit a particular interpretation (animals).

These recent developments prove that neural networks are not necessarily complete black boxes and that it is possible to explain their output through a variety of formal methods. Filter visualization and feature saliency in image classification Deep Convolutional Neural Networks are the most prominent example of how well

¹<https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html>

the predictions can be explained. As Deep Dream has demonstrated, it is possible to visualize how the input image is gradually filtered through a number of channels and how complex features emerge within the neural network, starting with primitives like edges, through more complex shapes and patterns, up to recognizable features of the target classes like ears, wheels and feathers and how they eventually affect the final output (see Section 2.2.4). This work presents efforts to extend this techniques for the needs of the current model as they should offer a new and valuable scientific view on how the emergence of the complex contours of intonation may take place somewhere between the discrete cognitive concepts, i.e. the linguistic input features, and their very physical continuous realization – the output values of fundamental frequency.

The above example of how CNNs process visual information closely resembles the neural processing of information in the human visual cortex (see Section 2.2.2). Simple and complex neurons in that area of the brain were shown to perform similar computations as the basic layers of the artificial network, i.e. linear filtering and maximal pooling. Eventually that information triggers activations in various areas of the brain responsible for recognition of patterns. Recent neurobiological studies show that this resemblance can also be found in the auditory cortex (see Section 2.2.2) which provides additional arguments for the use of these networks for the needs of the current problem. CNNs are becoming more and more popular as actual models of human visual and auditory processing in the field of computational neuroscience.

Therefore, a Deep Temporal Convolutional Network (TCN) was chosen as the model architecture for the current study. Although it is the Recurrent Neural Networks that became the *de facto* standard for modeling speech and language in the recent years, TCNs have already proven to outperform them in a variety of tasks, especially with the recent TTS systems based on TCNs demonstrating unprecedented quality of output [320, 274, 239, 14, 111, 248]. As already reviewed in Section 2.2.3, TCNs also demonstrate a number of purely technical advantages over their recurrent counterparts. They exhibit longer memory and provide a flexible receptive field size with the same capacity. Moreover, TCNs do not suffer from the vanishing gradient problem, allow for effective computational parallelism, have comparably low memory requirements for training, allow variable input lengths and provide numerous output explanation methods.

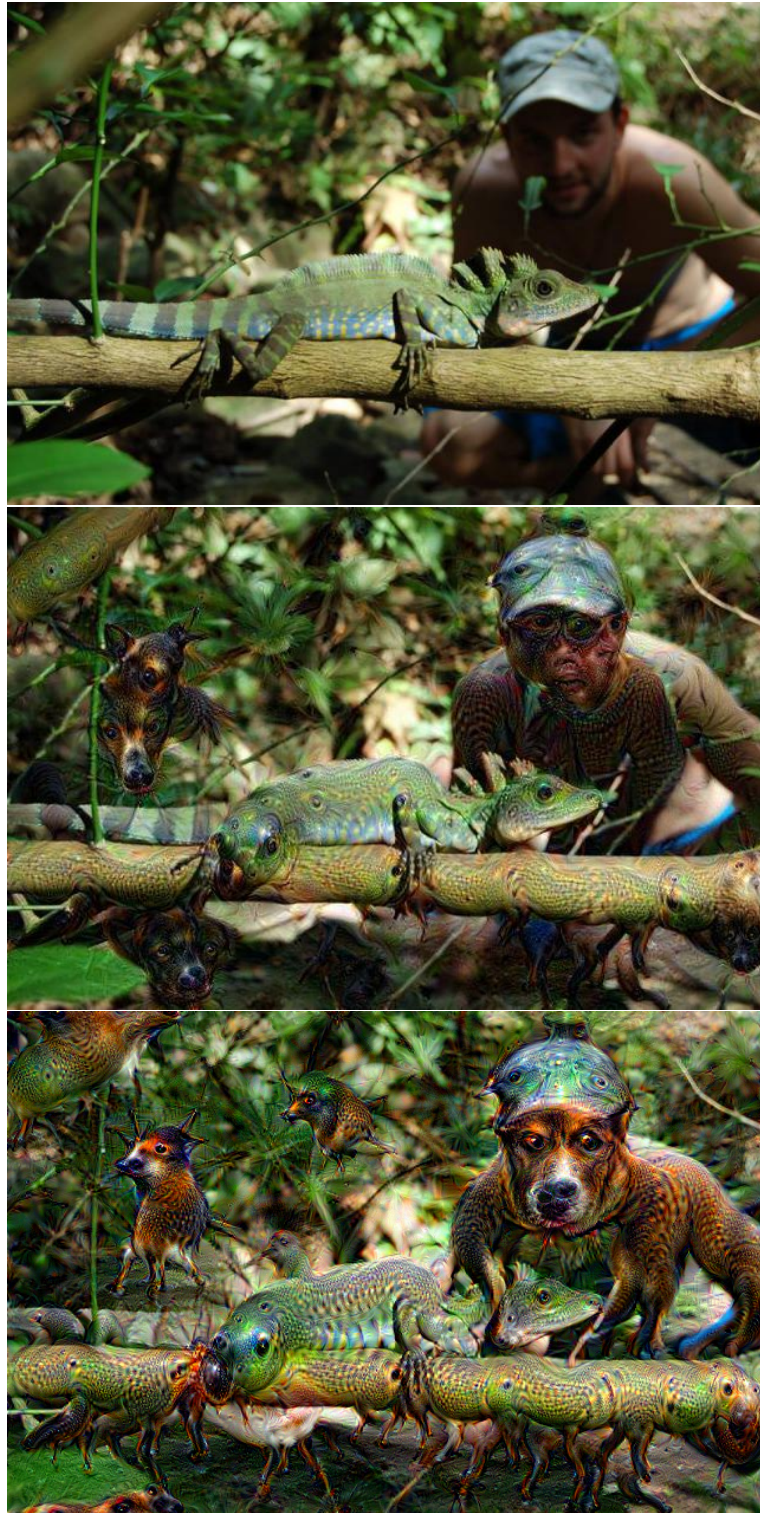


Fig. 3.1.: Google's Deep Dream example.

3.3 Deep Temporal Convolutional Neural Network as a Scientific Model

The motivation behind the choice of the TCNs for the current work should be evident by this point. However, adopting it for the purpose of scientific explanation is controversial, especially in linguistics. The application of DNNs for scientific purposes can be argued from many different perspectives, as listed and supported with multiple examples by Cichy and Kaiser [51]:

Notion 1 (Deductive-nomological [134]). *An explanation consists of an explanans and an explanandum. The explanandum must follow logically from the explanans (the deductive part).*

Notion 2 (Inductive-statistical [133]). *The explanandum follows probabilistically rather than logically from the explanans (the inductive part), and the explanans contains a statistical regularity (the statistical part).*

Notion 3 (Causal mechanistic [263]). *The explanation enumerates the causal processes and how they interact up and lead to the phenomenon.*

Notion 4 (Unificationist [100, 174]). *Explanation amounts to unify different phenomena in a common account, showing connections and relationships between phenomena whose relation was previously unclear.*

Notion 5 (Pragmatic [4, 314]). *Pragmatists stress that whether an explanation is successful, depends irreducibly on facts about the interests, goals, and beliefs of those providing or receiving explanation.*

Although the inherent nature of the current model is deeply probabilistic and purposefully does not easily fit within (1) which relies on the deterministic viewpoint and is the backbone of the classical phonological inquiry, it does present strong arguments from all other perspectives.

The emphasized outstanding results produced by DNN-based speech synthesis systems, provide uncontested pragmatic (5) arguments. These networks have been shown to generate human-like speech, learning even how to imitate additional non-speech sounds, like breathing, mouth movements and filled pauses. Moreover, some of them are even able to generate raw waveforms of the audio signal one sample at

a time and straight from an unprocessed textual input (see Section 2.2.3). The lack of a synthesizer frontend, where traditionally the text had to be first converted into a number of linguistic units and relations, removes the attachment to and bias of any phonological theory that the conversion rules were always necessarily based upon. Additionally, this provides further evidence of the capabilities of the network for making out very complex implicit patterns in the input data. These abilities surpass even that of human expert linguists. With these models defining the new state of the art in the practical domain of synthetic speech, the pragmatic argument (5) behind the adoption of TCNs for the current work was very strong.

Moreover, very few of the categories in language are actually of the well-defined type amenable to analysis in terms of necessary and sufficient conditions, and are more rather probabilistic fuzzy concepts. This is also supported by numerous neurobiological studies on speech as discussed in Section 2.2.2. Therefore, the models of language should also be implemented in frameworks that allow modeling of fuzzy probabilistic relations, as is the case with neural networks. From yet another point of view, TCNs rely on methods that allow generalization of some statistical trends found within some population based on a large representative number of samples of that population. The generalizations are encoded in the model parameters which are later employed in algorithmic predictions made for unseen data. This description suggests TCNs can be seen as statistical models and hence fit within the inductive-statistical notion (2). However, whether a particular model can be treated as a proper statistical model depends on its mathematical formulation and can be controversial in some cases. The currently accepted definition of a statistical model as given by Cox and Hinkley [58], Cox and Barndorff-Nielsen [57], Lehmann and Casella [199], Bernardo and A. F. M. Smith [27], and McCullagh et al. [218], states that:

“[...] A statistical model is a set of probability distributions on the sample space S . A parameterized statistical model is a parameter set together with a function $P : \theta \rightarrow P(S)$, which assigns to each parameter point $\theta \in \Theta$ a probability distribution P_θ on S .”

Now consider how the formulation of the TCN used in the current work as cited by the original WaveNet paper fits in that definition [239]:

“The joint probability of a waveform $\vec{x} = \{x_1, \dots, x_T\}$ is factorised as a product of conditional probabilities as follows:

$$p(\vec{x}) = \prod_{t=1}^T p(x_t | x_1, \dots, x_{t-1}) \quad (3.1)$$

Each audio sample x_t is therefore conditioned on the samples at all previous timesteps.

Similarly to PixelCNNs [312], the conditional probability distribution is modeled by a stack of convolutional layers. There are no pooling layers in the network, and the output of the model has the same time dimensionality as the input. The model outputs a categorical distribution over the next value x_t with a softmax layer and it is optimized to maximize the log-likelihood of the data *w.r.t.* the parameters. Because log-likelihoods are tractable, we tune hyper-parameters on a validation set and can easily measure if the model is overfitting or underfitting.”

Additionally, a special method was developed to allow for meaningful introspection of the model parameters. For this purpose, a feature relevance calculation algorithm was applied. Feature relevance is the measure of how particular parts of the input, and the intermediate latent features on every layer of the network, influence the final prediction at every time step. This information is considered a viable explanation in many scenarios including domains like self-driving cars [170], finance [184] and medical imaging [337]. In this work, it makes it possible to produce rich information about the importance of the input linguistic dimensions of the sentence and their intermediate latent generalizations made by the network, over the output fundamental frequency values, and thus enables the current model to be used for meaningful inductive-statistical (2) exploration of the mappings between linguistic entities and the contours of the fundamental frequency which can not only help prove the main hypothesis (Hypothesis 1), but also provides a valueable methodological contributions (see Hypothesis 4).

The current neural architecture of the model was also based on strong neurobiological evidence of how intonation is processed in the human brain and should support the same mechanisms that allow gradual emergence of the intonation contour from discrete linguistic categories. In this way, it allows to provide a causal mechanistic (3) explanation of the phenomenon. Because the model attempts to provide means for explaining the nature of the mappings between the phonological and phonetic levels of intonation in a neural-based modeling framework, which was additionally placed within a neural source-filter speech synthesizer, which acts as a model of speech production, it also fulfills the unificationist notion (4; also see Section 3.2.1).

Thereby the three levels of success required for a theory to be a scientific model [49] are achieved, although by different means, in the current model, i.e.: observational adequacy, descriptive adequacy and explanatory adequacy.

3.4 Methods

3.4.1 Dataset

Even the state-of-the-art neural network architecture is only as good as the data it has been trained on. In order to build a robust general predictive model of intonation of any language one needs to cover all possible contexts in which some specific distinguishable intonation contours appear, or at least the majority of these, so that the model can effectively generalize that information and use it to infer intonational contours of unseen examples. In practice, this is of course impossible. Even the relatively narrow subproblem of expressive intonation modeling is still unsolved due to the vast range of possible emotional states which are resistant to partitioning and defining. In this work, through the choice of the dataset, the model is limited to a relatively simple problem of neutrally read sentences, where any significant emotional, discourse and social contexts, as well as any other factors that can impact intonational contours, was significantly minimized.

The dataset used in the current work came into existence as a speech corpus built primarily for the purpose of the Polish BOSS unit selection synthesizer [72, 69, 73, 76]. The quality of this strictly data-based speech synthesis technology depends mainly on the phonetic coverage of the underlying database rather than on the number of examples included in it. A comprehensive set of acoustically relevant speech segments needs to be included in the inventory that the synthesis engine later draws from to construct new utterances. Although it is possible to contrive a single sentence that would cover all possible phonemes of that language (e.g.: *That quick beige fox jumped in the air over each thin dog. Look out, I shout, for he's foiled you again, creating chaos.* contains all English phonemes), building a synthesizer based exclusively on such an example would yield poor results at best due to the underrepresentation of the phonetic contexts a speech segment might appear in a language and that would influence its articulatory properties and hence acoustic realization. The bigger the speech corpus, and the better the resulting coverage of phonetic contexts, the better the quality of the resulting synthetic speech. However, unit selection databases traditionally attempt to cover the highest number of such

Tab. 3.1.: Stress and accent labels used in the original Polish BOSS speech corpus

Stress and accent type labels	
%	rising accent realized by F_0 rise on post accented syllable/syllables or F_0 interval between accented and post accented vowels
,	rising accent realized by F_0 change (rise on accented syllable)
"	falling accent realized by F_0 fall on post accented syllable/syllables or F_0 interval between accented and post accented vowels
&	falling accent realized by F_0 change (fall on accented syllable)
	rising-falling accents with rise-fall shape of F_0 movement on accented vowel
*	level accent realized by F_0 interval between preaccented and accented vowels; near zero slope of fundamental frequency
<	level accent realized only by differences in duration between preaccented, accented and postraccented vowels

contexts with the lowest possible number of recorded utterances minimizing the physical size of the database. The database becomes an integral part of the resulting synthesis engine and in most cases its large size is the single most significant factor limiting the implementational potential of the whole synthesis system.

To attain the best possible results with the least amount of data, the current corpus is following the ECESS guidelines for building speech corpora for neutral speech [72, 310, 87]. As recommended, the recording prompts include not only specially selected and frequent phrases, but also triphone coverage sentences to cover most contextual variations of segments in Polish, and some short excerpts from literary novels. To specially address prosody, which has always been one of the biggest challenges in unit selection synthesis, the design principles additionally put significant focus on the prosodic coverage including a specially planned set of distinctive intonation events discussed in more detail in the original work by Demenko et al. [72].

On this basis, 4 hours of speech read by a professional radio speaker in studio conditions and supervised by an expert phonetician were recorded. After an initial quality assessment, some 115 minutes of recordings were eventually hand selected to comprise the final database used in the current work.

The whole corpus can be divided into 6 parts, each addressing some different aspect, i.e.:

- (A) Phrases with 367 most frequent consonant structures addressing the consonantal nature of Polish,

Tab. 3.2.: Prosodic phrase boundary labels used in the original Polish BOSS speech corpus

Stress and accent type labels	
-5, .	Intonation on the first word in a sentence with falling accent F (or level accent L). In most cases it is used for declarative sentences or wh-questions. Mark on the first phoneme of the first word in the sentence
-5, ?	Intonation on the first word in a sentence with rising accent R. It can be used in different complex sentences. Mark on the first phoneme of the first word in the sentence
5, .	Intonation on the last word in sentence with falling accent F (or level accent L). In most cases it is used for declarative sentences or wh-questions. Mark on the first phoneme of the last word in the sentence
5, ?	Intonation on the last word in a sentence with rising accent R. In most cases it is used for yes-no questions. Mark on the first phoneme of the last word in the sentence
5, !	Intonation on the last word in a sentence with falling accent F. In most cases it is used for exclamatory sentences. Mark on the first phoneme of the last word in the sentence.
2, ?	Intonation on the last word in the phrase with rising accent R. In most cases it is used for continuation phrases. Mark on the first phoneme of the last word in the phrase.
2, .	Intonation on the last word in the phrase with falling accent F (or level accent L). In most cases it is used in declarative phrases in complex sentences. Mark on the first phoneme of the last word in the sentence.

- (B) All Polish diphones realized in 114 grammatically correct but semantically nonsense phrases.
- (C) 676 phrases with CVC triphones realized in non-sonorant voiced context and in various intonation patterns.
- (D) 1923 6-14 syllable-long phrases with CVC triphones realized in sonorant context and various intonation patterns.
- (E) 2320 utterances with 6000 most frequent Polish vocabulary items.
- (T) 15 minutes of prose and newspaper articles.

The database was segmented and labelled using the BOSS *blf* format [175, 40]. Labels contain information about segment, syllable, intonational word and phrase boundaries as well as prosodic labels for different accents and phrase boundary types, based on the British models of intonation [60, 15, 245, 157, 155, 154, 156]. The stress and accent labels used in the database are presented in Table 3.1 and the phrase boundary type labels are listed in Table 3.2. The phoneme inventory was based upon the modified Polish SAMPA alphabet [158, 78, 90, 188, 321, 322]. Prosodic word and syllable boundaries are marked with # (or #_ for orthographic boundaries) and . . . Additionally, \$p is used to denote a pause. A special symbol \$j was also used to mark unintelligible segments. An example BLF label file for the sentence "Smutek malował się na jej twarzy" (*Sadness was showing on her face*) is presented in Listing 3.1.

```

1  0 # $p
2  3360 #s -5, .
3  5502 m
4  6331 |u
5  7892 .t
6  8772 e
7  9769 k
8  11200 #m
9  12061 a
10 13021 .l
11 13715 *o
12 14720 .v
13 15459 a
14 16211 w
15 17119 #s' 2, .
16 18909 e
17 19680 #n
18 20724 a
19 21684 #j

```

```

20 22436 e
21 23040 j
22 24000 #t 5, .
23 25600 f
24 26573 *a
25 28713 .Z
26 30080 y
27 32811 #p

```

Listing 3.1: An example BLF label file for the sentence "Smutek malował się na jej twarzy" (*Sadness was showing on her face*) taken from the original Polish BOSS corpus.

3.4.2 Feature Extraction

In order to serve as an input for training the neural network the dataset labels had to be first normalized to an acceptable format. Using a set of tools developed by the author as part of one of the previous works [182], the labels were automatically converted into a notation that explicitly reflects the structural relations between the various levels of speech segments and prosodic events.

For this purpose, separate labels were first extracted for phones, syllables, words, phrases and intonational events and saved in the Festival label format [31, 297, 32] as shown in Table 3.3. The original segmental labels needed to be converted in some cases as they contained special characters which are used internally by the HTK toolkit [331, 330] which is used in this research extensively.

The intonational labels were converted into a different system based on ToBI [278] as shown in Table 3.4. Although the corresponding ToBI labels might be controversial in some cases, their choice is actually insignificant for the training process. What matters the most is labelling consistency. Even if the actual ToBI mark for a given class of intonational events is incorrect, the neural network does not assign any meaning to the label itself as it only serves as a category distinction and could as well be represented with ordinal numbers or letters.

Using a set of Festvox synthetic voice building scripts [12] and some HMM-based Speech Synthesis System (HTS) tools [335], the labels were then converted into the HTK full-context label format [330] which allows for easy data normalization and feature extraction. The HTK full context label file is an explicit declarative transcript of all the structural information in the utterance (see Table 3.5).

Each row represents a phone in its quintphone context followed by a series of special expressions that encode all positional, quantitative and qualitative features that

Tab. 3.3.: An example of the intermediate Festvox-based label structure for "Smutek malował się na jej twarzy" (*Sadness was showing on her face*).

End time	Segment	Syllable Label	Stress	Word	Phrase	Int. Event
0.21	pau	pau	0	pau		
0.343875	s					
0.3956875	m					
0.49325	u	s.m.u	1			
0.54825	t					
0.6105625	e					
0.7	k	t.e.k	0	smutek	1	
0.7538125	m					
0.8138125	a	m.a	0			
0.8571875	l					
0.92	o	l.o	1			
0.9661875	v					
1.0131875	a					
1.0699375	w	v.a.w	0	malowaw	1	
1.1818125	si					
1.23	e	si.e	1	sie	3	L-
1.29525	n					
1.35525	a	n.a	1	na	1	
1.40225	j					
1.44	e					
1.5	j	j.e.j	1	jej	1	
1.6	t					
1.6608125	f					
1.7945625	a	t.f.a	0			
1.88	rz					
2.0506875	y	rz.y	1	tfarzy	4	L-L%
2.2506875	pau	pau	0	pau		

Tab. 3.4.: Tone accent and phrase boundary labels conversion scheme.

Original Polish BOSS label	ToBI
"	H*
&	L*
%	L*+H
‘	L+H*
*	-
<	-
5, . or 5. ! and last accent is not ‘ and not &	L-L%
5, . or 5. ! and last accent is ‘ or &	H-L%
5, ? and last accent is not ‘ and not &	L-H%
5, ? And last accent is ‘ or &	H-H%
2, .	L-
2, ?	H-

were extracted during the conversion. This record type is understood by HTK HED [330] *questions*, which are special regular expression-like matching masks that, when applied to a record in an HTK label, return a boolean value. For example, the following row outputs `True` if the immediate left context of a given segment is a voiced fricative:

```
QS "L-Voiced_Fricative" {*^v-,*^z-,*^zi-,*^rz-}
```

This question, in turn, checks if the current syllable is not farther than on the 13th position from the beginning:

```
QS "Pos_C-Syl_in_C-Phrase(Fw)<=13" {*&?-*,*&10-*,*&11-*,*&12-*,*&13-*}
```

Another example checks if the current syllable is 3 syllables away from the previous accented syllable:

```
QS "Num-Syl_from_prev-AccentedSyl==3" {*;3-*}
```

A list of such questions can be input to the HTK HED script and allows to convert the HTK labels into one-hot encoded feature vectors whose length is equal to the length of the question list. An overview of the types of questions that were formulated for the current work are listed in Table 3.6 and 3.7.

Table 3.6 contains the segmental part of the questions. They were added to account for their impact on microprosody and for any possible articulatory constraints that segmental articulation contexts might have on the production of intonation. These

Tab. 3.5.: An example fragment of the HTK label file for the sentence "Smutek malował się na jej twarzy".

Segment start	Segment end	Full-context label
...	...	...
2100000	3438750	xx^pau-s+m=u@1_3 A:xx_xx_xx B:1-0-3@1-2&2-2#0-0\$0-0!xx-3;xx-5 u C:0+0+3 D:content_1 E:content+2@2+1&1+0#1+xx F:content_3 G:xx_xx H:3=2^1=6 NONE I:4=2 J:12+8-6
3438750	3956870	pau^s-m+u=t@2_2 A:xx_xx_xx B:1-0-3@1-2&2-2#0-0\$0-0!xx-3;xx-5 u C:0+0+3 D:content_1 E:content+2@2+1&1+0#1+xx F:content_3 G:xx_xx H:3=2^1=6 NONE I:4=2 J:12+8-6
3956870	4932500	s^m-u+t=e@3_1 A:xx_xx_xx B:1-0-3@1-2&2-2#0-0\$0-0!xx-3;xx-5 u C:0+0+3 D:content_1 E:content+2@2+1&1+0#1+xx F:content_3 G:xx_xx H:3=2^1=6 NONE I:4=2 J:12+8-6
...	...	...

are mostly phonological and phonetic categories and features of a given segment and its 4 nearest neighbours - the *quintphone*.

Tab. 3.6.: List of segmental HTK HED question types used for feature extraction

Question type	Segments
Vowel	{i, y, e, a, o, u, schwa}
Consonant	{gs, p, b, t, d, k, g, ki, gi, f, v, s, si, z, zi, sz, rz, x, c, dz, cz, drz, ci, dzi, m, n, ni, ng, l, r, w, ww, j, jj}
Stop	{gs, p, b, t, d, k, g}
Nasal	{ww, jj, m, n, ni, ng}
Fricative	{f, v, s, si, z, zi, sz, rz, x}
Front	{e, i, y, f, v, p, b, m, w, ww}
Central	{schwa, a, t, d, s, si, z, zi, n, r, l, t, d, sz, rz, cz, drz, c, dz, ci, dzi}
Back	{o, u, k, g, ki, gi, ng, x, gs}
Front Vowel	{e, i, y}
Central Vowel	{a, schwa}
Back Vowel	{o, u}
High Vowel	{i, y, u}
Medium Vowel	{e, o}
Low Vowel	{a}
Rounded Vowel	{o, u}
Unrounded Vowel	{a, e, i, y}
XVowel (e.g. AVowel)	{i, y, e, a, o, u, schwa}
Unvoiced Consonant	{gs, p, t, k, ki, f, v, s, sz, x, c, cz, ci}
Voiced Consonant	{b, d, g, gi, v, z, zi, rz, dz, drz, dzi, m, n, ni, ng, l, r, w, ww, j, jj}
Front Consonant	{f, v, f, p, b, m, w, ww}
Central Consonant	{t, d, s, si, z, zi, n, r, l, t, d, sz, rz, cz, drz, c, dz, ci, dzi}
Back Consonant	{gs, k, g, ki, gi, ng, x}
Fortis Consonant	{gs, cz, f, k, p, s, sz, t, ci, c, ki}

Lenis Consonant	{drz, v, g, b, rz, z, d, dzi, dz, gi, zi}
Neither F or L	{m, n, ni, ng, l, r, w, ww, j, jj}
Voiced Stop	{b, d, g}
Unvoiced Stop	{p, t, k, gs}
Front Stop	{b, p}
Central Stop	{d, t}
Back Stop	{g, k, gs}
Voiced Fricative	{v, z, zi, rz}
Unvoiced Fricative	{f, s, si, sz, x}
Front Fricative	{f, v}
Affricate Consonant	{dz, drz, dzi, c, cz, ci}
silences	{pau}
Phone X identity (e.g. a)	{a}

The non segmental features, in turn, are mostly related to the relative position and length of segments at various levels. The types of non segmental features are listed in Table (3.7).

The resulting feature set covers phonetic and articulatory qualities of the central phone and its 4 most immediate neighbours, accent placement and type as well as information about the objective and relative position of a given segment, syllable, word and phrase in their enclosing units of various levels and in relation to its beginning or end or some other intonationally prominent event or unit within the utterance. That yields a total of 1296 binary features for each segment. For the full list of features see Appendix A.

Next, $\log F_0$ values were extracted from all recordings using the the SPTK toolkit [283] implementation of the Robust Algorithm for Pitch Tracking (RAPT) [293]. The F_0 detection range was limited to 55-200 Hz based on a brief inspection of the speaker's voice range in order to prevent pitch doubling and pitch halving errors as much as possible. The voice is naturally low but is also characterised by abrupt inclines phrase-finally, especially in questions. Pitch extraction was conducted using a 240-point frameshift and a 0.005s frame length with a 48Hz sampling frequency of the recordings. The extracted information was converted to the log scale and saved as the target label for training the neural network. Given the resulting $\log F_0$ values for voiced sections and null elsewhere, a single additional binary voiced/unvoiced indicator was added to the feature set.

Tab. 3.7.: List of non-segmental HTK HED question types used for feature extraction.

Number of preceding/succeeding segments in the previous/current/next syllable is equal to/less than or equal to 0-7
Previous/current/next syllable is stressed
Previous/current/next syllable is accented
Previous/current/next syllable has accent X (where X is one of the ToBI accents described above)
Number of preceding/succeeding segments in the next syllable is equal to/less than or equal to 0-7
Forward/backward position of the current syllable in current word is equal to/less than or equal to 0-7
Forward/backward position of the current syllable in current phrase is equal to/less than or equal to 0-20
Number of stressed syllables before/after the current syllable in current phrase is equal to/less than or equal to 0-12
Number of accented syllables before/after the current syllable in current phrase is equal to/less than or equal to 0-12
Number of accented syllables before/after the current syllable in current phrase is equal to/less than or equal to 0-7
Number of syllables from previous/next stressed syllable is equal to/less than or equal to 0-5
Number of syllables from previous/next accented syllable is equal to/less than or equal to 0-16
Current syllable nucleus is a non-vowel, vowel, front vowel, central vowel, back vowel, high vowel, medium vowel, low vowel, rounded vowel, unrounded vowel, [i], [e], [a], [o], [u], [y], [schwa]
Number of syllables in the previous/current/next word is equal to/less than or equal to 0-7
Forward/backward position of the current word in the current phrase is equal to/less than or equal to 0-13
Number of content words before/after the current word in the current phrase is equal to/less than or equal to 0-9
Number of words from previous/next content word is equal to/less than or equal to 0-5
Number of syllables in the previous/current/next phrase is equal to/less than or equal to 0-20
Number of words in the previous/current/next phrase is equal to/less than or equal to 0-15
Forward/backward position of the current phrase in the utterance is equal to/less than or equal to 0-4
Number of syllables in the utterance is equal to/less than or equal to 0-28
Number of words in the utterance is equal to less than or equal to 0-13
Number of phrases in the utterance is equal to/less than or equal to 0-4

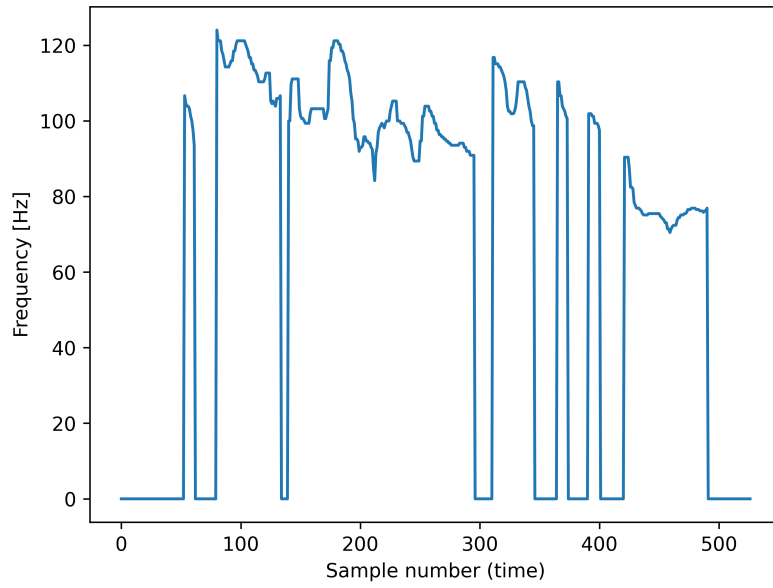


Fig. 3.2.: Extracted F_0 values for "Słyszałam odgłos zbliżającego się pociągu" (*I heard the sound of an approaching train*).

In deep neural network research, the data is often customarily normalized to a $[0, 1]$ or a $[-1, 1]$ range. However, the extracted features are one-hot-encoded so they are naturally expressed in the $[0, 1]$ range and need no further normalization. Target normalization was not performed as initial experiments showed no particular improvements.

Phone feature vectors were naively reproduced for each time point within the time boundaries of that phoneme to match the length of the target vector. As TCNs cannot deal with variable length input, all of the data samples were extended to match the maximum length within the database using zero-padding to account for the missing values.

The resulting 2-dimensional vectors whose length and width represent the number of samples in the data and number of features respectively were served to the training algorithm in batches of 64 resulting in a 3-dimensional input vector, i.e. $\text{batch size} \times \text{number of features} \times \text{number of samples}$.

Although a number of extra features were also engineered as part of the current work, they were eventually not used but are still available in the code repository² and can be used in future work. These include phoneme, syllable, word and phrase forward and backward durations, mel-generalized cepstral coefficients and $\log F_0$ backward window features that represent n -previous $\log F_0$ values at each sample. The duration features, F_0 backward window and MGCC [302] were abandoned as the author suspects that the information they convey can be implicitly modeled by the network. Although all of these feature classes are quite appealing in their very own way, the single most important reason for abandoning them were the computational memory and time constraints of the current work. The current feature vector size is $64 \times 1900 \times 1297$ of 8-byte Python booleans, which results in 1.2617216 gigabytes of memory required for just a single input vector. The consequential memory requirements of training a deep network with such input are much higher. Additionally, as described later in more detail, the currently implemented training process makes extensive use of parallel processing. This results in even higher memory requirements as each parallel thread operates on their own input vector significantly extending memory consumption.

3.4.3 Model Implementation

The current model as well as the data loading, training, testing and evaluation were implemented using the Python [315] API for Keras [47, 123] and Google TensorFlow [1], which are free and open-source software libraries for differentiable programming and neural networks. Additionally, since these libraries provide no out-of-the-box support for TCNs, an external general implementation [255] of Temporal Convolutional Networks [194, 20, 19] was employed for this purpose. This implementation is based on the recent convolutional architectures for sequential data [239, 164, 108], but is also distinct from all of them and, as the authors point out, it was designed from first principles to combine simplicity, autoregressive prediction, and very long memory. It differs from the original WaveNet architecture as it introduces no skip connections across layers, no conditioning, and no context stacking, or gated activations. It has, however, significantly longer memory.

With the help of the above libraries a model was built combining classical densely connected layers and a special TCN block. The TCN block itself is, in turn, composed of smaller blocks of 6 residual layers each. Figure 3.3 illustrates the detailed architecture of the residual block used in the current work. The residual blocks

²https://github.com/mrslacklines/intonation_synthesis

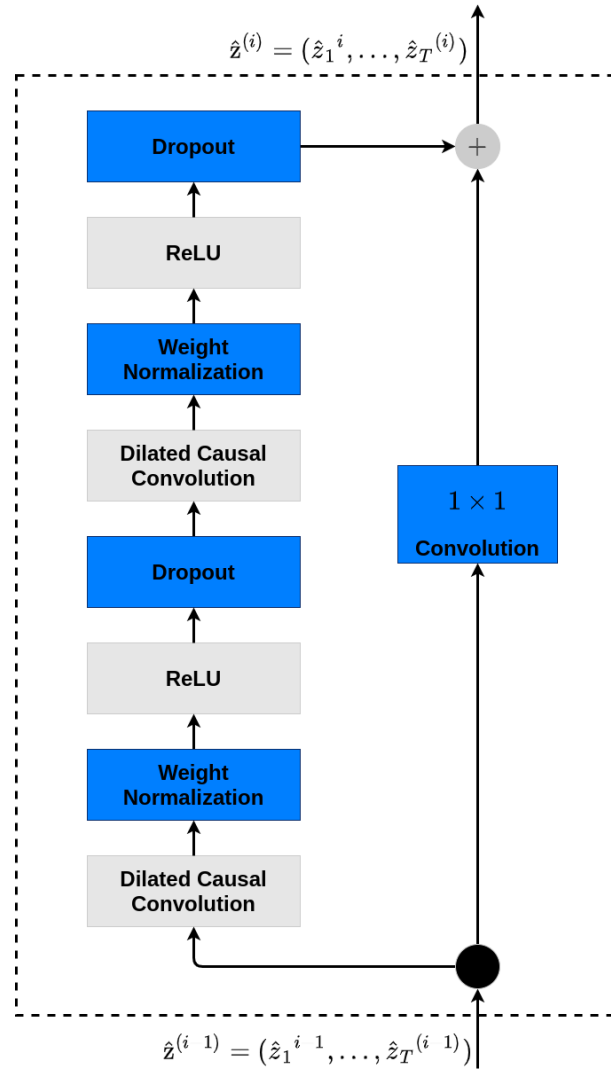


Fig. 3.3.: Architecture of the Residual Block used in the current network architecture.

are additionally stacked on top of each other to achieve a large receptive field with just a few layers, while preserving the input resolution throughout the network as well as computational efficiency (as cited in the original WaveNet paper [239]). Each residual block contains a chain of a dilated causal convolutional layer, a batch normalization layer, a Rectified Linear unit activation layer and a dropout layer, all repeated twice in that order. The convolution layers contain 64 convolution filters of length 2 and apply the standard 2^n dilations setup known from the original WaveNet implementation (i.e. $[1, 2, 4, 8, 16, 32]$).

The data output by the convolutional block undergoes batch normalization in order to further improve the speed, performance, and stability of the network. After that,

the normalized data is passed through the Rectified Linear Unit (ReLU) [235, 285] which was chosen as the main activation function for the network. Finally, in order to prevent overfitting, 0.2 of the values output by the previous operations is removed in the dropout block [285]. This whole sequence is repeated twice in each residual block. The data is also passed in parallel through a simple identity convolution block whose output makes up the actual residue in the output data. The results of both computational paths are summarized and in that form comprise the final output of the whole block.

Additionally, a single 1×1 convolutional layer was prepended to the whole pipeline as an optimization technique that reduces the dimensionality of the input vector to 64×1 before the costly convolutions take place.

Two fully connected dense layers with ReLU activation functions and sizes 64 and 1 were used for the final regression and dimensionality reduction of the TCN output. This setup closely mimics the aforementioned arrangements and functions of the simple and complex cells in the human brain (see Section 2.2.2). The detailed implementation of the network architecture is available as in the public project repository at https://github.com/mrslacklines/intonation_synthesis.

3.4.4 Model Training

The code was packaged in a Docker [219] container and deployed to the Amazon Web Services (AWS) Sagemaker [163] computational cloud service. The training data was deployed onto AWS S3 servers [217]. This setup allowed the training experiments to run on 64 CPUs and 8 NVIDIA Tesla V100 GPUs with 488 and 128 gigabytes of memory. At the time of implementation of this work, NVIDIA Tesla V100 Tensor Core was the most advanced data center GPU ever built to accelerate AI [213, 214]. Such a powerful hardware setup was required not only to account for the complexity of the network itself but also for the large dimensions of the input vector. A single batch of data, which is a $64 \times 1900 \times 1297$ vector of 8-byte boolean values, occupies 1.2617216 gigabytes of memory, and the model comprises of a total of 449,409 parameters (446,337 trainable and 3,072 non-trainable).

An average training session took between 2 and 10 hours. Although a multi-GPU model was implemented and is available in the attached repository it showed no significant speed improvements since it was the data loading and preprocessing that became the bottleneck of the whole process. In order to speed up the underlying input/output operations they were implemented to run in parallel on multiple CPUs

where each worker was reading and preprocessing a single file from a batch. This improved the overall performance greatly but not to an extent that would allow the benefits of a multi-GPU training as it was still the I/O operations that were the slowest.

The network was first randomly initialized using the He initializer technique [131] which draws samples from a truncated normal distribution centered on 0 with

$$\sigma = \sqrt{\frac{2}{\text{Fan-in}}}$$

where Fan-in is the number of input units in the weight vector.

Adam [173] optimization algorithm was used to train the network instead of the classical Stochastic Gradient Descent procedure (SGD) [39]. Adam has become the *de facto* standard in training Deep Neural Networks as it is computationally efficient and has little memory requirements. It is also invariant to diagonal rescale of the gradients and well suited for problems that are large in terms of data or parameters. Additionally, the authors of the method show that it is appropriate for non-stationary objectives and for problems with very noisy or sparse gradients. Another advantage is that the hyper-parameters have intuitive interpretation and typically require little tuning. The method computes individual adaptive learning rates for different parameters from estimates of first and second moments of the gradients instead of keeping a fixed value throughout the whole training, as was the case with SGD. Specifically, the algorithm calculates an exponential moving average of the gradient and the squared gradient. The parameters β_1 and β_2 control the decay rates of these moving averages.

In the current work, an initial learning rate of 0.1 and $\beta_1 = 0.9$, $\beta_2 = 0.999$ were used. The ε parameter was set to $1e - 07$. This is the small constant added to all values for numerical stability as it prevents the division by zero. Mean Squared Error (MSE) was used as the loss function. This is the most popular method for evaluating the synthetic F_0 values against the original values and although the initial plan was to experiment with other loss functions that might be able to better capture the dynamics of the contours, the primary choice has proven to perform well enough for the current needs and was therefore kept.

In each training instance the dataset was randomly divided into a training, testing and validation subsets using an 8:1:1 ratio. Network optimization was set to run for 200 epochs. During each epoch, the whole training data subset was fed to the optimization algorithm in batches of 64. Loss was calculated both for the training batch and for the unseen validation data subset with the aim to optimize the latter

and using the former only as an additional performance metric. The training was also set to stop automatically if after 20 consecutive epochs the validation loss did not improve by at least 0.001. The model parameters configuration that performed best on the validation set was kept. The parameters were saved in a HDF5 [96] file which is also published in the attached repository.

3.4.5 Inference

An additional Docker container was prepared to be run as a SageMaker job on the AWS cloud for the purpose of results generation and evaluation of the trained model. This inference job initialized a new instance of the model with the parameters serialized by the training process to a HDF5 file and iterated over the whole holdout set of 191 sentences generating prediction for each of them.

Given the predicted $\log F_0$ values and the ground truth, objective errors were calculated including the Mean Squared Error, Mean Absolute Error, Root Mean Squared Error and the Normalized Root Mean Squared Error.

The predictions themselves were saved to binary files after being trimmed to match the length of the ground truth array and get rid of the zero-padding. Each $\log F_0$ prediction was also converted to the frequency scale and plotted against the ground truth. As an extra post-processing step, mainly for the sake of the subjective evaluation of model performance (described in the next sections) the predicted values were time-aligned with the ground truth arrays and any non-zero values output by the F_0 model, appearing at intervals where no F_0 values were found in the ground truth arrays³, were deleted. In cases, where the model failed to predict a non-zero value where the ground truth array rendered non-zero values, the last value predicted by the model was forward-filled, except for the onset where it was back-filled. This procedure was expected to minimize any glitches appearing at voiced/unvoiced boundaries during resynthesis and help improve the overall quality of the synthesized output. The predictions were also plotted against the ground truth values and the plots were saved for further visual inspection.

³Actually, this were not the original ground truth F_0 values from the original dataset but the F_0 values extracted by the NSF tools, described in the next section.

3.4.6 Feature Relevance Analysis

In addition to simple inference, a feature relevance analysis was performed for each of the test sentences. In order to extract a representation of the relations between the input features and the output, the Layer-wise Relevance Propagation (LRP) algorithm was used. Although this technique has already become a standard in the domain of image recognition, only a few studies have applied it to time-series data. **To the best of the author’s knowledge no such attempt has been made to explain the relevance of linguistic features in the process of generating any aspect of speech.** For this purpose, an implementation of the LRP-Z (LRP Zero) method from the Innvestigate [6] Python library was employed. Innvestigate was chosen as it supports time-series data almost out-of-the-box through the `neuron_selection_mode` parameter which can be set to `all` and can hence generate relevance maps *w.r.t.* all output neurons as opposed to only the maximally activated one as is the case in most classification problems (and the competing Python library TSViz [277]). However, the original code had to be adjusted by the author in order to support Tensorflow 2.0 and eager execution mode for which some of the operations used in the original Innvestigate library are not yet supported⁴.

3.4.7 Neural Source-Filter Resynthesis

Subjective evaluation of the model predictions was performed alongside the objective evaluation. Although the corpus used in the current study is supposed to contain only neutral read speech it is impossible to ensure that each individual sentence is read by the speaker with a prototypical intonation contour. Each sentence can be read in a number of different ways varying not only by scale of how specific contours are realized, so for example how far does the speaker raise their voice when realising an `L*+H` tone on the accented syllable, but also by the choice of tonal accents used for that same accented syllable without change in meaning or expressiveness of the whole sentence. The actual interpretation of a sentence is a whole another area susceptible to different realizations. The neural model used in this study does not consider any information regarding these and although they should be conveyed in the actual annotations of the corpus it is also impossible to ensure that all labels are always correct as they are also a subject to perceptual analysis of the annotator.

⁴These changes were staged as a pull request to the original Innvestigate repository and the author was pleased to note that it has been used by some other users in the community. See <https://github.com/albermax/innvestigate/network/members>.

Therefore, only a subjective perceptual evaluation of the predicted contours can help gain a full insight into the actual performance of the model. However, the design of such a perceptual evaluation was also non-trivial. Although a number of well known methods for the evaluation of audio and, more specifically, speech exists, the main challenge was to design such a procedure that would enable an evaluation of the generated intonation by itself, without any influence of the quality of the carrier speech. Usually in such experiments speech is resynthesized using one of the well known algorithms PSOLATM [311], TD-PSOLATM [291], LPC or any of the more contemporary source-filter resynthesis methods like STRAIGHT [166]. However, most of these generate a number of auditive spectral artifacts that can easily serve as unwanted cues for discrimination of synthetic samples during testing. One method to handle this issue is to resynthesize both the natural benchmark speech samples and the evaluated synthetic ones to ensure similar artifacts are present in both subsets. This is also not ideal as it may lower the perceived quality of the actual benchmark and in an experiment that relies mostly on a comparative evaluation, this becomes a significant problem.

In order to mitigate the listed issues, the newest state of the art vocoding technique was employed. After the WaveNet synthesizer demonstrated the capabilities of neural vocoders this technique became very popular as the results it produces easily approach the naturalness of human speech [239]. However, as noted by Xin Wang [319]:

“As an autoregressive (AR) model, WaveNet is limited by a slow sequential waveform generation process. Some new models that use the inverse-autoregressive flow (IAF) can generate a whole waveform in a one-shot manner. However, these IAF-based models require sequential transformation during training, which severely slows down the training speed. Other models such as Parallel WaveNet [313] and ClariNet [247] bring together the benefits of AR and IAF-based models and train an IAF model by transferring the knowledge from a pre-trained AR teacher to an IAF student without any sequential transformation. However, both models require additional training criteria, and their implementation is prohibitively complicated.”

The cited paper introduces a new technique based on a much simpler assumption. The Neural Source-Filter (NSF) requires three components only; “a source module that generates a sine-based signal as excitation, a non-AR dilated-convolution-based filter module that transforms the excitation into a waveform, and a conditional module that pre-processes the acoustic features for the source and filter modules.” Despite its simplicity it was shown to match or even outperform WaveNet while generating the waveforms 100 times faster. An additional advantage of this technique

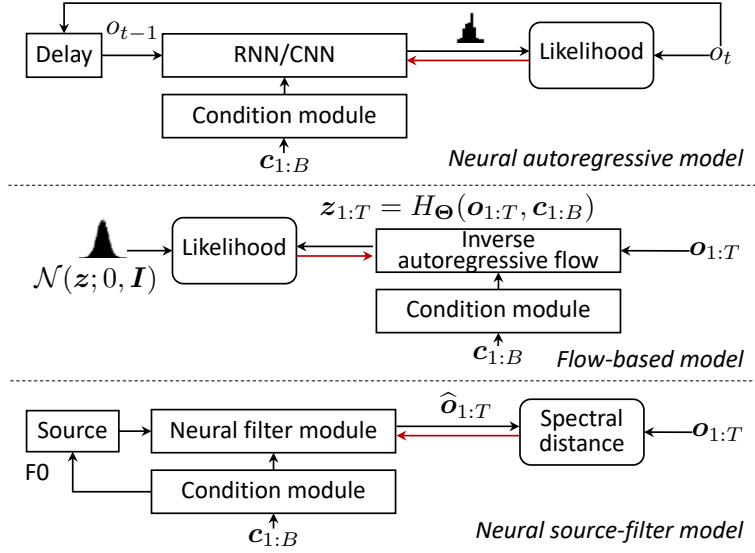


Fig. 3.4.: Three types of neural waveform models in training stage. Adopted from X. Wang et al. [319] (©[2019] IEEE).

is the fact that the component of this method can be easily divided into source and filter providing a better linguistic abstraction than some of the end-to-end neural models that model both aspects within an integrated framework.

Figure 3.4 provides a comparison of the training process of all three of the mentioned approaches to neural vocoders, the autoregressive model (e.g. WaveNet), the model based on the inverse autoregressive flow and the NSF as the last one. $\hat{o}_{1:T}$ and $o_{1:T}$ denote generated and natural waveforms, respectively. $c_{1:B}$ denotes input acoustic features. Red arrows denote gradients for back propagation.

What is characteristic of NSF is that during training it uses spectral-domain distances for loss calculation instead of the more traditional metrics like MSE or the likelihood. The NSF was implemented in three main variants, i.e.:

- b* with the network structure similar to ClariNet [247] and Parallel WaveNet [313]
- s* which is the b-NSF with simplified neural filter modules
- hn* an extension of the s-NSF implementing an explicit generation of the harmonic (periodic) and noise components.

The current study is using the latter as it seems best suited for the purpose of evaluation of an F_0 prediction model that disregards the unvoiced regions. The hn-NSF was also found to perform best in terms of modeling unvoiced segments as compared to its two other variants. The harmonic/noise module merges the two

components in different ratios for the voiced/unvoiced sections of the generated speech using two pairs of low- and high-pass filters.

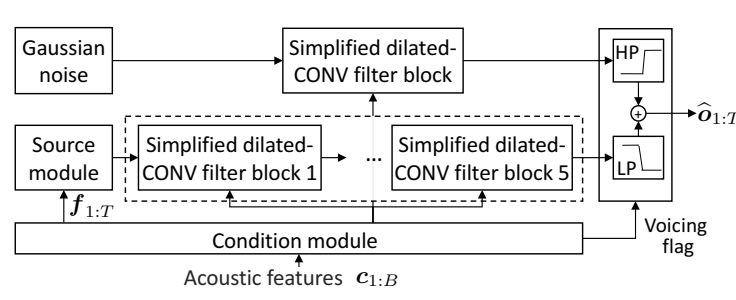


Fig. 3.5.: Diagram of harmonic-plus-noise NSF (hn-NSF) model. Adopted from [319] (©[2019] IEEE).

For the purpose of subjective evaluation of the F_0 model built in the current work NSF was trained on the whole training subset of the current corpora. First, mel cepstral features were extracted using a 16000 sampling rate, 400-point frame length and 80-point frame shift. The FFT was performed using 1024 points. The mel bands were constrained to 80 and the minimum and maximum were set to 0 and 8000 Hertz respectively. Although the test set already contains extracted F_0 values, the extraction was rerun using the tools included in the NSF framework to ensure data consistency. The extraction algorithm was configured to run with minimum and maximum set to 60 and 400 Hertz, and a 35-point frame length and a 10-point frameshift were used. The dataset and the training script were deployed in the Amazon Web Services cloud and the hn-NSF model was trained using a single Tesla V100 GPU, 8 vCPUs and 32GB of ram. The training took 4.8 hours without the inference. The trained model was used to resynthesize all recordings in the test set, once with their natural F_0 , and once with F_0 values predicted by the model built in this work resulting in total of 191 recording pairs. These were then used for subjective evaluation of F_0 model performance using two testing procedures as described below.

3.4.8 Perceptual Evaluation

In the current work, two traditional testing procedures were used for the purpose of subjective evaluation of the intonation generated by the implemented F_0 model, ABX discrimination test and a simple MOS test. The experiment was implemented as a web application using a lightweight Python backend framework Flask [118] for building the REST API and the JSPsych [66] Javascript library as both the

frontend and the experimental engine. The JSPsych is a state-of-the-art library for running a wide range of laboratory-like behavioral experiments in a web browser. It enables an easy modular design for a range of experiments and provides a number of out-of-the-box functionalities like reaction timing and low latency audio playback.

The backend REST API exposed 5 endpoints, i.e.:

- simple healthcheck endpoint returning one of Wittgenstein's famous quotes: "Die Grenzen meiner Sprache bedeuten die Grenzen meiner Welt." [325],
- the main endpoint used for rendering the experiment through a template HTML in which the main JSPsych script was included,
- a save endpoint, that was called at the end of each experiment and was used for saving the results, cleaned by some additional controller methods, into a MongoDB-compatible database [125],
- a static URL serving a brief, light description of the whole study serving as a departure screen for participants,
- an additional endpoint for getting a JSON-serialized [243] list of all current results in the database.

In order to provide high availability, reliability and accessibility of both, the application and the data itself, especially under some increased traffic and use, the whole application was deployed to an Amazon Web Services-powered cloud infrastructure. The application was packaged as a single Docker container and deployed as an AWS Elastic Beanstalk application. This solution provided an auto-scaling mechanism (AWS Elastic Load Balancer) [25] that was spawning new EC2 instances of the application when an increased traffic was experienced, and scaling their number down when the traffic was decreasing. The database was also implemented as an AWS cloud solution using their MongoDB-compatible DocumentDB⁵ service. Similarly to the application itself, the database was replicated according to the experienced load. Another advantage of this solution was that the state of the database was automatically stored during a daily snapshot occurring at midnight when the lowest traffic was anticipated. The application and the database as well as all intermediate solutions described upto this point were communicating via a Virtual Private Cloud that is inaccessible from the external networks. The only entrypoint was through the actual HTTP port of the API exposed via the AWS ELB infrastructure.

⁵<https://aws.amazon.com/documentdb/>

In order to provide high accessibility to the audio recordings used as the stimuli in the experiment, they were deployed as a public data bucket on the AWS Simple Storage Service (S3)⁶. The experiment was, in turn, designed to preload all the audio files as the initial step to provide the lowest latency possible.

Additionally, the application was registered under the following public domain: <http://fonetyka.cudaniewidy.org>, which helped increase user trust and resulted in an increased number of participants. The application is maintained and can be still accessed at the given url for reference.

The whole infrastructure is presented in Figure 3.6.

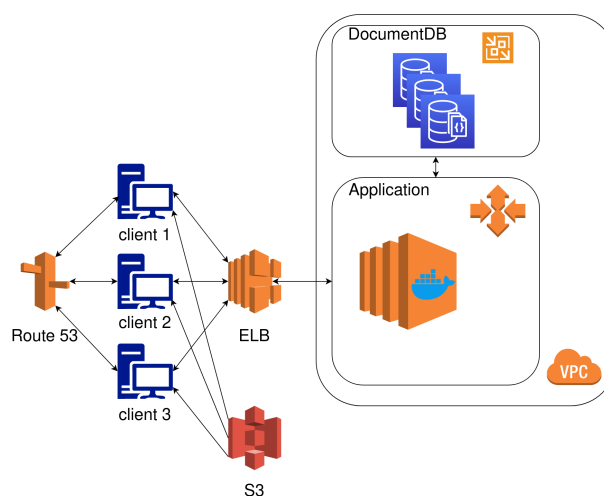


Fig. 3.6.: Infrastructure of the web experiment framework.

The experiment was first tested using a staging instance residing at a different URL than the production instance. Only after a series of corrections and validations under different user scenarios the application was deployed to the production instance that was sent out to a group of IT-professionals as well as some people not working with technology. Only a few of the included subjects can be considered trained phoneticians, linguists or audio-specialists (< 5). The subjects were instructed not to open the link on mobile devices as this was not supported since a hardware keyboard is required for input. This also limited the number of users participating in the experiment in outdoor environments under bad background noise and with numerous distractions.

Over a 100 responses were collected during a week-long trial. After that period a database snapshot was taken and saved and the application was left running

⁶<https://aws.amazon.com/s3/>

collecting further results. The flow of a single experiment trial is presented on 3.7 below.

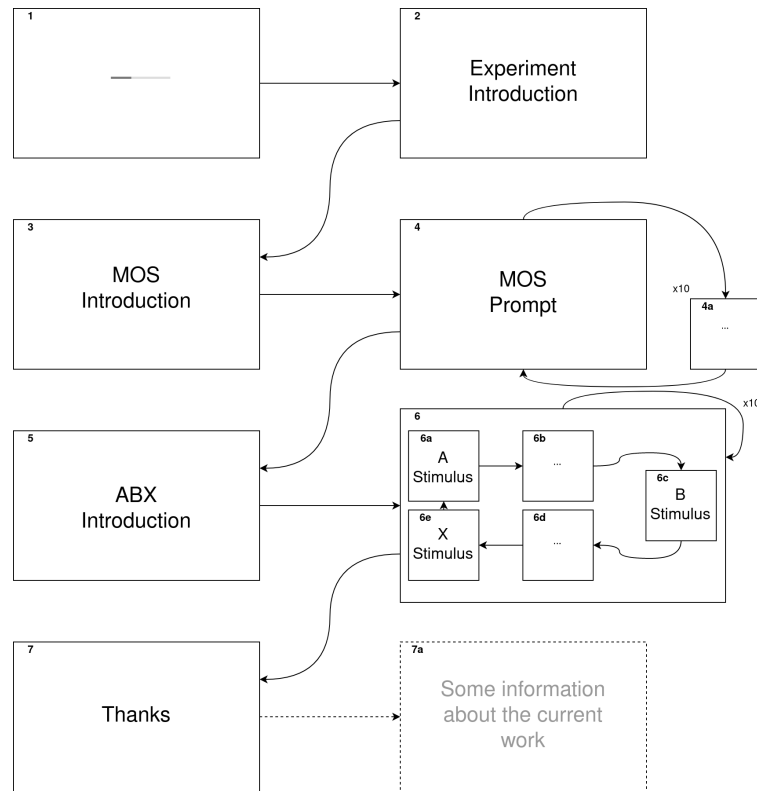


Fig. 3.7.: Perceptual evaluation experiment trial scheme.

When a subject accessed the provided link, first an animated progress bar was displayed (1). During that time 20 filenames from the test set were randomly selected from the list, 10 for a MOS evaluation and the other 10 for an ABX test. This stage ensures the trial to be a double-blind and by preloading the audio to memory it enables low latency playback. Once all images were successfully loaded an introductory screen was displayed (2). Here, the subjects were informed that the experiment consists of two parts; a rating of naturalness of a series of 10 recordings and a discrimination procedure of another 10 triplets of recordings. The screen also emphasizes that all recordings will be played automatically, each only once, and that the playback cannot be stopped.

After clicking the button the user was taken into the introductory screen of the first part of the experiment, the MOS test. Here, another introduction is displayed (3), this time addressing the MOS-part of the experiment, introducing the rating scale and instructing the subject to use the numerical keys on the keyboard after each recording to rate it accordingly. Additionally, a hint was provided that in some of

the stimuli the natural “melody of speech” was replaced with a synthetic one. When the subject clicked the **Next** button the MOS procedure was started. Each stimulus was preceded by a 1000ms pause during which a **...** (4a) was displayed on the screen. After the pause a single recording was played chosen from the shuffled set of 10 natural and synthetic (5/5) stimuli. The subject could rate the stimuli from the immediate start of playback with a **1 - 5** keypress, which ended the playback if attempted mid-stimulus. During playback a small call to action was displayed with the available keys listed once again (4). After 10 iterations of the above procedure an introductory screen for the ABX trial was displayed (5). Similarly to the previous part, simple instructions along with available keys **a** and **b** were listed here followed by a **Next** button that started the ABX part of the experiment. Each of the 10 iterations (6) of the ABX part consisted of two version of the same sentence, one with the synthetic F_0 predicted by the model and one with natural unchanged F_0 , played one after another in random order (6a, 6c) and each preceded by a 1000-ms pause (6b, 6d). The A and B templates were immediately followed by the X stimulus which was one of the previous recordings chosen randomly by the script. At that time a call to action was displayed (6e) for the subject listing the available keys **a** and **b**. After this part a typical thank you screen was displayed (7). This final screen included a hyperlink to a short description of the current work⁷ (7a).

⁷Available at <http://fonetyka.cudaniewidy.org/departure>

Results

„If the facts conflict with a theory,
either the theory must be changed or the facts.

— Baruch Spinoza

The results of F_0 inference conducted for each of the 191 utterances in the test set randomly selected from the original database and not included in the training data, were plotted against the ground truth values as shown in Figure 4.1.

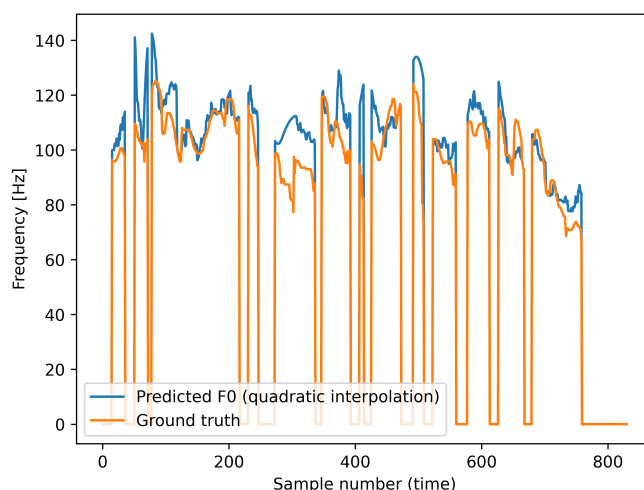


Fig. 4.1.: Result of F_0 prediction for "Lokatorzy znaleźli się w podbramkowej sytuacji i musieli się wyprowadzić" (*The tenants found themselves in a difficult situation and had to move out*).

As can be seen, the model seems to demonstrate very good performance on average with some occasional differences in contours from the ground truth contour. The plots show the interpolated results where missing values were back- or forward-filled depending on their location as described in the previous chapter of this work. Some parts of the predicted contour can also be seen as detached from the original one but still preserving its general characteristics and dynamics. Such displacements can

be seen as errors when using objective evaluation metrics but might turn out to be rather insignificant under subjective perceptual evaluation.

Given the small size of the training data, incomparable to the size used for training speech models nowadays, which usually include hundreds of hours of recorded speech, and given the fact that the values were inferred only from a narrow choice of very high-level linguistic features such as phoneme and syllable identity, position and context, the model seems to perform outstandingly well and seems rather stable across the whole test set as can be seen in Figure 4.2.

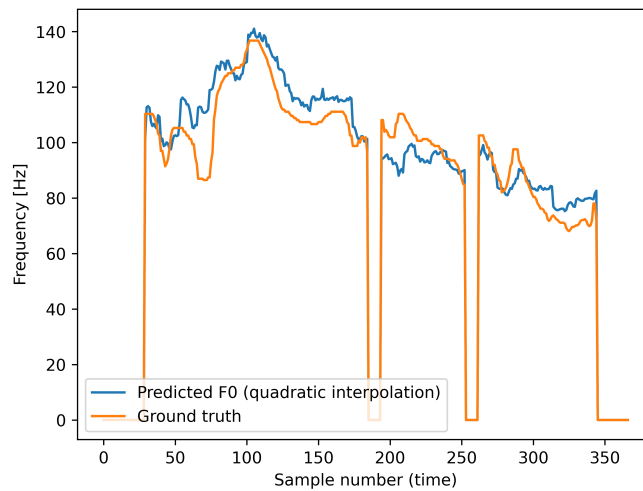


Fig. 4.2.: Result of F_0 prediction for "Powodzenie nie jest gwarantowane" (*Success is not guaranteed*).

In some of the best examples in the test set the results seem truly impressive, e.g. in Figures 4.3, 4.4, 4.5, 4.6 and 4.7.

Both good and bad examples can also be found for the results of the technique used to fill missing values in the inferred F_0 to match the ground truth, e.g. at the phrase onset in Figure 4.7. The values for the short initial fragment around $t = 50$ were copied and filled from the first predicted value around $t = 80$ resulting in a single flat bar. The interpolated regions can also be quite easily identified on the first example in this chapter (Figure 4.1), around $t = 275$ and $t = 500$. Another such example is seen in Figure 4.8:

The last example (Figure 4.8) also shows that the model was unable to predict the phrase final frequency raise realising the phrase boundary tone typical for most

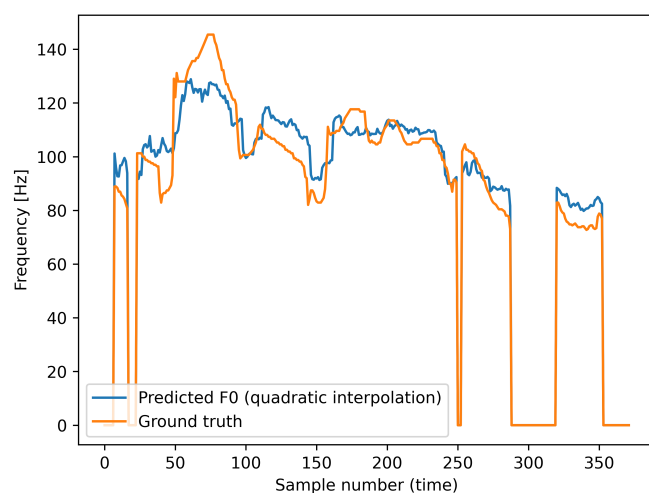


Fig. 4.3.: Result of F_0 prediction for "Gaduła była bardzo nieznośna" (*Gabby was very annoying.*).

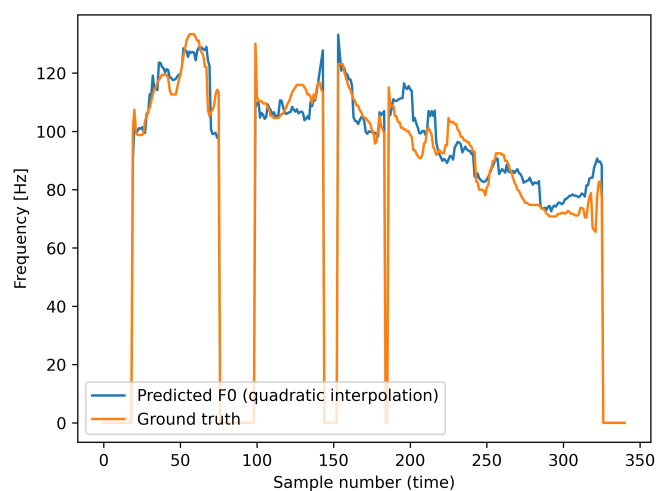


Fig. 4.4.: Result of F_0 prediction for "Może przyniosą też gorzałę" (*Maybe they will bring booze too.*).

questions in Polish. After an analysis of the other questions in the test set it turned out that this was a systematic problem.

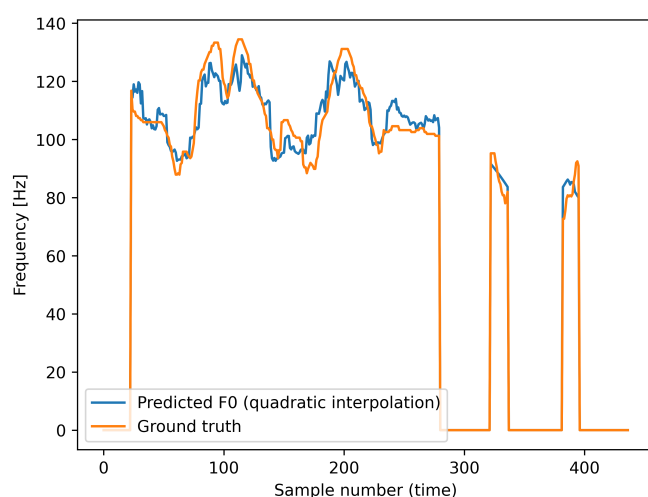


Fig. 4.5.: Result of F_0 prediction for "To jest ważna godzina dla nas wszystkich" (*This is an important hour for all of us*).

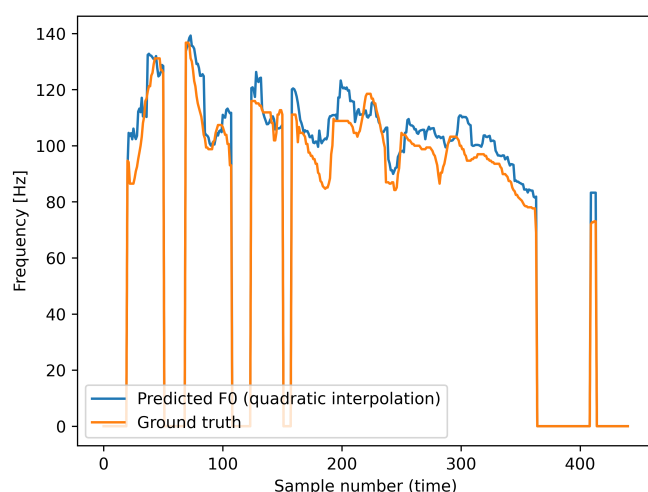


Fig. 4.6.: Result of F_0 prediction for "Myślę, że chleb razowy będzie najlepszy" (*I think that a wholemeal bread will be the best*).

Some examples, although these are rather rare, show improper ground truth F_0 values, mainly due to errors in the extraction algorithm as can be seen in Figure 4.9. Because no F_0 smoothing and no F_0 stylization was applied as a preprocessing step, these errors include mainly pitch halving and pitch doubling. These appear

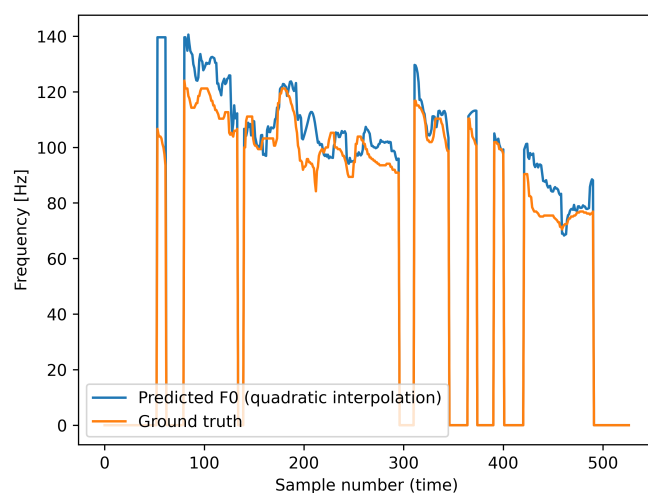


Fig. 4.7.: Result of F_0 prediction for "Słyszałam odgłos zbliżającego się pociągu" (*I heard the sound of an approaching train*).

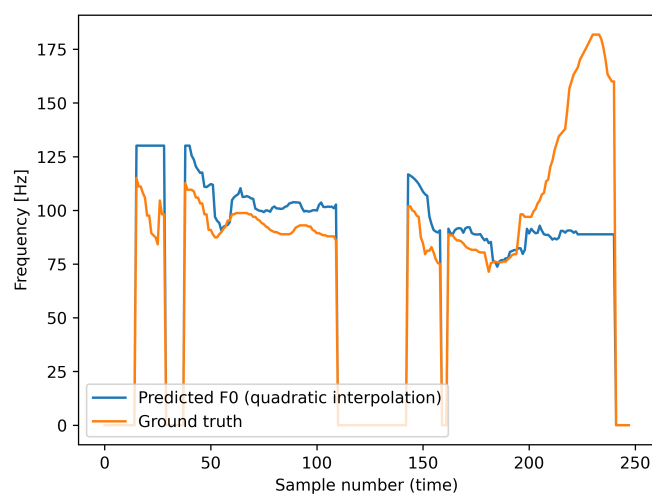


Fig. 4.8.: Result of F_0 prediction for "Czy to był łatwy dobór?" (*Was it an easy choice?*).

mainly in parts containing some inadequately expressive realizations by the speaker, as visible phrase-finally in Figure 4.10, or regions characterized by vocal fry, also phrase-finally.

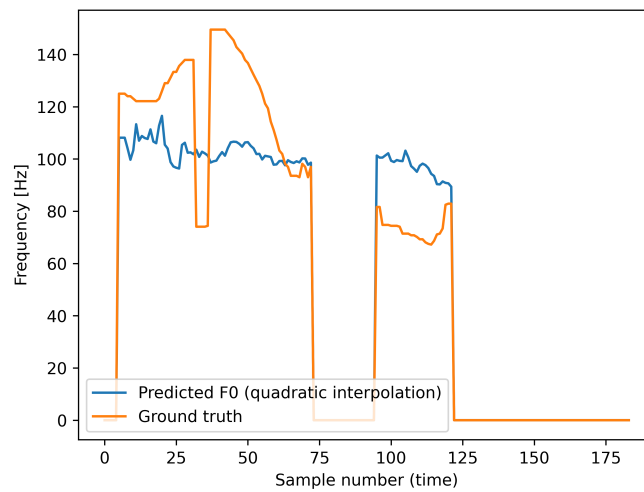


Fig. 4.9.: Result of F_0 prediction for "To Majka" (*This is Majka*).

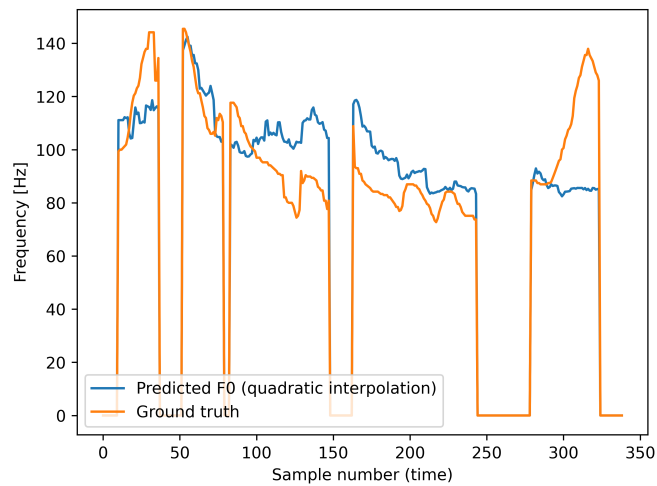


Fig. 4.10.: Result of F_0 prediction for "Na czym polega kandyzacja?" (*How does candying work?*).

4.1 Objective Evaluation

Although preliminary inspection of the predicted F_0 values shows some tendencies, a formal method for calculating the errors was also applied. Table 4.1 lists the mean values of each of the calculated errors for the whole test set.

Tab. 4.1.: Mean MOS scores in the test dataset for $\log F_0$ predictions.

MAE	0.015382
MSE	0.002495
NRMSE	0.009534
RMSE	0.048128

A number of more sophisticated metrics do exist for objective evaluation of synthesized speech, including the Peak Signal-to-Noise Ratio (PSNR) [150] and various spectral distortion measures (see for example Samuelsson et al. [266], for a comparison of the base methods), but this study is focused on intonation and such metrics can only be applied to a complete speech signal. In the current case, applying them to resynthesized speech would yield results overshadowed by the influence of the quality of the resynthesis method itself. They were, therefore, ignored here and the industry standard Mean Squared Error and Mean Absolute Error were calculated instead [200].

4.2 Subjective Evaluation

Prediction results were also used in two types of perceptual evaluation, as described in the previous chapter. The whole trial included responses from 90 different participants. Each response included MOS ratings for 10 randomly picked sentences from the test set and 10 results of ABX discrimination tests for pairs of synthetic/natural variants of other randomly picked test sentences. The random sampling for both experiments was performed independently with repetitions – two variants of that same sentence, synthetic and natural might have been picked for the MOS trial and then also reappear in the ABX trial for that same subject. Although the probability for such a case was rather low given the 191 sentences in the test set. The MOS scores for all results are presented in Figure 4.11.

Resynthesized speech samples containing synthetic F_0 values received a lower overall mean score than the resynthesized samples with natural F_0 . The mean score for the

synthetic F_0 was 2.9 and the mean for the natural F_0 reached 3.7. The synthetic stimuli were also characterized by a slightly bigger standard deviation 2.44 as compared to 2.31 for the natural ones. As shown in Figure 4.12, mean response time was also slightly longer for synthetic stimuli than for the natural ones with 5725.7ms and 5283.9ms respectively.

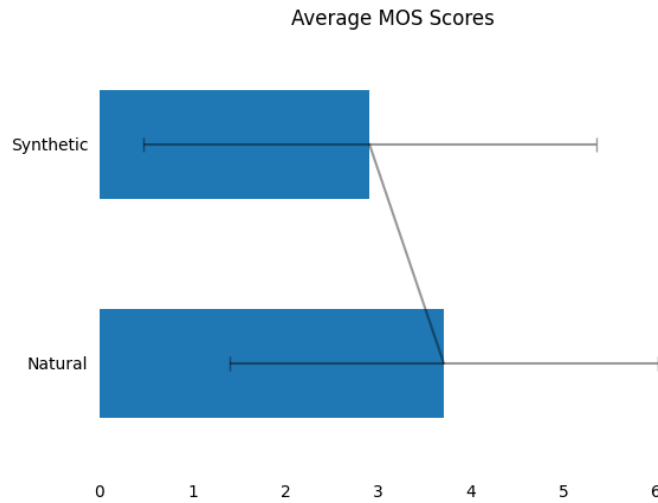


Fig. 4.11.: Mean Opinion Score-based evaluation results.

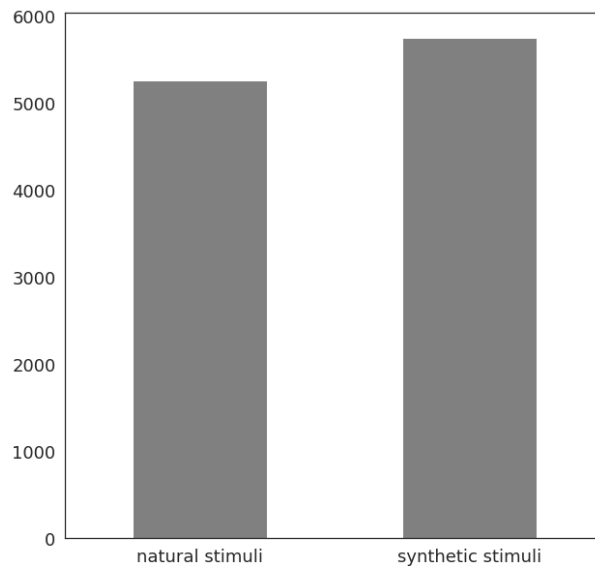


Fig. 4.12.: Mean Opinion Score-based evaluation mean response times for synthetic and natural stimuli.

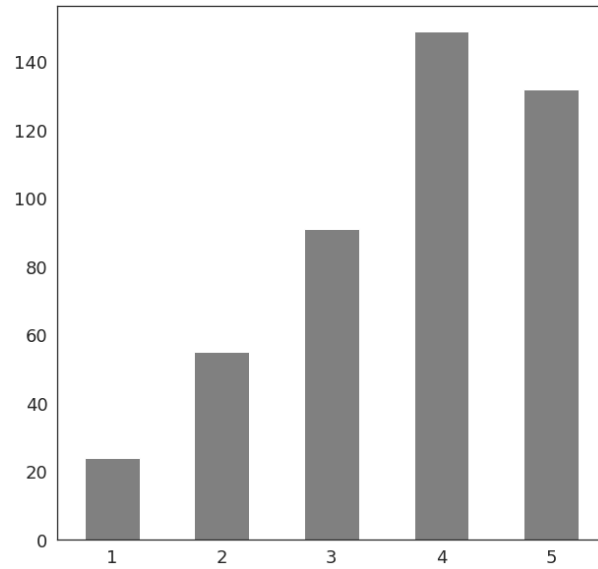


Fig. 4.13.: Mean Opinion Score-based evaluation total numbers of specific scores for natural stimuli.

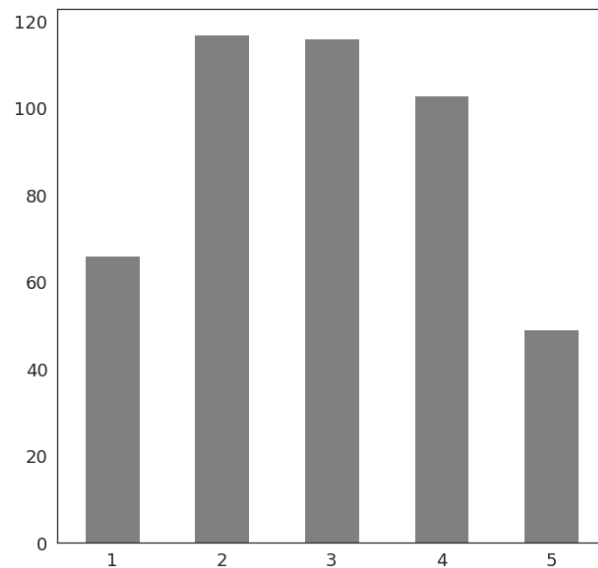


Fig. 4.14.: Mean Opinion Score-based evaluation total numbers of specific scores for synthetic stimuli.

Total numbers of specific scores for both types of stimuli are presented in Figures 4.13 and 4.14. It clearly shows that the score distribution for natural stimuli is leaning towards the higher values. However, the distribution of synthetic stimuli scores is closer to symmetric normal distribution without a strong trend towards the lower scores.

A correlation matrix of all MOS parameters is presented in Figure 4.15. It clearly shows a positive correlation between the type of stimulus and the answer. The value is small but significant. A negative correlation can also be seen between the amount of time elapsed from the beginning of the experiment and the given answer, which is to be expected as the subjects naturally gain experience towards the end of the experiment and become more critical. No strong correlation, however, exists between the other parameter pairs; stimulus type and response time, time elapsed and response time and answer and response time, which is coherent with some of the data in the other figures.

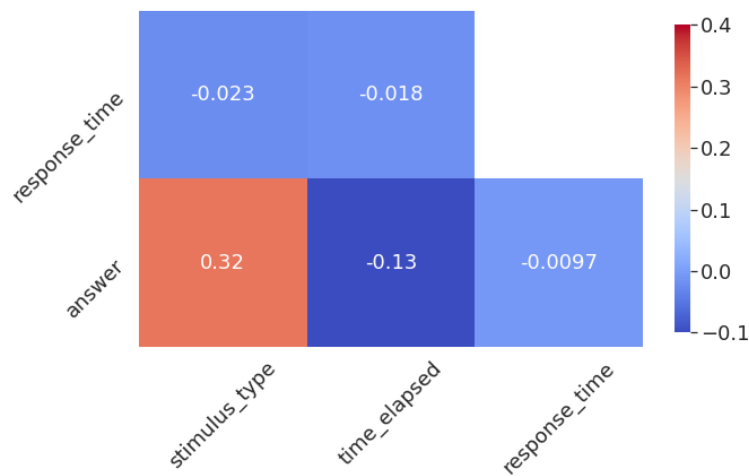


Fig. 4.15.: Mean Opinion Score-based evaluation parameters correlation matrix.

The results of the second part of the perceptual experiments consisted of an ABX discrimination procedure between pairs of resynthesized utterances with synthetic and natural F_0 contours. Even with a relatively low confidence of 95% assumed ($p = 0.05$), it was not possible to reject the following null hypothesis with a $\tilde{\chi}^2$ test:

Null Hypothesis (H_0). *There are no perceptually significant differences between resynthesized recordings with synthetic and natural F_0 .*

Hypothesis (H_a). *There are perceptually significant differences between resynthesized recordings with synthetic and natural F_0 .*

Figure 4.16 shows the mean ABX test score per listener and its standard deviation plotted against the assumed 95% confidence level, which lays at the level of 80% mean ABX test score.

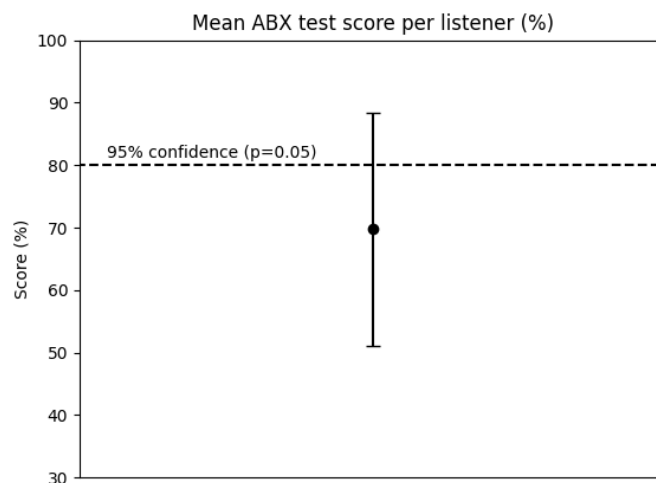


Fig. 4.16.: ABX experiment results. Mean score per listener (%).

A histogram representing numbers of individuals with n correct answers is presented in Figure 4.17 for reference.

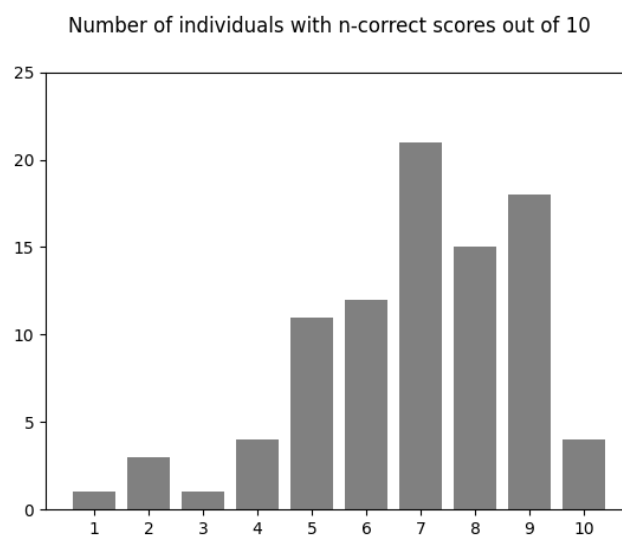


Fig. 4.17.: ABX experiment results. Number of individuals with n -correct answers out of 10.

Also in case of the discrimination experiment there were no significant differences between the mean response times in case of the main (X) stimuli being synthetic or natural as shown in Figure 4.18.

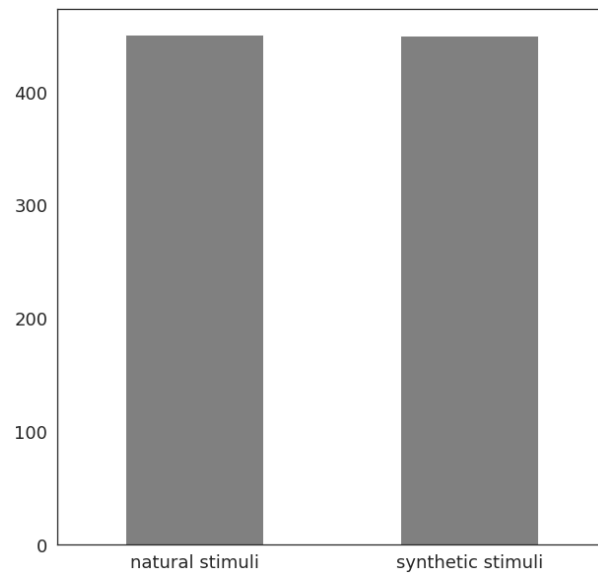


Fig. 4.18.: ABX experiment mean response times for synthetic and natural stimuli.

Similarly, the number of correct and incorrect answers was almost identical in both cases as illustrated in Figure 4.19.

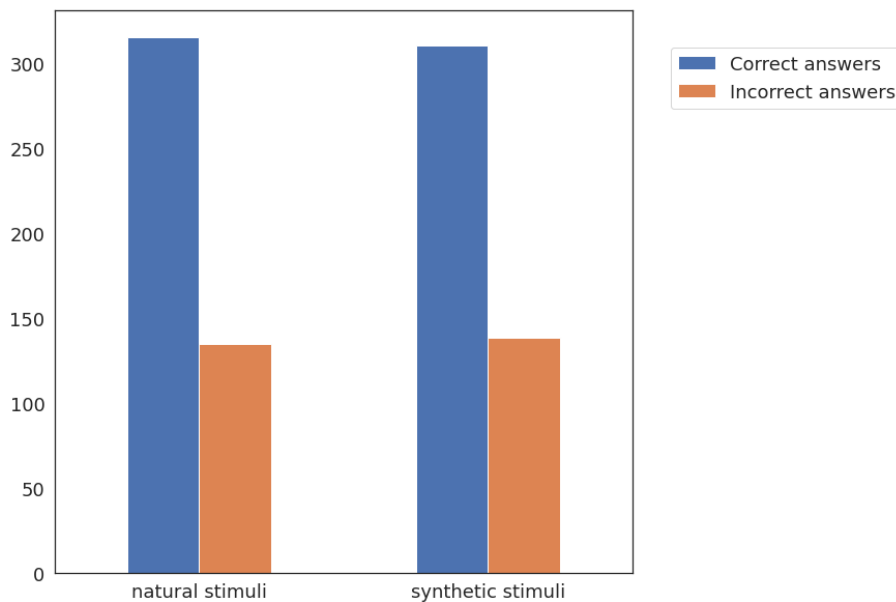


Fig. 4.19.: ABX experiment number of correct and incorrect answers in case of natural and synthetic stimuli.

These same observations can be made for the correlation matrix for the experiment parameters presented in Figure 4.20. No strong correlation was found between the stimulus type, response time and whether the answer was correct. Some weak negative correlation exists for the response time and the correctness of the answer as well as for the amount of time elapsed since the start of the experiment and the correctness of the answer.

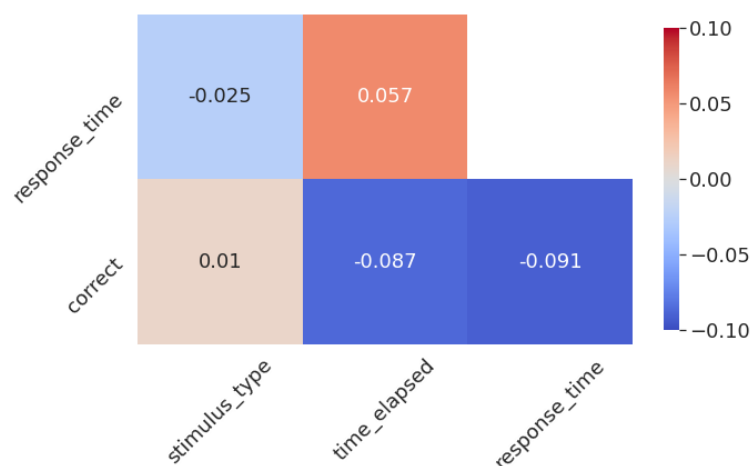


Fig. 4.20.: ABX experiment parameters correlation matrix.

Raw unprocessed results containing detailed answers, response times and other metrics are available at the experiment code repository at https://github.com/mrslacklines/listening_experiments/blob/master/results/results.json.

4.3 Feature Relevance

Each of the generated relevance maps was plotted as a heatmap with the x -axis corresponding to the sample number (discrete time) and the y -axis representing a feature on a given numerical index ranging from 0 to 1297 as shown in Figure 4.21.



Fig. 4.21.: Fundamental frequency predictions for "Słyszałam odgłos zbliżającego się pociągu" (*I heard the sound of an approaching train*) aligned with feature relevance heatmap.

The relevance, represented on a continuous colormap scale, was additionally *log*-normalized symmetrically around the 0-value, with linearity threshold set to 3, and the linear part scaled using a 0.001 factor for better readability. Additionally, both the predicted and the ground truth $\log F_0$ values were plotted above the relevance maps with time-axes aligned for better representation and reference.

The mapping of the numerical indexes to actual feature names is available in Appendix A and all of the relevance plots are available in the original project repository in the results folder¹.

As described in Section 2.2.4 the algorithm calculates a value that considers not only which parts of the input can account for the outputs but also the actual activations representing the relevance of a given feature value at a given point to an actual prediction. Positive relevance indicates positive evidence for the predicted F_0 values and negative relevance indicates evidence against the predicted values.

Because the large number of low-level features makes identifying general trends difficult, mean relevance was also calculated for a number of different higher-level feature groups. The definitions of these groups along with a comprehensive description of the rationale behind each of them is available as Appendix B of this work. Figure 4.22, for instance, shows the relevance of features grouped by the general information type they convey.

As can be seen in Figure 4.23 and 4.24 the highest positive relevance values were invariably reported for the Voiced/Unvoiced feature and are marked by a band of the most intense dark green on the heatmap. Additionally, the features corresponding to the position of the phrase in the utterance usually show mostly negative relevance. The other most relevant features groups are the relative word position, segment position in the syllable (i.e. current position or t), presence of syllable accent and the syllable accent type, as well as syllable length measured in segments, syllable position, word length measured in syllables, word position and word context. Some lesser positive relevance can be also observed quite regularly for phrase and utterance length, syllable neighbourhood, relative position of the syllable and all segment-related features. Information about the nucleus type of the syllable usually renders as irrelevant with only some occasional negative relevance appearing at boundary regions, which are generally characterized by very high variance of relevance values.

¹https://github.com/mrslacklines/intonation_synthesis/tree/master/intonation_synthesis/results

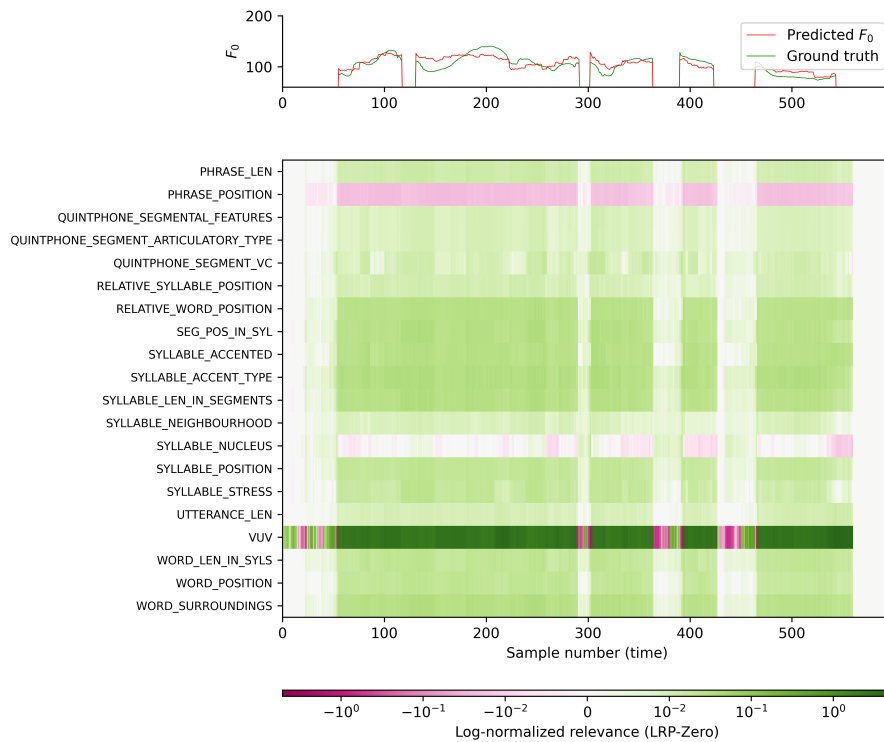


Fig. 4.22.: Fundamental frequency predictions for "Waluta jeny - w języku greckim - jest cenna" (*The currency yen, in Greek, is valuable*) aligned with relevance heatmap for general feature groups.

In Figure 4.25 the features were in turn grouped according to the level of utterance organization to which they pertain (disregarding the V/UV from now on). Similarly, the word and syllabic features turned out to exhibit the highest relevance. However, all the other feature groups, the phrasal, segmental and utterance levels, are also showing positive mean relevance.

When the features are grouped into categories regarding the various kinds of relation they convey, as in Figure 4.26, some other strong trends can be observed. The central visible darker band suggests that it is mostly the features that express various kinds of positional relations between linguistic units in the utterance that provide the most positive evidence for the F_0 contour prediction. Also the features related to both, the composition of a linguistic unit itself, and the composition of its parent unit were found to be highly relevant, although to a slightly lesser extent. *Composition* is used here as an umbrella term for all information about the number and kind of subordinate units, e.g. number of accented syllables in the including phrase

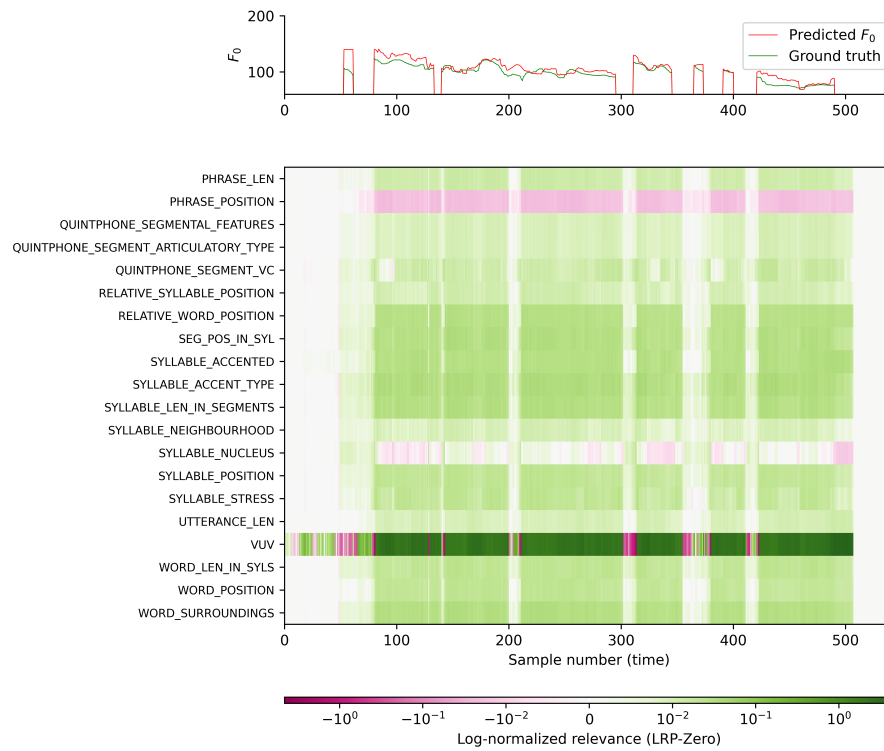


Fig. 4.23.: Fundamental frequency predictions for "Słyszałam odgłos zbliżającego się pociągu" (*I heard the sound of an approaching train*) aligned with relevance heatmap for general feature groups.

or length of syllable measured in number of segments. The relevance of relative position also shows regions with greater values. The least pronounced relevance was observed in case of qualitative features such as the phonetic identity of a segment or the type of syllable nucleus.

In Figure 4.27 the two previous perspectives were combined and the mean relevance was plotted for features grouped by both, the relation type and the level of utterance structure they relate to. Here, the greatest relevance was reported for the segment position in the utterance, length of syllable measured in number of segments, position of word in relation to their immediate accented neighbours and length of the parent unit in measured in number of words. The position of the syllable, word composition and absolute position were also rendered as comparatively significant but to a much lesser extent. The only feature group that provided evidently negative evidence was again the information about phrase position in the utterance.

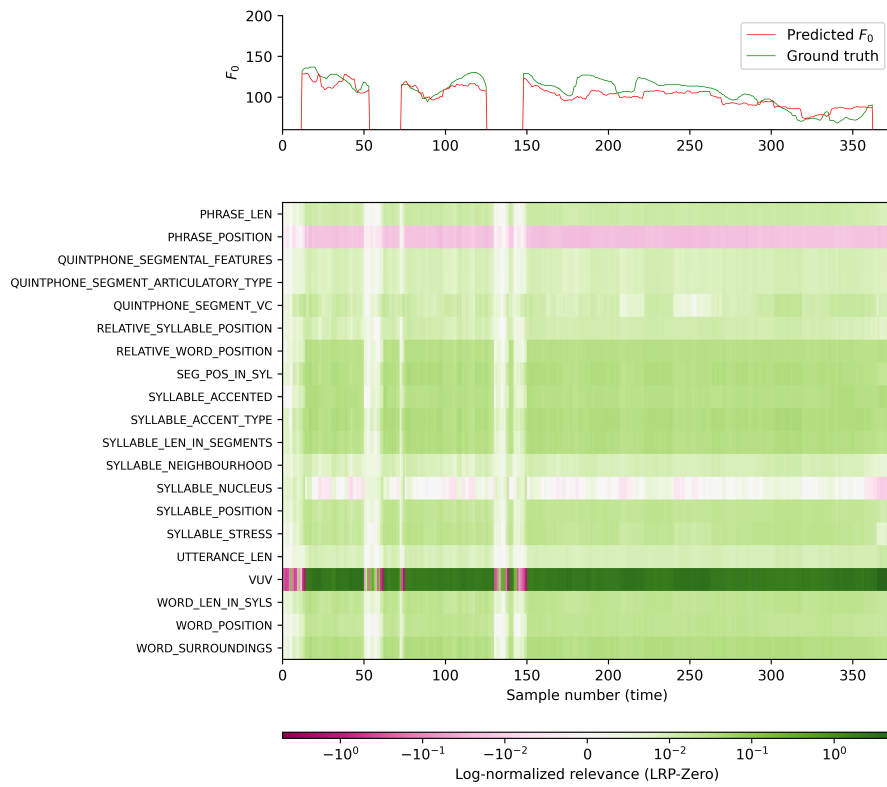


Fig. 4.24.: Fundamental frequency predictions for "Tego z pewnością żaden nie zrobi" (*Certainly none will do that*) aligned with relevance heatmap for general feature groups.

Given the infinite number of linguistic contexts and possible F_0 contour configurations it is impossible to infer the general relevance of a feature only with a look at a single relevance map for a single utterance, even when the features are grouped into meaningful high-level categories. In order to show general trends in feature relevance across many different contexts, a number of calculation methods were used to aggregate the values in all 191 individual relevance matrices. Aggregation methods included regular element-wise summation (results are illustrated in Figure 4.28), sum of absolute values and sums of only positive and only negative values, as well as a corresponding mean matrix for each of these summation methods.

All aggregated results show a natural trend for higher values towards the beginning of the time axis where the number of examples was naturally higher. Already at this point, some general tendencies can be observed in some more intense red and green bands for some of the individual features. Figure 4.29 therefore shows which

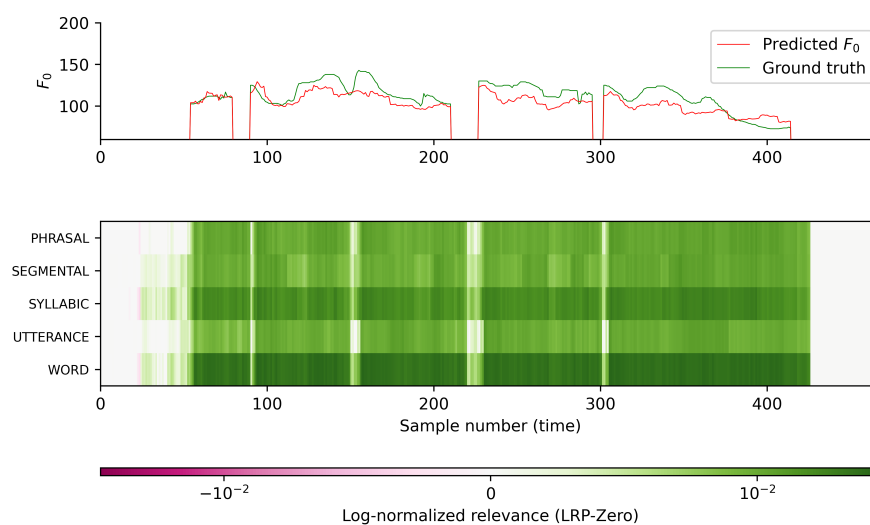


Fig. 4.25.: Fundamental frequency predictions for "Musisz dojrzeć do tego, by to zrozumieć" (*You have to grow up to understand it*) aligned with relevance heatmap for features grouped by the level of utterance.

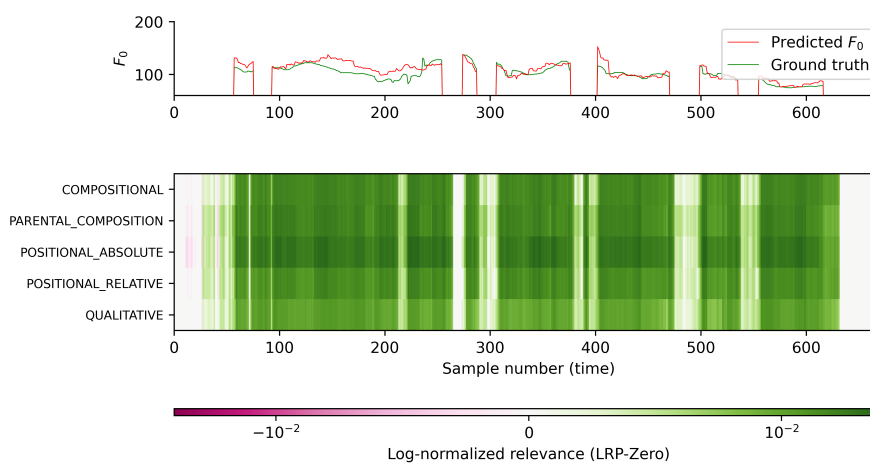


Fig. 4.26.: Fundamental frequency predictions for "Przepłynęłam na grzbiecie siedemnaście długości basenu" (*I swam seventeen pool lengths on my back.*) aligned with relevance heatmap for features grouped by the type of relation.

of the individual features provided the highest and lowest evidence for predictions across the whole dataset on average. Figure 4.30, in turn, illustrates which of the individual features were the most prominent across all test samples, regardless of

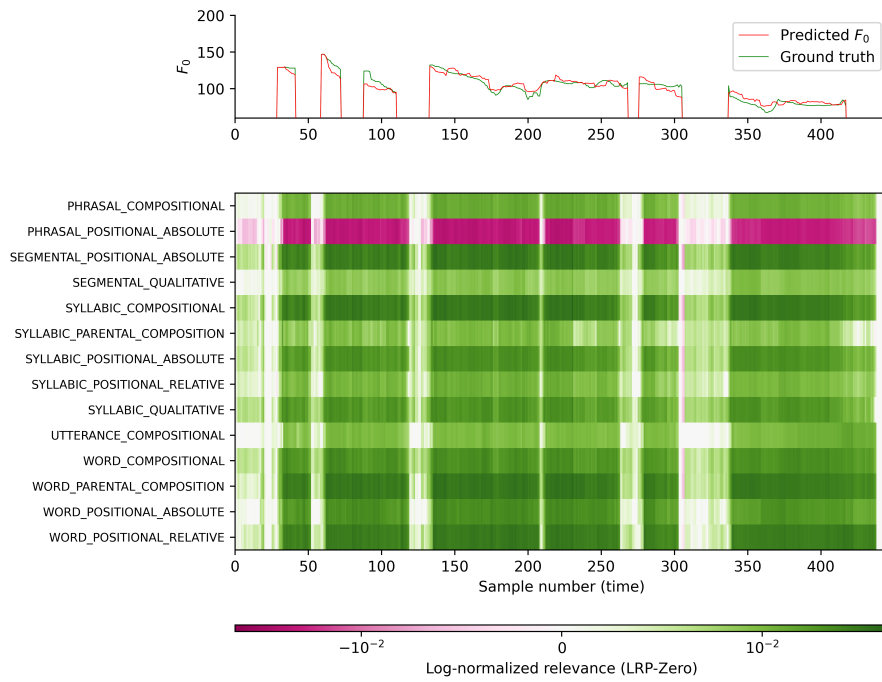


Fig. 4.27.: Fundamental frequency predictions for "Ciocia pracuje w urzędzie państwowym" (*Aunt works in a government office*) aligned with relevance heatmap for features grouped by the type of relation and the level of utterance.

whether the evidence it provided was positive or negative. Separate plots for overall mean negative and positive prominence (absolute relevance) are shown in Figure 4.31 and Figure 4.32.

Mean relevance was then calculated across the whole x -axis (time) per feature group using the same grouping scheme as in case of the individual relevance results (see Appendix B). Based on the results the feature groups were ranked using a number of different levels of data abstraction. In order to avoid excessive scaling of the y -axis because of huge relevance of the Voiced/Unvoiced feature additional versions of all plots excluding this single feature were also created.

The plot of the mean relevance across the whole test set (in Figure 4.33) shows similar trends to those that could be observed in case of individual relevance plots. After removing the V/UV feature, in order to prevent scaling of the other values, it clearly shows (see Figure 4.34) that the single most relevant feature was the syllable accent type (so the ToBI tone). The next five most relevant features exhibit very similar values. These are the length of syllable in number of segments, presence of

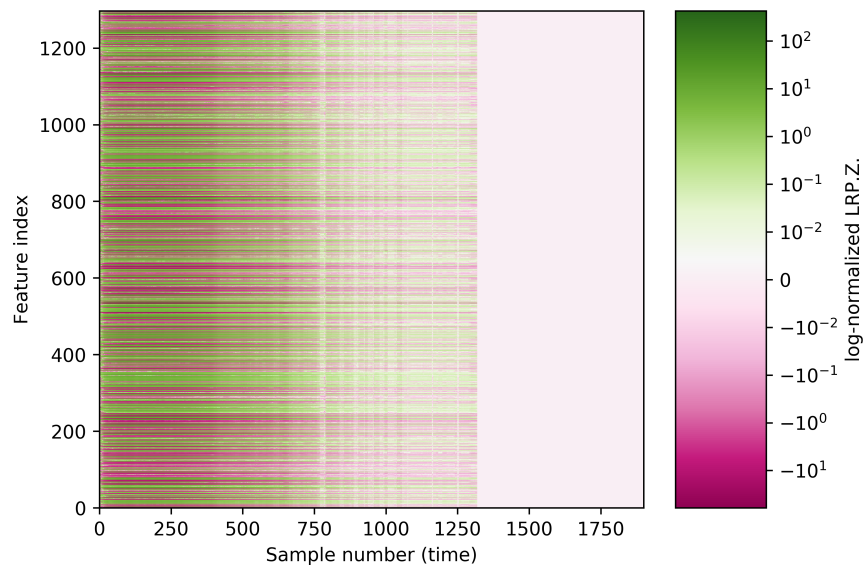


Fig. 4.28.: Relevance sum across the whole test set.

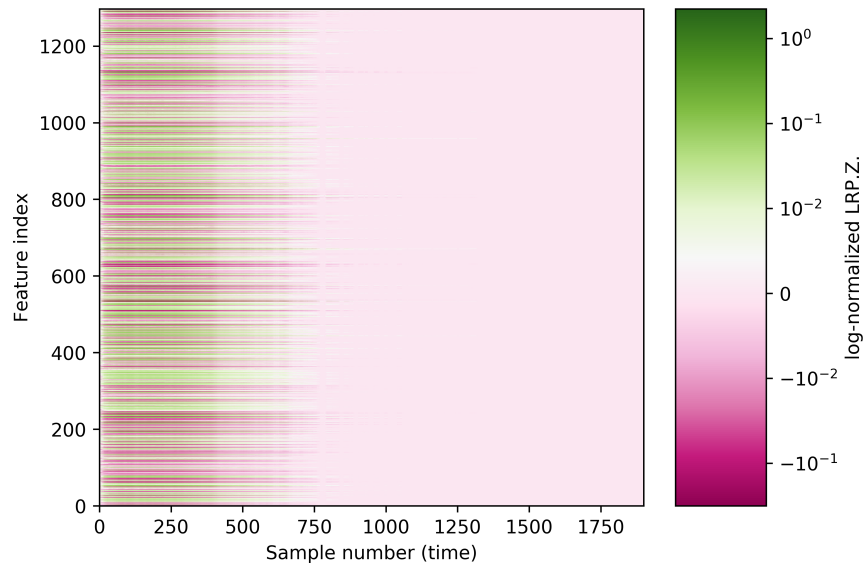


Fig. 4.29.: Mean relevance of features based on regular sum.

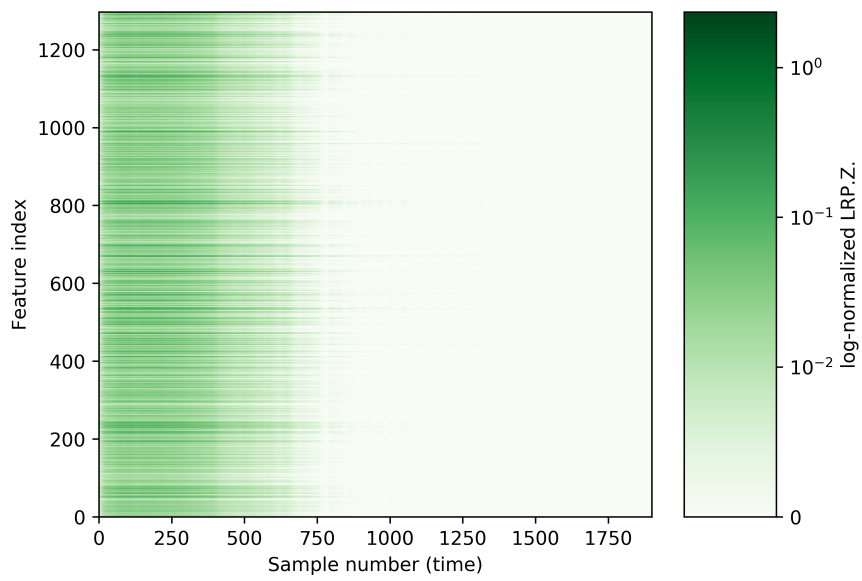


Fig. 4.30.: Absolute mean relevance of features (based on absolute sum).

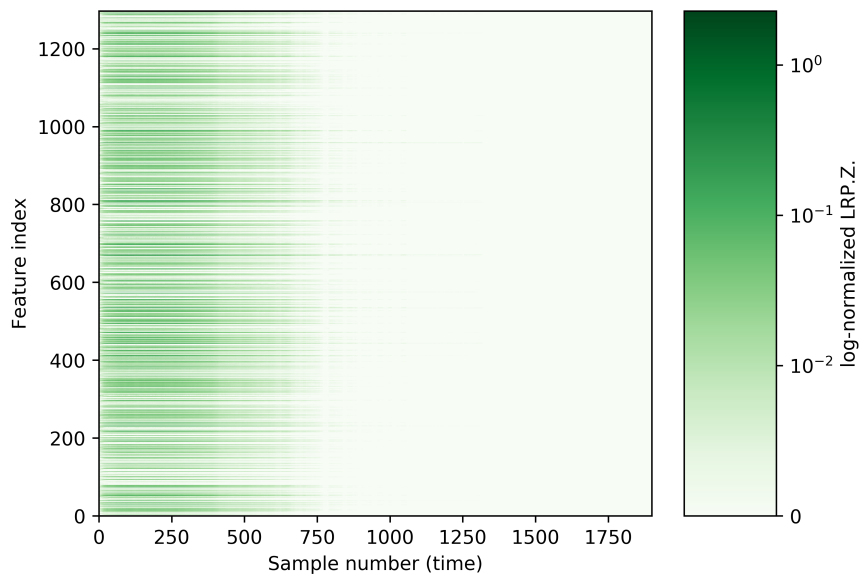


Fig. 4.31.: Mean relevance of features calculated for positive-only values.

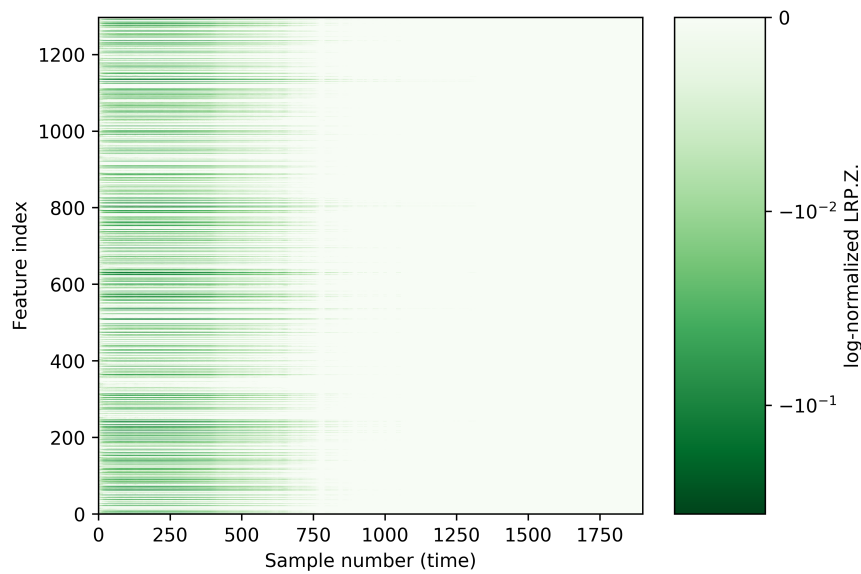


Fig. 4.32.: Mean relevance of features calculated for negative-only values.

syllable accent, segment position in syllable, the number of other accented (content) words before and after the current word (word surroundings) and the distance to these words (relative word position). Then, with slightly lower mean relevances, syllable and word position features follow, along with length of word in number of syllables and the information about syllable stress on the farther positions. The other features, or feature groups, are ranked even lower and these include most of the segmental and utterance features with phrase length having the highest mean and the information about the type of syllable nucleus and the position of the phrase in the utterance having visibly lower relevances.

In Figure 4.35, that same ranking based on the mean relevance is presented but this time with a finer granularity. Here, some other tendencies can be spotted. These also seem much more apparent as lower granularity evidently made the group results appear much more normalized introducing a damping factor for the more relevant members across the groups. Here, the most relevant feature group seems to be the number of accented syllables after the current syllable in the current phrase, immediately followed by the distance from the next accented (content) word expressed in number of words and by the forward position of the syllable in word. The syllable accent type reappears with an identical name as this whole feature group is small and was not further broken down. It was included as is in the general

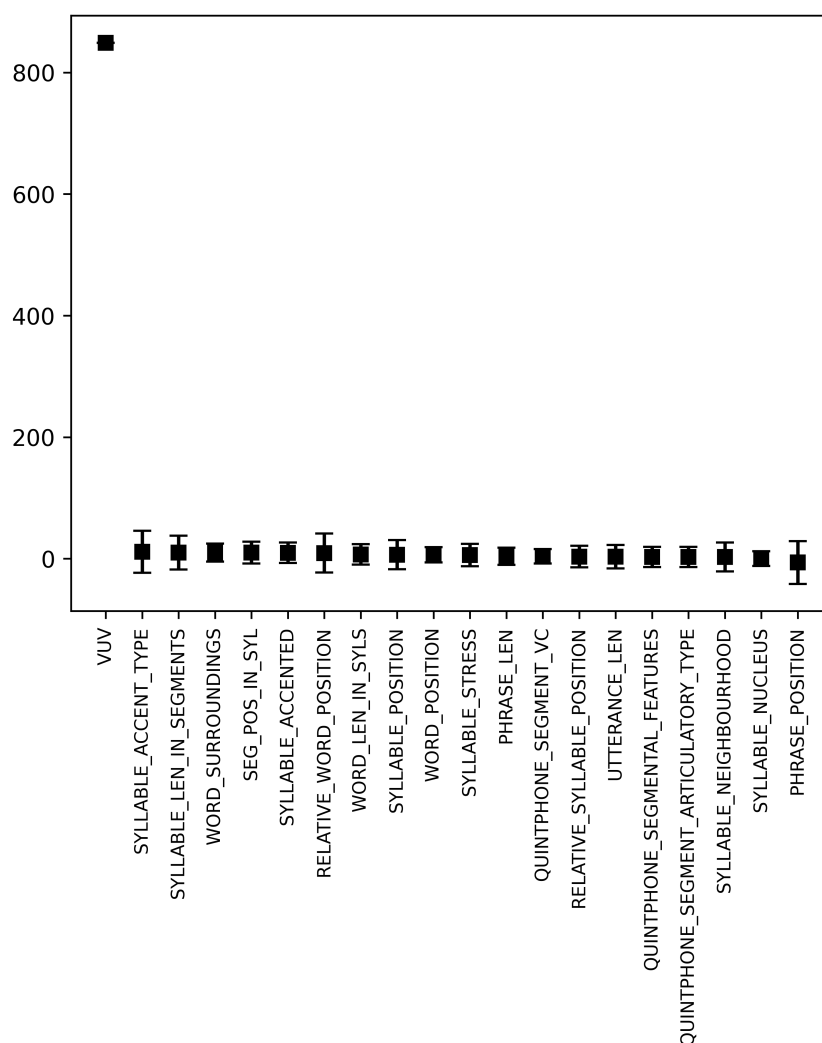


Fig. 4.33.: Regular mean relevance per feature group. The y -axis scaling due to the high relevance of V/UV renders comparatively flat plots for other features.

ranking in Figures 4.33 and 4.34), as the accent type label relevance was in some of the most interesting features and the groups were defined arbitrarily. This rendered the possibility that some of the most relevant features might be hidden within some of the more general groups and to address this intuition also the relevance for a very fine and detailed grouping was considered and the results can be seen in Figure 4.36.

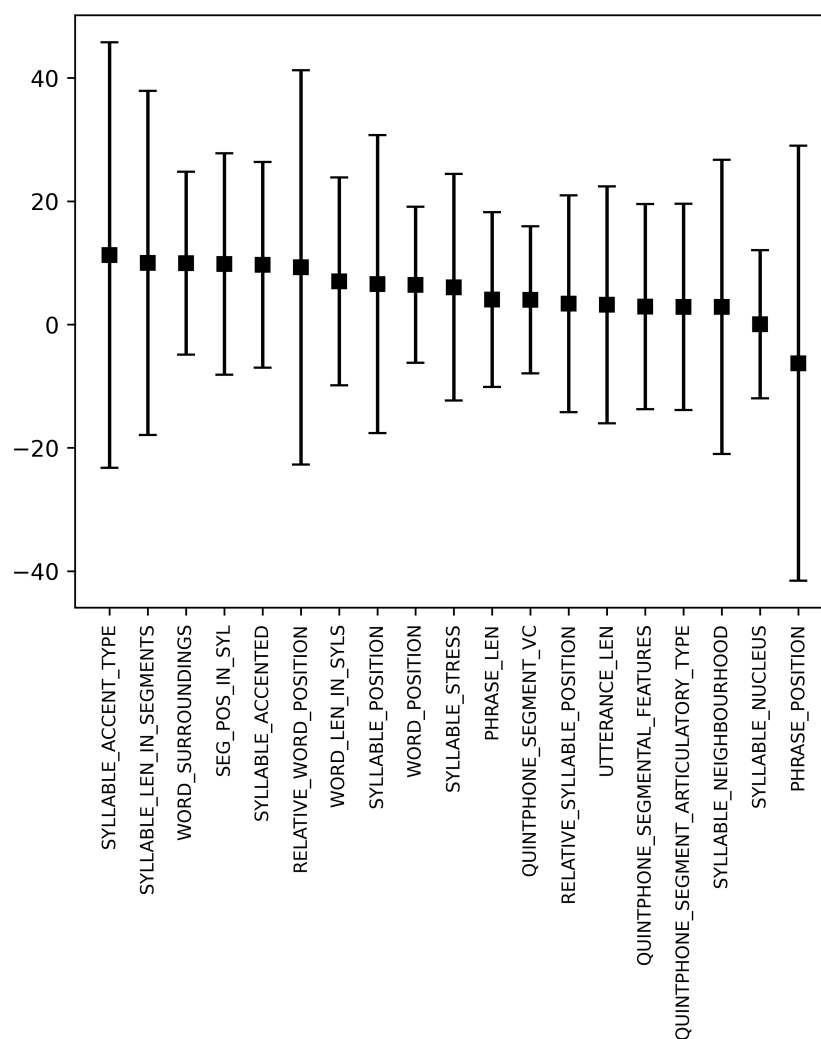


Fig. 4.34.: Regular mean relevance per feature group with the most relevant feature (V/UV) excluded.

The tendencies are even more vivid in this case but the results, at least for some of the groups, might appear harder to interpret later. However, the general tendencies are kept as demonstrated by some of the most and least relevant information.

Also in case of the mean rankings, additional plots were generated for some other aggregates of the relevance. In Fig 4.37, a feature ranking based on mean absolute relevance is presented. In this way the features were ranked according to their total contribution, positive or negative. Here, the mean absolute relevance of the syllable

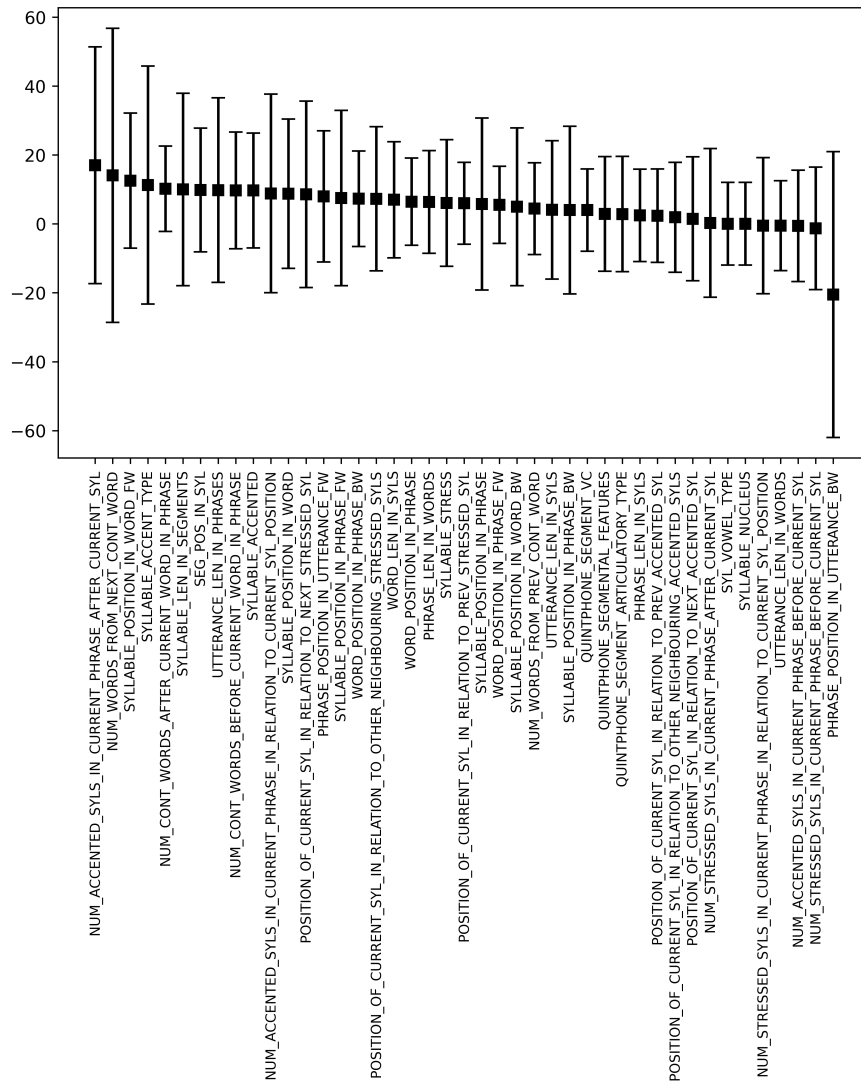


Fig. 4.35.: Regular mean relevance per feature group with the most relevant feature (V/UV) excluded. Medium granularity of feature groups.

accent type is ranked second immediately after the position of the phrase which was seen as providing strong negative evidence on most of the individual plots but also in the mean-based rankings. The inventory of the most relevant features is kept but the order is slightly changed as a result of the features providing mostly positive values being favored over the ones that might have higher mean relevance but also experiencing small ranges of negative relevance that evens out the overall absolute mean.

Additionally, Figures 4.38 and 4.39 present feature rankings based on the mean calculated with only positive or only negative values. These methods were included as a result of the tendencies observed with all of the previous aggregation functions, and in order to account for some other interesting tendencies that might have not been caught by the other aggregation methods. This includes the very high positive-only mean for the feature that usually delivers all-negative evidence – the position of the phrase, as can be seen in Figure 4.38. The following feature groups are similar to those seen in all of the previous figures.

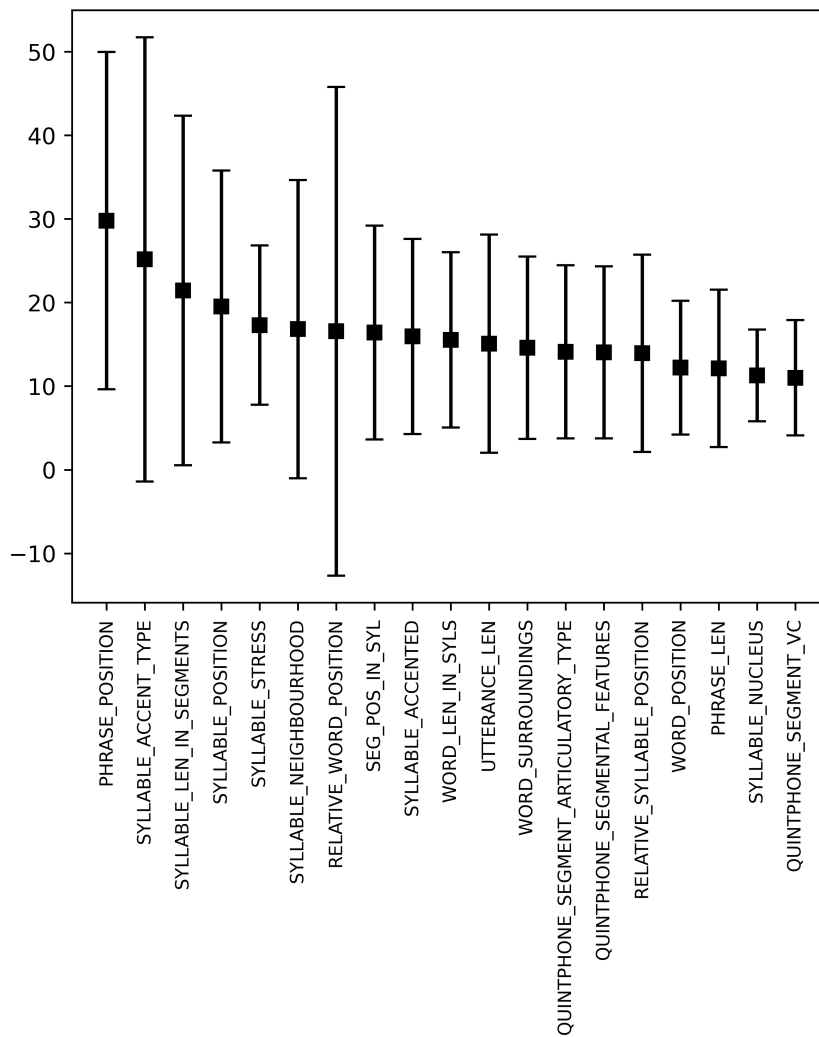


Fig. 4.37.: Absolute mean relevance per feature group with the most relevant feature (V/UV) excluded.

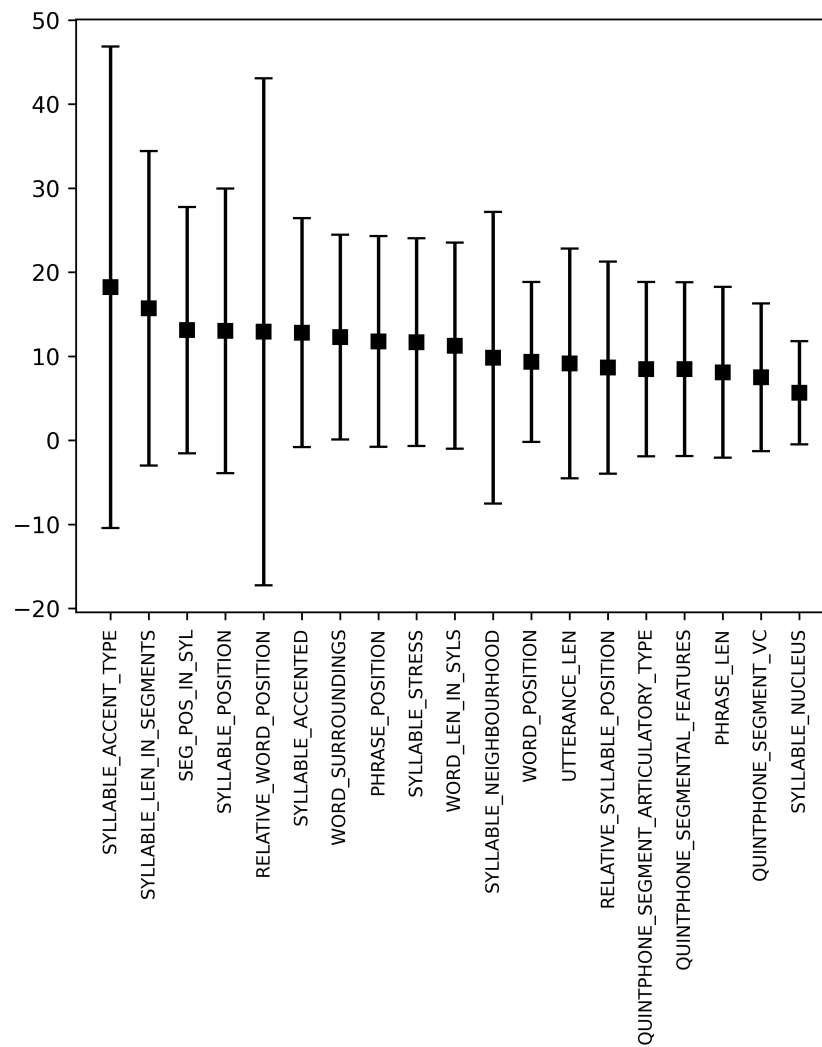


Fig. 4.38.: Positive relevance-only mean per feature group with the most relevant feature (V/UV) excluded.

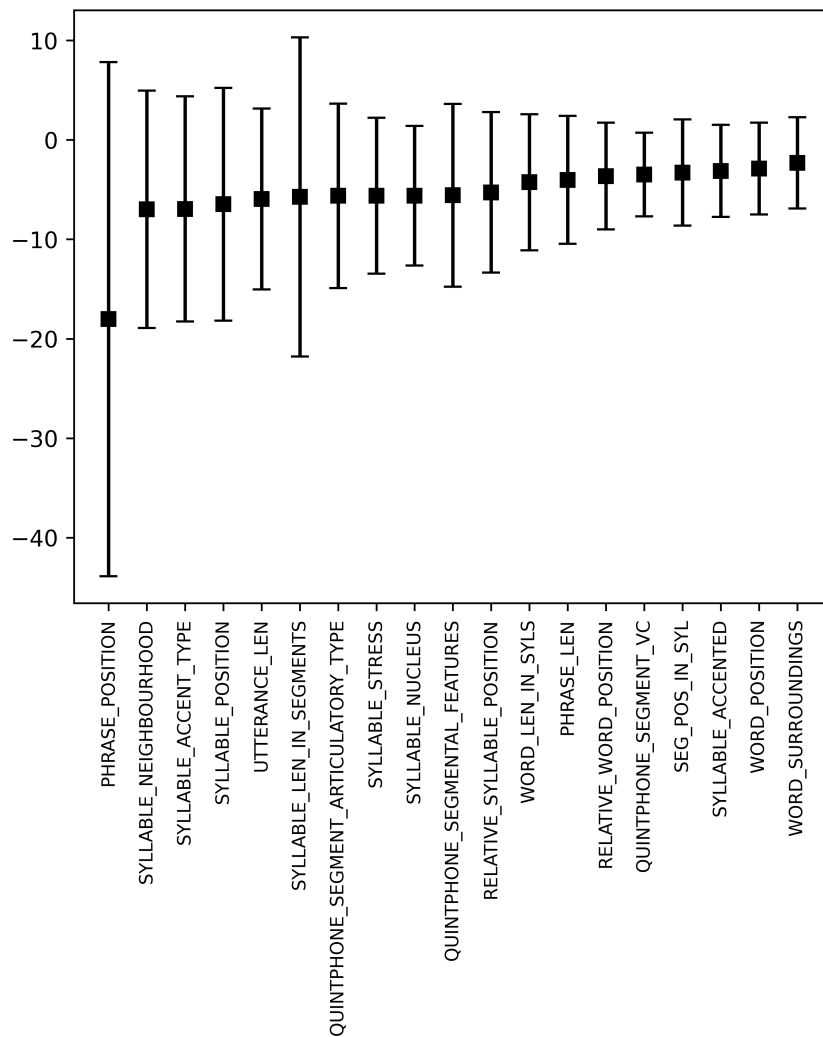


Fig. 4.39.: Negative relevance-only mean per feature group with the most relevant feature (V/UV) excluded.

Discussion and Conclusions

” *Think not of what you see, but what it took to produce what you see.*

— Benoit Mandelbrot

As with most current deep neural architectures, the results produced by the intonation model built in this work are also expectedly good. This can be observed in generally surprisingly adequate contours of the predicted F_0 values, closely tracking the ground truth values. A good overall quality can be generally anticipated in such systems but the actual results depend heavily on the dataset. It is not the quality of the data alone, but also its adequate size. Most of the current state-of-the-art neural models of speech are trained on thousands of hours of recordings, as in the case of Wavenet [239] or the original NSF paper [319]. This allows the model to be trained on a large number of examples and to catch general tendencies much more accurately. The original purpose of the current dataset was to serve as a foundation for a unit selection synthesizer. From this perspective it can be considered quite large as it was designed with high coverage of segmental and suprasegmental contexts of the Polish language and manually labelled by trained phoneticians. However, as Machine Learning dataset it is rather small when compared to the size of the datasets typically used for deep learning. The data is also not free from labelling errors, incorrect segmentation and even some occasional mispronunciations by the speaker himself. The statistical nature of the model, however, allows to assume that with most part of the data being correctly labelled, occasional errors should not influence the performance of the trained model.

On the other hand, it is rather rare to find datasets with such detailed multi-level hand labels, and of sizes comparable to that used in the industry by giant data companies like Alphabet (Google). Usually, they are built using either very simple text annotations, or some automatic labelling methods (or both). It is also becoming increasingly popular to feed the deep neural models with raw data skipping the once obligatory feature extraction and engineering stage altogether. This is because these models, and especially CNNs, are very efficient at discovering and extracting features. As mentioned before, they are able to extract increasingly complex features

from just the raw arrays of pixels linking these to specific categories. The model built in this study can potentially demonstrate similar performance if trained on raw text transcription of the utterances as the features used here can in theory all be derived from the textual input. The advantage of feeding the model raw data is that it can decide on its own what information is actually useful for solving the problem in question and can extract and process that information in the most optimal way. It should be relatively easy for such a deep architecture to extract positional features of segments, words, phrases and even syllables. Maybe it could even come up with a more optimal notion of syllable than that used by the human annotators. Preengineered features, like the ones used in this work, add a layer of encoding, which in some cases might provide the model with just the information that it needs, and in some other cases might mask that information making it harder for the model to discover and extract.

Despite the obvious shortcomings, the current dataset contains features that are easily interpretable and were derived based on our current linguistic theory of language and intonation in the very general sense. This presented the opportunity to build a model that maps from abstract linguistic (and quasilinguistic – as we cannot really relate the number of syllables in the current phrase to any specific linguistic theory or phenomenon) categories and features to concrete physical realizations of the log-fundamental frequency. It was assumed that this mapping modeled within a biologically-inspired neural architecture might be a sufficient approximation of how intonation might be processed in different parts of the human brain and provide a valuable perspective on the mapping between cognitive categories and their physical realizations. Although the source of the biological inspiration is the human (or rhesus monkey) auditory cortex and the model was built for synthesis (so for speech production rather than perception), it was assumed that the general tendencies shown here should be still applicable. When people read speech silently, most will hear the words being *played back* in their heads and these utterances will seem to have distinguishable intonation contours as if somebody was actually reading them aloud¹. This fact might support a number of theories but regardless of which theory might be closer to truth, most will assume that production does somehow relate to the perception in one way or another. Therefore, instead of building a model of intonation production that is close to the actual biological implementation, this work aimed at showing how specific characteristics of the human brain might support the

¹The author was not able to find a proper literature reference to support this statement as it is virtually impossible to devise a scientific experiment that could confirm this mental phenomenon. However, the reader can quite easily observe this through a simple introspection.

emergence of complex categorical (and of course gradual) ideas from a continuous flow of frequency values and the other way around.

In this study, the functional performance of the model as evaluated by objective and subjective measures, was operationalized as an indicator of its adequacy. While this general methodological approach is rather incontrovertible in phonetics, it does suffer from the lack of proper evaluation methods.

The objective measures that are usually used are nothing more than a calculation of some variant of the Mean Squared Error. This method renders insufficient in most problems as it lacks the ability to address some domain-specific degrees of freedom and at the same time does not penalize some of the errors that are more significant than others given the domain. In case of intonation, that same speaker intent can be realized with a range of configurations – a gradient of variants of a single specific contour, or even a number of some completely different contours. Additionally, even intonation contours that seem adequate at the macro level might be perceived as very unnatural if they contain lots of high-frequency variation. Such residual noise is observed in the case of the current model. It might be interpreted as the network's attempt at modeling the microprosody.

In some cases the abrupt shifts in frequency, although rather insignificant in relation to the total frequency range of the speaker, cause unnatural robotic glitches and distortions that seem to have had a much greater influence on the perception of naturalness than the low-frequency contour fit with that of the ground truth. In some other cases they seem to have no particular negative effect of the perceived naturalness.

New objective evaluation metrics are often proposed (e.g.: the tangential distance or the warping method [53]), as this is a significant issue not only for the evaluation of models, but also for the training itself where the metrics are used for the calculation of model inference loss. However, even the state-of-the-art methods are still mostly variants of the idea behind the original MSE. One possible alternative to consider would be a metric based on the autocorrelation algorithm but this solution also has its limitation. Metrics based on measuring the spectral distortion provide a good solution in the general context of evaluating speech synthesizers, they cannot be implemented for the current model of intonation for the obvious reasons. Evaluating spectral distortion in the resynthesized utterances with the predicted F_0 values would mostly reflect the performance of the resynthesis method instead of the F_0 predictions. Therefore, a few variants of the standard MSE metric were used here as a standard solution.

The subjective evaluation of the model was designed in a way that should minimize the effect of the synthesis quality on the listeners and isolate the differences between the predicted and ground truth F_0 values. Initial experimentation with standard resynthesis methods did not yield satisfactory results as it generated a high level of audible spectral distortion in outputs with synthetic F_0 source and an original spectral filter, as opposed to the resynthesis results for the original F_0 and original spectral filter. These initial experiments (not described in this work) led the author to the Neural Source-Filter resynthesis method [319]. The flexibility of the neural vocoder offered much more stability even in case of very unnatural F_0 values and helped address the spectral distortions that could affect the listener. These were significantly reduced, but still present in some cases. On the other hand, some of the utterances resynthesized from all-natural inputs also showed some slight distortions which helped even out the perceived spectral quality of the two categories of stimuli further and eliminate any extra cues that could help the subjects identify synthetic speech. The huge advantages of this resynthesis method are obvious but there is also a disadvantage. The neural vocoder might render slightly different values as compared with the input, whereas the more traditional methods always provide a one-to-one mapping. The changes are usually small and were treated as the necessary compromise in this study. One could also look at the vocoder as being a part of the model corresponding with the actual process of speech production framed within the source-filter theory. It is a rather non-controversial assumption that the articulatory stage of speech production might impose constraints on the realization of the planned utterance and especially its intonation. These can be easily observed in involuntary out-of-tune singing. In the current work, these articulatory or physiological constraints are represented implicitly by the probabilistic model. However, the NSF vocoder introduces a more explicit representation of the influence of the actual articulation on the dynamics of the fundamental frequency.

The problem of missing and superfluous F_0 values (as compared to ground truth), especially when appearing within vowel segments, appeared to decrease the results both in the objective and subjective evaluations. It was decided that missing values will be back- and forward- filled if they appear at the very beginning or end of the utterance. Quadratic interpolation was used for missing values appearing in all other places. Superfluous values were simply deleted. The reason for such a mismatch of values between the ground truth and the predicted vectors output by the model is of purely engineering nature. Different pitch extraction algorithms and implementations were used at different stages as part of the different frameworks used throughout this work as a necessary compromise. In many cases, it was simply

easier to use built-in tools, then to replace them. The discrepancies were minimal and were also treated as necessary.

Given all these distortions and areas for potential improvements, the subjective evaluation, performed with a statistically significant number of subjects using a professionally designed experimental framework, still showed that the null hypothesis (Hypothesis 4.2 in Section 4.2) cannot be rejected with high confidence. This suggests that the resynthesized model predictions appear to have no statistically relevant audible differences in terms of naturalness and that the model was able to effectively predict natural F_0 values in a number of arbitrary and previously unseen inputs given their basic phonological form. **These results were operationalized as an indicator of the model's robustness and provided the necessary support for proving hypotheses 3 and 2 of this thesis, namely that the set of linguistic features used in this thesis contains information which is sufficient for synthesis of natural sounding intonation in the context of statistical-parametric speech synthesis (3), and that the Deep Temporal Convolutional Neural Network used in this thesis can effectively model the mappings between these features and the F_0 contours of an utterance (2).**

The evaluation results also demonstrated the high implementational potential and high pragmatic and functional value of the model in the context of a statistical-parametric speech synthesis of Polish read speech and rendered it potentially useful in other related domains such as modeling of expressive speech. As already outlined in the introductory chapters of this work, a number of such models exist as part of the current state-of-the-art speech synthesis systems. These systems exhibit a range of approaches to modeling intonation with DNNs. They differ both in the specific neural architectures and paradigms used and in the placement of the intonation model within the processing pipeline. Some implement end-to-end models which take textual input and output raw waveforms with an integrated framework where intonation is modeled implicitly. Some on the other hand, introduce dedicated models that handle various levels of speech, including intonation, in separate parallel processing pipelines. In both cases, many different deep neural architectures can also be found, including prototypical TCN's as well as ensemble approaches which include non-convolutional DNN's such as the LSTM. Regardless of the underlying solutions, very rarely are purely linguistic information used as inputs. The architectures are also rarely backed with neurobiological research results. This is due to the fact that these models are built with performance and applicability in mind rather than scientific modeling and explanation of the modeled processes. The current model was also developed in the context of speech synthesis but here linguistic

relevance and explanatory power were of primary interest. Although the choice of explicit linguistic features and network architecture selected merely on the basis of neurobiological evidence, usually yields much worse results for that same problem than some more proven heuristics, it also produces scientific output that is potentially much more valuable to the root discipline of the modeled phenomenon.

As previously noted, deep neural models are not always accepted as scientific models. Although, as shown with extensive argumentation in Section 3.3, the current model fits not only within the pragmatic and statistical-inductive notions of a scientific model, as demonstrated with the results of the evaluation and ensured through formal definition of the model, but also within the unificationist and causal-mechanistic notions. The unificationist interpretation assumes that all levels of intonation processing are relevant, from the organization of abstract psychological concepts in our mind, through the nature of the underlying processing in the human brain and the concrete realizations of the glottal vibrations translated to the frequency of the fundamental frequency and its role as a source in speech production. All of these levels, however not without some necessary idealizations, are addressed in the current framework with a methodology that attempts to stay free of any linguistic dogmas and adopting a physicalist approach that puts the neural aspect of intonation at the center of the problem. Thus, the model also demonstrates strong causal-mechanistic foundations.

Given the above argumentation for the scientificity of the current modeling method, the positive evaluation of the models output naturalness, and with the model architecture being strongly inspired by the computational and structural characteristics of the actual human auditory cortex as discussed in Section 2.2.2, it is also safe to infer that the general neural processing scheme of the human brain might similarly allow the continuous F_0 contours of an utterance to gradually emerge from its discrete phonological (mental) features through a series of successive probabilistic mappings into intermediate latent representations, thereby proving Hypothesis 1 of this work.

The explanatory power is further ensured with the application of the Layer-wise Relevance Propagation explainability algorithm to the results of model inference. The nature of relevance, as defined within that method, also preserves all of the model's formal assumptions, as it is realized within the model itself using the same methods that are used for model training and does not introduce any new operations but simply reverts the operations comprising the model under an additional conservation principle. In this way, the method can be included as part of the model itself.

The model built in this work along with the full source code and methods for running the training, inference, evaluation and explainability analysis along with full results in the form of relevance plots, resynthesis samples and comma-separated data files is openly available at https://github.com/mrslacklines/intonation_synthesis. Additionally, the perceptual evaluation experiment application can still be accessed at <http://fonetyka.cudaniewidy.org/experiment> for reference. The source code of the application is also made freely available at https://github.com/mrslacklines/listening_experiments, and with slight changes can be used to perform arbitrary ABX/MOS experiments.

It was anticipated that the relations between the linguistic inputs and the actual F_0 values might not be straightforward enough, especially given the relatively deep neural architecture, to produce easily interpretable results. This was partly observed in the unprocessed individual results where relevance was generated for the predictions for a single individual utterance sample. With a high number of low-level features and the specificity of context, some of the most relevant features were rather unexpected and hard to interpret, as in the case of features representing specific lengths and positions showing much higher relevance than their counterparts within that same feature class. The surprising aspect was that these specific values represented by these highly relevant features showed no interpretable relation to the actual context in which they appeared. However, the exploding relevance values for the Voiced/Unvoiced feature provided an initial indication of the adequacy and correctness of the applied method. The unexpected values, on the other hand might be useful for the network, as for example, it might have randomly learned to favor a specific positional feature more than the other. For a human, some specific values might appear more meaningful than some others (e.g. 0 and 17, where 0 has that special meaning of being at the beginning or end of something and 17 is *just-a-number*). The network, on the other hand, learns starting from a completely random state and is not initially conditioned in any specific way, unlike humans. Therefore, the unexpected relevance of apparently meaningless features was not questioned.

Additionally, the relevance maps were produced with respect to the output layer of the network only. The inherent feature of the network is to use the trained convolutional filters at intermediate layers to extract latent features of different abstraction levels from combinations of features from previous layers. This might also explain some of the apparently unjustified high relevance values for unexpected features as they might provide valuable input to some of the trained convolutional layers. Although an introspection into these features appearing at those intermediate layers and even visualization of the convolutional filters is possible with the current

methods, they were not included here given the various constraints of the current work.

Under these assumptions, it was decided to create feature groups that would not only withhold the unnecessary, at least at the time of the initial validation of the model, low-level feature characteristics, but also provide a more general insight into which parts of the immediate input is relevant for the model when predicting F_0 values in specific contexts, from the perspective of more relatable higher-level categories. Since it was impossible to assume what kind of granularity and grouping schema would work best, the features were structured into three different groupings. The first one, containing all feature groups, uses the most granular grouping possible. Here, only feature variants differing by quantitative measures were grouped together, so that a feature representing the current word being one word away from the next content word, i.e. `QS Num-Words_from_next-ContWord==1{*+1/F:*}` and a feature representing the current word being two words away from the next content word are included in a single group, but also keeping the information about forward and backward position in separate groups. Similarly features expressing quantitative features like position or length expressed with equality and inequality, as well as features pertaining to the different contexts included (i.e. second neighbour to the left, immediate left neighbour, current, immediate right neighbour, second neighbour to the right) were kept in separate groups. This grouping was supposed to provide some generalization and better readability but keeping the lowest possible level of abstraction. The second grouping scheme included the different contextual variants of the features, as well as variants expressed with different mathematical relations (equal, greater than) in single groups but keeping the forward-backward variants separated. This scheme was meant to provide almost the highest possible level of abstraction given the feature in a way that still allows to reflect the influence of the past and future utterance segments on the current articulation. The last, most general grouping scheme groups all feature variants into the most intuitive categories such as `RELATIVE_WORD_POSITION`. Specific information about group names and their content is available as Appendix B of this work, in the form of a Python script where the structure is defined through a nested list of lists. In this way, a practically infinite number of quantitative analyses can be performed, even on the small holdout set of test utterances. This contributes to building a quantitative methodology that allows to study the relevance of different configurations of linguistic features in predicting specific F_0 contours.

With feature grouping, higher abstraction was achieved and the results became easier to interpret within more familiar categories. Strong trends started appearing

and were easier to relate to some of the known theories and assumptions regarding the linguistic factors and relations on the intonation contour of an utterance. This was achieved through the calculation of different aggregates for each F_0 sample in each individual utterance for all groups. These included mean, sum, positive- and negative-only mean and sum as well as absolute sum and mean, as each of these calculation methods was suspected to reveal different tendencies in the data. Mean, being the most obvious choice, was used to show the regular mean relevance, just generalized for a given group of features. Sum was also included as it was readily available at the time of mean calculation to provide an accumulative measure but turned out to be useless by itself with feature groups containing a non-equal number of features causing the groups containing more features to have naturally higher sum values. Positive- and negative-only means were supposed to provide insight into what kind of features provide the most positive and most negative evidence for given predictions. Because simple calculation of mean was suspected to flatten out features that in some contexts provide strong positive evidence, and in some other strong negative evidence resulting in relevance results similar to some other features that provide little evidence in any context, mean based on absolute sum was also calculated to show which of the features are the most prominent in general. The tendencies demonstrated by the absolute mean were anticipated to be similar to what can be achieved with some of the sensitivity-based explanation methods, which favor features that are used for any specific prediction, regardless of their activation level.

The relevance in individual test samples shows a number of interesting tendencies. First of all, when the features were grouped according to the level of segmentation they relate to, word and syllabic features clearly rendered as the classes that provide the most positive evidence on a regular basis. Closer inspection of syllabic features at a more granular level with additional partitioning based on the type of relation they encode, showed that it was mostly the high relevance of compositional features of syllables, followed by syllable qualitative features and features relating to the absolute position of the syllable that contributed to the high overall relevance in that group. Compositional features group, in case of syllables, include the length of the syllable as expressed in its segmental building blocks (i.e. phonemes). Qualitative features group all information about the presence and type of syllable accent and the presence of stress. It is important to emphasize, that these information are always provided not only for the current syllable but also for its most immediate right and left neighbours, comprising a syllabic equivalent of a triphone.

The above is true also for other segmentation levels. Each segment, regardless of whether this is a syllable, a phoneme, or a phrase, always encodes (through its features) various information about themselves, and additionally that same information on their most immediate neighbours. Segments, additionally encode information on their second most immediate neighbours extending the represented context to a quintphone.

These facts can be interpreted as evidence for the central role of the syllable as a metrical unit and a primary anchor for intonation on the segmental level as usually assumed in most of the phonological models. The length of the syllable determines how a specific F_0 contour will be realized in the time domain, and its absolute position towards the end of the word, phrase and utterance translates to its potential role within the intonational phrase. The lexical stress in Polish is very regular and is placed on the penultimate syllable, which is best represented with the absolute position of the syllable in the enclosing word. Additionally, syllable position towards the end of the phrase and utterance determines its base frequency through its position on the usually descending base frequency line of the whole phrase or utterance.

Interestingly, the relative positional features did not show relevance significantly higher than the mean. Also the features encoding various qualities and quantities of the parent segments, so the word, phrase and the utterance itself, were not ranked very high in case of the syllable. The absolute position group lists features representing the numerical position of the syllable, both forward and backward in their enclosing segments, mainly the word and the phrase, whereas the relative position group includes information about the position of the syllable in relation to their most immediate accented and stressed neighbours.

Features related to (phonological) words also show constantly high positive relevance across the whole utterance in all test samples. This fact also seems to confirm some of the traditional assumptions of most phonological intonation models. Although generally, from the perspective of the relation type, it is the absolute position that appears to be most relevant on average given all features in the feature set, in case of the word-related features, which appear to be the most relevant feature group from the perspective of the linguistic level, it was the relative positional features that appeared to show the highest relevance. Similarly high values were rendered by the features related to the composition of the parent unit. The relative position and the composition of the parent unit represent features that encode the distance to the closest accented (content) words expressed in number of words and the total number of accented words in the enclosing phrase. The absolute position of the word and

its composition were showing significantly lower contribution for the predictions. This might reflect the way the network is encoding the structure of the phonological phrase as in Jassem [154], Demenko [70]) or some equivalent structure that relates the intonational contour of the phrase to its shallow informational structure.

As expected, segmental (i.e. phonemic) features show high relevance in determining the position in the utterance as they provide comparatively the most accurate data, being the lowest most granular segmentation level in the current data. Another interesting observation was made in case of the phrasal features. The information about the composition of the phrase was regularly providing positive evidence for the prediction but at a rather insignificant level if compared to the syllabic features for instance. This group of features includes the number of syllables and words that constitute the phrase. Although this information might seem intuitively important, it is redundant in relation to some of the previously mentioned groups that already encode this information and are much more specific.

More interestingly, the whole group of phrasal features that indicate the position of the phrase in the utterance was shown to provide almost exclusively negative evidence for the predictions along the whole utterance. Very short intervals of rather high relevance may however be observed in regions relating to pauses in the utterance, initial, final but also intermediate. Although it is difficult to interpret that trend it might be some way the model is determining whether a current pause marks the end of a phrase and if the next voiced segment should implement new phrasal intonation contour or not. Mean group relevance results seem to confirm the tendencies discussed above. The Voiced/Unvoiced feature is naturally the single most relevant with other values residing at much lower levels. The information about the syllable accent, length, the relative and absolute position of the word and syllable and current position in the phrase seem to clearly stand out with noticeably high mean relevance whereas the mean relevance of the phrase position appears to be the only feature group with obviously negative mean relevance but also exhibiting some of the highest standard deviations that accounts for the occasional high positive relevance observed in the individual relevance results.

The total mean results also show that the features representing the absolute and relative position of a linguistic unit, its composition and the composition of its enclosing unit are the most relevant and the qualitative features which include mostly information about phoneme identity are the least relevant.

The additional aggregation methods provide additional confirmation. For example the mean of positive-only values shows that the position of the phrase does exhibit

high relevance but only in some specific contexts. It also shows that the feature groups with the highest positive-only relevance mean are also showing the highest standard deviation which might suggest their high contextuality or unreliability. The negative-only relevance mean shows on the other hand that, except for phrase position, some of the feature groups scoring lowest in the mean relevance are also among the ones providing the most negative evidence for F_0 predictions, e.g. the composition of the utterance (so length), type of syllable nucleus and most segmental (phonemic) feature groups. The negative-only mean also shows that feature groups that can be almost universally observed as relevant are also showing negative values in some contexts. These include, amongst other, the relative position or length of the word and the absolute position and accent type of the syllable. In terms of the linguistic level-related grouping, it was the syllabic features that show highest negative-only mean, where the word-related features showing the lowest (so closest to zero). The relations with the least negative evidence where the groups including the compositional and relative position features, and the absolute position was the group with the highest negative-only mean relevance. All these observations may suggest a number of possible factors. First of all, the nature of the model is probabilistic and it is natural that it accounts for some unreliability and variance in evidence provided by some features. This might also be indirectly taken as a cue for the probabilistic nature of prosody altogether. Another interpretation is, as already mentioned, that some features are strictly contextual or categorical and that they only provide positive evidence in some specific contexts which can be determined by their dependence on evidence provided by other features. The absolute mean-based analysis rendered the position of the phrase as the most prominent feature (after V/UV). It is followed by the syllable accent type, syllable length and the absolute position of the syllable with rapidly decreasing absolute mean relevance in each of these categories. Most other feature groups show similar values except for the absolute word position, length of phrase, syllable nucleus type and information whether the segment is a vowel or a consonant, which exhibit an apparent decrease in prominence.

The inventory of individual features used in this work is rather wide and necessarily redundant. This results in a large number of degrees of freedom within which model training can effectively converge. The number of potential model parameter configurations and, hence, features that become most relevant for the many versions of the models trained with the current framework is practically infinite. However, the general trends presented here should hold even across many different instances of the model trained within the current framework, especially at the higher levels of abstraction among more general feature groups.

Although the quantitative results present high value on their own it is necessary to consider their theorygenic potential. The above discussion clearly shows that it is possible to isolate a group (or rather a group of groups) of features that are significantly more prominent in terms of relevance than the others. Of course, the line has to be drawn arbitrarily in some cases as the differences of mean results between neighbouring groups are often minimal. Nonetheless, the most significant information for the contours of F_0 are:

i Word

- a) Position of each accented word relative to other accented words in the current phrase

ii Syllable

- a) Presence of accent
- b) Accent type, if present (ToBI)
- c) Length, expressed in number of segments

And to a lesser extent:

iii Word

- a) Length, expressed in number of syllables
- b) Absolute position in phrase

iv Syllable

- a) Absolute position in the enclosing units

v Segment

- a) Absolute position in the enclosing units

Many other features and feature groups demonstrate steady positive relevance and although their mean value is much lower than that of the features enumerated above, their actual contribution to the accuracy of the final predictions might be high. The segmental features, like phoneme identity, are a good example. Although Polish is traditionally said to be syllable-timed, the current data might suggest that the syllable duration might not be that regular and might play a role in how the contour of intonation is articulated. A syllable comprising 3 phonemes (so with length equal 3), might have a slightly different total duration, or at least might provide a shorter interval for the realization of an intonational event, if these phonemes are a plosive,

a fricative and a short vowel like schwa than if those phonemes are two vowels and a nasal. This information does influence the shape of the predicted F_0 contour significantly as it changes the shape of any intonational event realized during the articulation of that syllable. On the other hand, given the inherent characteristics of the TCN network, this information can also be inferred just from the features listed above, the past inputs the network has already processed and the rate of change of the features in (i- iv), as well as any other latent features extracted as a result of many consecutive convolutions. The memory of the network and its causal nature as well as its ability to perform many complex transformations of the input, which resembles the properties of the human sensory cortices, allows to assume that the features showing constant positive, but not very prominent, relevance might be used as a source of confirmation of some of the information the network was able to infer on its own. These are, of course, only wishful speculations, but justified ones nonetheless.

If the above, preliminary assumptions can be further refined and confirmed with revised experiments, it should allow us to suspect that the phonological transcription of any (non-expressive and neutrally read) utterance can be limited to just these information and that it can be used to effectively reproduce a natural sounding neutral intonation for that utterance.

It is also possible that the syllable accent type, the ToBI tone, could be omitted and the actual type of tonal accent could be inferred by the network from context as humans usually do when reading an unseen sentence. However, humans have access to a much wider range of linguistic information that are not included in the current data, such as some representation of the grammatical relations in the sentence, including deep and shallow semantic dependencies between the words. On the other hand, some of the CNN-based end-to-end TTS systems have recently demonstrated that it is possible to synthesize human-like speech from just raw text. This suggests that all these relations can be extracted and modeled implicitly by the network. Humans, also rarely have access to all of the words that appear in the sentence being read, and usually seem to make out the general sense of the utterance on the fly. The resulting intonation however, always turns out natural.

The above discussion and interpretation touches only the most pronounced of the tendencies observed in the data but the actual amount of data generated in the current study is much greater and should constitute a valuable source of linguistic inquiry. This fact along with the additional argumentation for adopting the current Deep Temporal Convolutional Neural Network as a scientific model of intonation provided in Section 3.3 confirms that a Deep

Temporal Convolutional Network can become an explanatory scientific model of mappings between linguistic features and the intonation of an utterance, also proving the contributory methodological Hypothesis 4 of this work.

All of the speculations presented here naturally require further validation through other experiments and revisions of the current framework, as discussed in the next chapter.

Future Plans and Challenges

” *Set thy heart upon thy work, but never on its reward.*

— Ved Vyasa
The Bhagavad Gita

The huge potential of the current methodology should naturally stimulate further research and revisions of the method itself as this work contains a number of inevitable compromises. The single most important of them is the size of the dataset used in this study and the most important improvements can be introduced in this area. Any following experiments should focus on implementing a specially designed training and evaluation material. This is not an easy task and the design of linguistic resources is a complicated problem on its own. However, given that these networks can learn to generate perfectly sounding speech from just the raw text and that most of the relevant linguistic features identified in this study can be extracted straight from the textual input with some simple text processing, it might be actually easy to obtain a much bigger, representative dataset. It could for instance be semi-automatically downloaded from the numerous internet services that provide millions of hours of audio and video streams with time-aligned subtitles. These services allow to search and retrieve specific content that can be easily limited for the needs of any domain of research, with neutral read speech included. In this way, a dataset comprising hundreds or even thousands of hours of speech can be obtained almost effortlessly as compared to the costly and time-consuming procedure of professional audio recording sessions and labelling. Including multiple speakers in the dataset would also introduce an important improvement and a yet another level of the needed generalization for the model.

On the other hand, the current dataset proved effective but was used as-is and the results could probably be further improved with some basic data work. That could include hand inspection of some of the labels, especially the ToBI tones that were converted automatically from the original transcriptions. Also the feature set should be checked for correlations between specific features to identify any redundant information that might distort the final interpretation of the results.

In the current study the model is a result of a single training session. Training multiple instances of the same model architecture within many training sessions could also provide some valuable insight into the variability of the most relevant features. In the current work, some of the most relevant among the individual features seemed as very unintuitive choices, e.g. *if there is 13 syllables towards the end of the phrase*. With multiple training sessions, it could be validated if the network randomly converged to use that specific feature in just that single session or if it always converges in a similar manner. The latter could suggest that there might be something special about that particular distance of 13 syllables that might need further investigation. The additional training sessions should also include at least some form of hyperparameter optimization. The number of layers in the network, as well as other parameters, such as the number and size of the convolutional filters, and the number of dilations were all chosen arbitrarily in this work. It would be beneficial to determine the most optimal configurations for these parameters that allow the network to reach its highest potential for the presented problem. A number of methods exist for performing such hyperparameter optimizations, e.g. Amazon Web Services SageMaker which was used as the cloud infrastructure provider for training the network provides such capabilities out-of-the-box. A number of Python libraries also exist, with TPOT¹ [193] being the most notable one. The mentioned library allows to easily plug virtually any modeling pipeline into an evolutionary optimization framework. On the other hand, this work focuses mainly on the linguistic aspect of the model, and although any improvements in the accuracy would improve the overall power of the model, it is only necessary to ensure model parameters are not overly constrictive. The pruning of model parameters that aims at improving the size of the model in memory and its computational complexity is much more of an engineering problem.

Any improvements to the training process, however, would be of little effect if the actual loss function is less than ideal. The current work uses the Mean Squared Error as the metric used for evaluating model predictions in relation to the expected values. This mathematical function is the de facto standard in most research related to the prediction or approximation of the fundamental frequency contour in speech. However, even the results of the current work alone, indicate a number of areas in which this specific metric fails to represent the actual target quality for the model. The ideal function should be able to mathematically encode all aspects that make the predicted intonation natural for a given input. Such a metric obviously does not exist as defining naturalness itself lies at the very heart of the whole problem of intonation modeling. Nonetheless, many improvements can be introduced even

¹<https://github.com/EpistasisLab/tpot>

to the MSE-based loss metrics. The current results show lots of high frequency variability in the predicted contours which is associated with some glitches in the resynthesized speech. Although that residual noise could be flattened out with some smoothing algorithms, it would probably also remove some of the much needed microprosody. A better method would include a compound loss metric that separates error calculation for the high and low frequency components of the signal. Such a decomposition would make the model aim to better represent each component on its own. Similarly the number of components to extract and how they are used in the calculation of the final loss could be a subject to optimization. In some works on intonation [226], the signal is separated into a phrasal component and accent component with different speaker-specific bandpass filters. Such an approach could also provide a valuable look at some of the superpositional theories of intonation (see Section 2.3). Some of the superpositional models, like the Fujisaki model [104], could even be included in the model itself as a separate processing stage that accounts for the influence of the biomechanics of the articulatory tract on the final contour of the intonation, whereas the neural network could be used to account only for the processing that takes place in the *brain*. The network could learn to generate such F_0 values that best serve as targets for approximating them with the Fujisaki model (with the Mixdorff [226] method for instance). In this case the loss would be calculated for the approximated contour. That would be a step towards better or more appropriate encoding of the articulatory or motoric information in the neural structures of the model.

The objective evaluation results were initially planned to be factored in as additional weights to the final relevance results of features. It was assumed that it is not sufficient to measure the relevance of a feature for any given prediction but it is important to measure its relevance given the accuracy of the prediction. However, the scope and limitations of this work caused a number of these ideas to be regrettably postponed.

Many improvements can also be made in terms of better representation of the actual neural processing of intonation during production in the model architecture. The current work is based only on some very general ideas of how the auditory cortex might process information. Although many neurobiological research papers are referenced in the introductory chapters, it would be beneficial to construct the model based on a single coherent neural theory and provide a better justification of each architectural features of the neural network as in the current work they are rather loosely and selectively interpreted. Instead of dealing with production, a recognition model could be built using a similar TCN architecture. Such a model could be used

to study the acoustic cues that the model learns to detect when recognizing Chinese lexical tones, for instance. This work was focused on building a single general model that allows to study the mapping between a wide range of linguistic categories and the resulting F_0 values through analysis by synthesis, but many other approaches are viable. Even the exact same modeling framework could be used to conduct a number of various experiments with the inclusion of linguistic features from other levels of language such as semantics, syntax, pragmatics or specific social contexts (so emotions), as these must also be computed in when reading sentences, but are much harder to represent as data labels. Möbius [228], for example, has shown that “nouns require higher amplitudes than other classes”.

Although the explanation methods are a relatively new problem domain that appeared only a few years ago with Layer-wise Relevance Propagation being one of the most recent algorithms, a number of improvements and good practices has already been introduced in [178]. The Zero-variant of the LRP algorithm used here is a default go-to that should demonstrate good results in most cases. After initial validation, which this work hopefully provides, the algorithm should be fine-tuned and a best fitting variant of the method should be selected as the potential improvements in the results are huge as demonstrated in Kohlbrenner et al. [178].

The verification of relevance results is also needed. One interesting way, that could present a lot of potential engineering value, would be to perform an ablation study given the relevance of features. The least relevant features could be systematically removed from the training data and after retraining the model on that pruned dataset the naturalness of the resynthesized speech could be again evaluated to confirm if the model was able to converge to a similar quality without the features that were identified as insignificant for the predictions. This would provide a method of validating the results of relevance analysis but could also help drastically limit the footprint and computational complexity of the model. In practical applications, that would translate to much shorter training time and much lower hardware requirements (so lower costs). It would also help reduce the time and effort needed for composing training datasets if it was confirmed which features constitute the minimum inventory required for training a high-quality model. Optimizations can also be introduced in the implementation itself. The current solution was rapidly prototyped and should definitely be refactored before new functionalities are added on top of it.

The subjective evaluation of the end results also shows many areas for potential improvements. It is rather hard to devise such a method that allows to only evaluate the naturalness of the intonation, without the influence of the spectral filter, in the

context of speech synthesis. The current method could probably use some noise or cocktail-party effect masking of the stimulus to further reduce the influence of the resynthesis itself on the perceived naturalness. The effect of the spectral distortion introduced by the resynthesis method could also be measured using one of the spectral distortion metrics and factored into the final results. Another way would be to abandon the resynthesis altogether and perform a purely psychoacoustic experiment with a synthesized fundamental frequency tone alone although this would not address the problem of intonation perception in speech, but would move the problem into the more general domain of pitch processing which could provide more insight into some processes but also considerably limit it for some other.

The last but probably the most significant piece of work that is planned for future is the analysis of the latent features and the actual filters that the network learned. The current work presents only the relevance of the input feature vector with respect to the final output of the network. Although the current results demonstrate which input features are relevant for specific predictions of the fundamental frequency they do not explain how they are relevant. The multiple hidden layers perform the convolution operation on the output of the preceding layers. The analysis of the actual convolutional filters and an insight into the emergence of the intermediate latent features would help understand how that input data is processed to convert high-level discrete categories into a flow of continuous values of F_0 . A number of methods could be used for this purpose including the original LRP used here. LRP can be applied to any of the network's layers but the amount of data that would be generated in case of analysing all of the hidden layers would greatly exceed the limits of this work. Nonetheless, it is a very promising direction of future research that the author wishes to undertake sometime in the near future.

Bibliography

- [1] Abadi, M., Barham, P., Chen, J., et al. “TensorFlow: A system for large-scale machine learning”. In: *12th USENIX Symposium on operating systems design and implementation (OSDI 16)*. 2016, pp. 265–283. URL: <https://www.usenix.org/system/files/conference/osdi16/osdi16-abadi.pdf> (cit. on pp. 9, 73).
- [2] Abdel-Hamid, O., Mohamed, A.-R., Jiang, H., et al. “Convolutional neural networks for speech recognition”. In: *IEEE/ACM Transactions on audio, speech, and language processing* 22.10 (2014), pp. 1533–1545 (cit. on p. 31).
- [3] Abe, I. “Intonational patterns of English and Japanese”. In: *Word* 11.3 (1955), pp. 386–398 (cit. on p. 13).
- [4] Achinstein, P. “Concepts of science: A philosophical analysis”. In: (1968) (cit. on p. 58).
- [5] Akansu, A. N. and Liu, Y. “On-signal decomposition techniques”. In: *Optical engineering* 30.7 (1991), pp. 912–921 (cit. on p. 32).
- [6] Alber, M., Lapuschkin, S., Seegerer, P., et al. “iNNvestigate neural networks!” In: *Journal of Machine Learning Research* 20.93 (2019), pp. 1–8. URL: <http://jmlr.org/papers/v20/18-540.html> (cit. on pp. 9, 40, 78).
- [7] Allen, J., Hunnicutt, M. S., Klatt, D. H., Armstrong, R. C., and Pisoni, D. B. *From text to speech: The MITalk system*. Cambridge University Press, 1987 (cit. on p. 17).
- [8] Alonso, J.-M. and Chen, Y. “Receptive field”. In: *Scholarpedia* 4.1 (2009), p. 5393 (cit. on p. 25).
- [9] Anderson, M., Pierrehumbert, J., and Liberman, M. “Synthesis by rule of English intonation patterns”. In: *ICASSP’84. IEEE International Conference on Acoustics, Speech, and Signal Processing*. Vol. 9. IEEE. 1984, pp. 77–80 (cit. on p. 52).
- [10] Angrick, M., Herff, C., Johnson, G. D., et al. “Interpretation of convolutional neural networks for speech regression from electrocorticography.” In: *ESANN*. 2018 (cit. on p. 35).
- [11] Angrick, M., Herff, C., Mugler, E., et al. “Speech synthesis from ECoG using densely connected 3D convolutional neural networks”. In: *Journal of neural engineering* 16.3 (2019), p. 036019 (cit. on p. 35).
- [12] Anumanchipalli, G. K., Prahallad, K., and Black, A. W. “Festvox: Tools for creation and analyses of large speech corpora”. In: *Workshop on Very Large Scale Phonetics Research, UPenn, Philadelphia*. 2011, p. 70 (cit. on p. 65).
- [13] Arbabzadah, F., Montavon, G., Müller, K.-R., and Samek, W. “Identifying Individual Facial Expressions by Deconstructing a Neural Network”. In: vol. 9796. Aug. 2016, pp. 344–354 (cit. on p. 40).
- [14] Arik, S. O., Chrzanowski, M., Coates, A., et al. “Deep voice: Real-time neural text-to-speech”. In: *arXiv preprint arXiv:1702.07825* (2017) (cit. on pp. 35, 56).

- [15] Arnold, G. F. and O'Connor, J. *Intonation of colloquial English*. Longman, London, 1973 (cit. on pp. 13, 45, 64).
- [16] Arras, L., Horn, F., Montavon, G., Müller, K.-R., and Samek, W. "Explaining predictions of non-linear classifiers in NLP". In: June 2016, pp. 1–7 (cit. on p. 40).
- [17] Bach, S., Binder, A., Montavon, G., et al. "On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation". In: *PloS one* 10.7 (2015), e0130140 (cit. on pp. 8, 37, 38, 40).
- [18] Baehrens, D., Schroeter, T., Harmeling, S., et al. "How to explain individual classification decisions". In: *The journal of machine learning research* 11 (2010), pp. 1803–1831 (cit. on p. 37).
- [19] Bai, S., Kolter, J. Z., and Koltun, V. "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling". In: *arXiv preprint arXiv:1803.01271* (2018) (cit. on pp. 33, 73).
- [20] Bai, S., Kolter, J. Z., and Koltun, V. "Convolutional sequence modeling revisited". In: (2018) (cit. on pp. 33, 73).
- [21] Bannert, R. "Variations in the perceptual modelling of macro-prosodic organization of spoken Swedish: prominence and chunking". In: *Reports from the Department of Phonetics, Umeå University, Phonum* 3 (1995), pp. 31–53 (cit. on p. 43).
- [22] Barber S. Carlson R., C. D. G. V. "Speech synthesis experiments with the GLOVE synthesizer". In: *Proceedings of Eurospeech 93*. Vol. 2. 1993, pp. 925–928 (cit. on p. 17).
- [23] Batliner, A. and Möbius, B. "Prosodic models, automatic speech understanding, and speech synthesis: Towards the common ground?" In: *The integration of phonetic knowledge in speech technology*. Springer, 2005, pp. 21–44 (cit. on p. 52).
- [24] Beckman, M. E. "The parsing of prosody". In: *Language and cognitive processes* 11.1-2 (1996), pp. 17–68 (cit. on pp. 28, 45).
- [25] Bellenger, D., Bertram, J., Budina, A., et al. "Scaling in cloud environments". In: *Recent researches in computer science* 33 (2011), pp. 145–150 (cit. on p. 82).
- [26] Bengio, Y., Simard, P., and Frasconi, P. "Learning long-term dependencies with gradient descent is difficult". In: *IEEE transactions on neural networks* 5.2 (1994), pp. 157–166 (cit. on p. 21).
- [27] Bernardo, J. M. and Smith, A. F. M. *Bayesian theory*. John Wiley & Sons, 1994 (cit. on p. 59).
- [28] Bertinetto, P. M. *Aspetti prosodici della lingua italiana*. Clesp, 1979 (cit. on p. 41).
- [29] Bharadhwaj, H. "Layer-wise relevance propagation for explainable deep learning based speech recognition". In: *2018 IEEE International symposium on signal processing and information technology (ISSPIT)*. IEEE. 2018, pp. 168–174 (cit. on p. 4).
- [30] Bierwisch, M. "Regeln für die Intonation deutscher Sätze". In: *Studia grammatica* 7 (1966), pp. 99–201 (cit. on p. 13).
- [31] Black, A., Taylor, P., Caley, R., and Clark, R. *The Festival speech synthesis system*. 1998 (cit. on pp. 17, 65).

- [32] Black, A. W. and Lenzo, K. *Building voices in the Festival speech synthesis system*. 2000 (cit. on p. 65).
- [33] Bloch, B. and Trager, G. L. “The syllabic phonemes of English”. In: *Language* 17.3 (1941) (cit. on p. 14).
- [34] Bloomfield, L. *Language*. 1935 (cit. on p. 44).
- [35] Bolinger, D. L. “A theory of pitch accent in English”. In: *Word* 14.2-3 (1958), pp. 109–149 (cit. on pp. 13, 45).
- [36] Bolinger, D. L. “Intonation: levels versus configurations”. In: *Word* 7.3 (1951), pp. 199–210 (cit. on pp. 13, 45).
- [37] Botinis, A. *Stress and prosodic structure in Greek*. Lund University Press, Lund, 1989 (cit. on p. 43).
- [38] Botinis, A., Granström, B., and Möbius, B. “Developments and paradigms in intonation research”. In: *Speech communication* 33.4 (2001), pp. 263–296 (cit. on p. 41).
- [39] Bottou, L. “Online learning and stochastic approximations”. In: *On-line learning in neural networks* 17.9 (1998), p. 142 (cit. on p. 76).
- [40] Breuer, S. and Hess, W. “The Bonn open synthesis system 3”. In: *International journal of speech technology* 13.2 (2010), pp. 75–84 (cit. on p. 64).
- [41] Carlson, R. Granström, B., H. “A multi-language text-to-speech module”. In: *Proc. of the 7th International Conference on Acoustics, Speech, and Signal Processing (ICASSP’82)*. Vol. 3. 1982, pp. 1604–1607 (cit. on p. 17).
- [42] Chang, N.-c. T. “Tones and intonation in the Chengtu dialect (Szechuan, China)”. In: *Phonetica* 2.1-2 (1958), pp. 59–85 (cit. on p. 13).
- [43] Chao, Y. R. “Tone and intonation in Chinese”. In: *Bulletin of the National Research Institute of History and Philology of the Academia Sinica* 4 (1933), pp. 121–134 (cit. on p. 13).
- [44] Chellapilla, K., Puri, S., and Simard, P. “High performance convolutional neural networks for document processing”. In: *Tenth International Workshop on Frontiers in Handwriting Recognition*. Suvisoft. 2006 (cit. on p. 28).
- [45] Cheung, C., Hamilton, L. S., Johnson, K., and Chang, E. F. “The auditory representation of speech sounds in human motor cortex”. In: *eLife* 5 (2016), e12577 (cit. on p. 31).
- [46] Cho, K., Merriënboer, B. van, Gulcehre, C., et al. “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation”. In: (June 2014) (cit. on p. 23).
- [47] Chollet, F. et al. *Keras*. <https://github.com/fchollet/keras>. accessed on 02 Feb 2021 (cit. on pp. 9, 73).
- [48] Chomsky, N. *Syntactic Structures*. Janua linguarum (Mouton, Paris).: Series Minor. Mouton, 1957 (cit. on pp. 2, 15).
- [49] Chomsky, N. “Current trends in linguistic theory”. In: *The Structure of Language*, New York: Prentice-Hall (1964), pp. 50–118 (cit. on p. 61).

- [50] Chomsky, N. “The Logical Basis of Linguistic Theory. I: Preprints of Papers for the Ninth International Congress for Linguists”. In: *Cambridge, Mass. S* (1962), pp. 509–574 (cit. on p. 45).
- [51] Cichy, R. M. and Kaiser, D. “Deep neural networks as scientific models”. In: *Trends in cognitive sciences* 23.4 (2019), pp. 305–317 (cit. on p. 58).
- [52] Cireşan, D. C., Meier, U., Masci, J., Gambardella, L. M., and Schmidhuber, J. “Flexible, High Performance Convolutional Neural Networks for Image Classification”. In: *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence. IJCAI’11*. Barcelona, Catalonia, Spain: AAAI Press, 2011, pp. 1237–1242 (cit. on p. 23).
- [53] Clark, R. A. J. and Dusterhoff, K. E. “Objective methods for evaluating synthetic intonation”. In: *Sixth European Conference on Speech Communication and Technology*. 1999 (cit. on p. 119).
- [54] Cohen, A., Hart, J., et al. “Perceptual analysis of intonation patterns”. In: *5e Congres international d’acoustique, Liège 7-14 septembre, 1965*. 1965 (cit. on p. 14).
- [55] Collier, R. et al. “The role of intonation in speech perception”. In: *Structure and process in speech perception*. Springer, 1975, pp. 107–123 (cit. on p. 46).
- [56] Cooney, C., Korik, A., Raffaella, F., and Coyle, D. “Classification of imagined spoken word-pairs using convolutional neural networks”. In: *The 8th Graz BCI Conference, 2019*. Verlag der Technischen Universitat Graz. 2019, pp. 338–343 (cit. on p. 35).
- [57] Cox, D. R. and Barndorff-Nielsen, O. E. *Inference and asymptotics*. Vol. 52. CRC Press, 1994 (cit. on p. 59).
- [58] Cox, D. R. and Hinkley, D. V. *Theoretical statistics*. CRC Press, 1979 (cit. on p. 59).
- [59] Cruttenden, A. “The origins of nucleus”. In: *Journal of the International Phonetic Association* 20.1 (1990), pp. 1–9 (cit. on p. 41).
- [60] Crystal, D. *Prosodic systems and intonation in English*. Vol. 1. CUP Archive, 1969 (cit. on pp. 45, 64).
- [61] Crystal, D. and Quirk, R. *Systems of prosodic and paralinguistic features in English*. 39. Walter De Gruyter Inc., 1964 (cit. on p. 13).
- [62] Cutler, A., Dahan, D., and Van Donselaar, W. “Prosody in the comprehension of spoken language: A literature review”. In: *Language and speech* 40.2 (1997), pp. 141–201 (cit. on p. 41).
- [63] Cutler, A. and Ladd, D. R. *Prosody: Models and measurements*. Vol. 14. Springer-Verlag Berlin Heidelberg, 1983 (cit. on p. 14).
- [64] Daneš, F. “Sentence intonation from a functional point of view”. In: *Word* 16.1 (1960), pp. 34–54 (cit. on pp. 13, 43).
- [65] de Heer, W. A., Huth, A. G., Griffiths, T. L., Gallant, J. L., and Theunissen, F. E. “The hierarchical cortical organization of human speech processing”. In: *Journal of Neuroscience* 37.27 (2017), pp. 6539–6557 (cit. on pp. 30, 31).
- [66] De Leeuw, J. R. “jsPsych: A JavaScript library for creating behavioral experiments in a web browser”. In: *Behavior research methods* 47.1 (2015), pp. 1–12 (cit. on p. 81).
- [67] De Saussure, F. “Course in general linguistics (trans. Roy Harris)”. In: *London: Duckworth* (1916), p. 14 (cit. on p. 42).

- [68] Delattre, P., Olsen, C., and Poenack, E. "A comparative study of declarative intonation in American English and Spanish". In: *Hispania* 45.2 (1962), pp. 233–241 (cit. on p. 14).
- [69] Demenko, G., Möbius, B., and Klessa, K. "Implementation of Polish speech synthesis for the BOSS system". In: *Bulletin of the Polish Academy of Sciences. Technical Sciences* 58.3 (2010), pp. 371–376 (cit. on pp. 7, 61).
- [70] Demenko, G. *Analiza cech suprasegmentalnych języka polskiego na potrzeby technologii mowy*. Wydawnictwo Naukowe UAM, 1999 (cit. on pp. 46–48, 127).
- [71] Demenko, G. "Intonation processing for speech technology". In: *Speech and Language Technology* 14/15 (2012) (cit. on p. 49).
- [72] Demenko, G., Bachan, J., Möbius, B., et al. "Development and evaluation of Polish speech corpus for unit selection speech synthesis systems". In: *Ninth Annual Conference of the International Speech Communication Association*. 2008 (cit. on pp. 61, 62).
- [73] Demenko, G., Klessa, K., Szymański, M., Breuer, S., and Hess, W. "Polish unit selection speech synthesis with BOSS: extensions and speech corpora". In: *International journal of speech technology* 13.2 (2010), pp. 85–99 (cit. on p. 61).
- [74] Demenko, G. and Kuczmarski, T. "Intonation modeling for neutral-style speech synthesis". In: *Speech and Language Technology* 14 (), p. 17 (cit. on p. 47).
- [75] Demenko, G., Szymański, M., Cecko, R., et al. "Development of Large Vocabulary Continuous Speech Recognition using phonetically structured speech corpus". In: *ICPhS*. 2011, pp. 568–571 (cit. on p. 48).
- [76] Demenko, G. and Wagner, A. "Prosody annotation for unit selection TTS synthesis". In: *Archives of acoustics* 32.1 (2007), pp. 25–40 (cit. on p. 61).
- [77] Demenko, G., Wagner, A., and Cylwik, N. "The use of speech technology in foreign language pronunciation training". In: *Archives of Acoustics* 35.3 (2013) (cit. on p. 48).
- [78] Demenko, G., Wypych, M., and Baranowska, E. "Implementation of grapheme-to-phoneme rules and extended SAMPA alphabet in Polish text-to-speech synthesis". In: *Speech and Language Technology* 7.17 (2003), pp. 79–97 (cit. on p. 64).
- [79] Denes, P. "A Preliminary Investigation of Certain Aspects of Intonation". In: *Language and Speech* 2 (Apr. 1959), pp. 106–122 (cit. on p. 14).
- [80] Deng, L. and Yu, D. "Deep learning: methods and applications". In: *Foundations and trends in signal processing* 7.3–4 (2014), pp. 197–387 (cit. on p. 28).
- [81] Di Cristo, A. "Intonation in French". In: *Intonation systems: A survey of twenty languages* (1998), pp. 195–218 (cit. on p. 41).
- [82] Di Cristo, A. *Soixante et dix ans de recherches en prosodie*. Vol. 1. Publ. de l'Université de Provence, 1975 (cit. on p. 41).
- [83] Di Liberto, G. M., O'Sullivan, J. A., and Lalor, E. C. "Low-frequency cortical entrainment to speech reflects phoneme-level processing". In: *Current Biology* 25.19 (2015), pp. 2457–2465 (cit. on p. 31).
- [84] Dłuska, M. *Akcent i aтона w języku polskim*. 1957 (cit. on p. 46).

- [85] Dłuska, M. *Prosodia języka polskiego*. Nakł. Polskiej Akademii Umiejętności, 1947 (cit. on p. 46).
- [86] Dusterhoff, K. E. and Black, A. W. "Generating f0 contours for speech synthesis using the tilt intonation theory." In: (1997) (cit. on p. 52).
- [87] *ECESS: European Center of Excellence on Speech Synthesis Web Page*. <http://www.eccess.eu/>. accessed on 13 Apr 2008 (cit. on p. 62).
- [88] Elsner, J. "Rozprawa o metryczności i rytmiczności języka polskiego". In: *Cz. I, Warszawa* (1818), pp. 51–52 (cit. on p. 46).
- [89] Emmeche, C., Køppe, S., and Stjernfelt, F. "Explaining emergence: towards an ontology of levels". In: *Journal for general philosophy of science* 28.1 (1997), pp. 83–117 (cit. on p. 2).
- [90] Esling, J. H. and Gaylord, H. "Computer codes for phonetic symbols". In: *Journal of the International Phonetic Association* 23.2 (1993), pp. 83–97 (cit. on p. 64).
- [91] Ethofer, T., Anders, S., Erb, M., et al. "Cerebral pathways in processing of affective prosody: a dynamic causal modeling study". In: *Neuroimage* 30.2 (2006), pp. 580–587 (cit. on p. 29).
- [92] Fant, G. and Martony, J. "Speech synthesis instrumentation for parametric synthesis (OVE II)". In: *Speech Transmission Laboratory Quarterly Progress and Status Report (KTH)* 2 (1962), pp. 18–24 (cit. on p. 17).
- [93] Faure, G. *Recherches sur les caractères et le rôle des éléments musicaux dans la prononciation anglaise: Études anglaises*. Didier (Paris), 1962 (cit. on p. 13).
- [94] Féry, C. *Intonation and Prosodic Structure*. Key Topics in Phonology. Cambridge University Press, 2016 (cit. on p. 41).
- [95] Flanagan, J., Ishizaka, K., and Shipley, K. "Synthesis of Speech From a Dynamic Model of the Vocal Cords and Vocal Tract". In: *The Bell System Technical Journal* 54 (Mar. 1975) (cit. on p. 16).
- [96] Folk, M., Heber, G., Koziol, Q., Pourmal, E., and Robinson, D. "An overview of the HDF5 technology suite and its applications". In: *Proceedings of the EDBT/ICDT 2011 Workshop on Array Databases*. 2011, pp. 36–47 (cit. on p. 77).
- [97] Fonagy, I. and Magdics, K. *L'intonation du Hongrois*. 1967 (cit. on p. 13).
- [98] Fonagy, I. "Accent et intonation du français moderne". In: *Grammaire Descriptive du Français Moderne*. 1952, pp. 62–82 (cit. on p. 13).
- [99] Francuzik, K., Karpiński, M., and Kleśta, J. "A preliminary study of the intonational phrase, nuclear melody and pauses in Polish semi-spontaneous narration". In: *Speech Prosody 2002, International Conference*. 2002 (cit. on p. 47).
- [100] Friedman, M. "Explanation and scientific understanding". In: *The Journal of Philosophy* 71.1 (1974), pp. 5–19 (cit. on p. 58).
- [101] Fry, D. "Linguistic theory and experimental research". In: *Transactions of the Philological Society* 59.1 (1960), pp. 13–39 (cit. on pp. 14, 15).
- [102] Fry, D. B. "The Present-day Tasks of the Phonetic Sciences". In: *Proc. 6th ICPhS*. Prague, Czechoslovakia, 1967, pp. 87–89 (cit. on pp. 13, 14).

- [103] Fujisaki, H. "Dynamic characteristics of voice fundamental frequency in speech and singing. acoustical analysis and physiological interpretations". In: *Dept. for Speech, Music and Hearing, Tech. Rep* (1981) (cit. on pp. 43, 47, 52).
- [104] Fujisaki, H. and Kawai, H. "Realization of linguistic information in the voice fundamental frequency contour of the spoken Japanese". In: *ICASSP-88., International Conference on Acoustics, Speech, and Signal Processing*. IEEE Computer Society. 1988, pp. 663–664 (cit. on p. 135).
- [105] Fukushima, K. "Neocognitron: A hierarchical neural network capable of visual pattern recognition". In: *Neural Networks 1.2* (1988), pp. 119–130. URL: <http://www.sciencedirect.com/science/article/pii/0893608088900147> (cit. on pp. 3, 23, 25).
- [106] Gårding, E. "On parameters and principles in intonation analysis". In: *Journal ISSN 280* (1993), p. 526 (cit. on p. 43).
- [107] Gårding, E., Botinis, A., and Touati, P. "A comparative study of Swedish, Greek and French intonation". In: *Working papers, Lund University, Department of Linguistics and Phonetics 22* (1982) (cit. on p. 43).
- [108] Gehring, J., Auli, M., Grangier, D., Yarats, D., and Dauphin, Y. N. "Convolutional sequence to sequence learning". In: *International Conference on Machine Learning*. PMLR. 2017, pp. 1243–1252 (cit. on p. 73).
- [109] Gers, F. A., Schmidhuber, J., and Cummins, F. "Learning to Forget: Continual Prediction with LSTM". In: *Neural Computation 12.10* (2000), pp. 2451–2471. eprint: <https://doi.org/10.1162/089976600300015015>. URL: <https://doi.org/10.1162/089976600300015015> (cit. on p. 22).
- [110] Gibbon, D. *Perspectives of intonation analysis*. Herbert Lang Bern, 1976 (cit. on p. 41).
- [111] Gibiansky, A., Arik, S., Diamos, G., et al. "Deep voice 2: Multi-speaker neural text-to-speech". In: *Advances in neural information processing systems*. 2017, pp. 2962–2970 (cit. on pp. 35, 56).
- [112] Golański, F. N. "O wymowie i poezji, powtórne wydanie, nowemi uwagami pomnożone, przez x". In: *Golańskiego SP nauczyciela wymowy, za przywileiem JK Mci., w Wilnie: W Drukarni JK Mci. i Rzeczypospolitej u XX Scholarum Piarum* (1788) (cit. on p. 46).
- [113] Goldsmith, J. A. "Autosegmental phonology. 1976". In: *Indiana University Linguistics Club* (1976) (cit. on p. 45).
- [114] Gonet, W. "Obstruent Voicing in English and Polish. A Pedagogical Perspective". In: *International Journal of English Studies 1.1* (2001), pp. 73–92 (cit. on p. 48).
- [115] Gonet, W. "Seven deadly sins in teaching English phonetics". In: *Materiały z konferencji "Dydaktyka fonetyki języka obcego w Polsce" Mikorzyn k. Konina*. 2004, pp. 1–15 (cit. on p. 48).
- [116] Gonet, W. "Success in the acquisition of English phonetics by Poles. A pilot study". In: *Dydaktyka fonetyki języka obcego w Polsce VI [Phonetics in Foreign Language Teaching 6]*, eds. W. Sobkowiak and E. Waniek-Klimczak (2006), pp. 70–88 (cit. on p. 48).

- [117] Graves, A. and Schmidhuber, J. “Framewise phoneme classification with bidirectional LSTM and other neural network architectures”. In: *Neural networks : the official journal of the International Neural Network Society* 18 (July 2005), pp. 602–10 (cit. on p. 22).
- [118] Grinberg, M. *Flask web development: developing web applications with Python*. O’Reilly Media, Inc., 2018 (cit. on p. 81).
- [119] Grønnum, N. *The groundworks of Danish intonation: An introduction*. Museum Tusculanum, 1992 (cit. on p. 43).
- [120] Groot, A. W. de. “L’intonation de la phrase néerlandaise et allemande considérée du point de vue de la linguistique structurale”. In: *Cahiers Ferdinand de Saussure* 5 (1945), pp. 17–31 (cit. on pp. 13, 43).
- [121] Group, H. W. *HTS Slides ver. 2.3*. [Online; accessed 23.01.2021]. 2021. URL: http://hts.sp.nitech.ac.jp/archives/2.3/HTS_Slides.zip (cit. on p. 18).
- [122] Guenther, F. “Involvement of Auditory Cortex in Speech Production”. In: *The Journal of the Acoustical Society of America* 123 (June 2008), p. 3579 (cit. on p. 28).
- [123] Gulli, A. and Pal, S. *Deep learning with Keras*. Packt Publishing Ltd, 2017 (cit. on pp. 9, 73).
- [124] Gussenhoven, C. et al. *The phonology of tone and intonation*. Cambridge University Press, 2004 (cit. on p. 28).
- [125] Györödi, C., Györödi, R., Pecherle, G., and Olah, A. “A comparative study: MongoDB vs. MySQL”. In: *2015 13th International Conference on Engineering of Modern Electric Systems (EMES)*. IEEE. 2015, pp. 1–6 (cit. on p. 82).
- [126] Hadding-Koch, K. “Acoustico-phonetic studies in the intonation of southern Swedish”. In: (1961) (cit. on p. 14).
- [127] Halliday, M. A. K. *Intonation and grammar in British English*. Vol. 48. Walter de Gruyter GmbH & Co KG, 2015 (cit. on p. 13).
- [128] Harris, Z. S. “Methods in structural linguistics.” In: (1951) (cit. on p. 44).
- [129] Hau, D. and Chen, K. “Exploring hierarchical speech representations with a deep convolutional neural network”. In: *UKCI 2011 Accepted Papers* 37 (2011) (cit. on p. 31).
- [130] Haugen, E. and Joos, M. “Tone and intonation in East Norwegian”. In: *Acta Philologica Scandinavica* 22 (1952), pp. 41–64 (cit. on p. 13).
- [131] He, K., Zhang, X., Ren, S., and Sun, J. “Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification”. In: *2015 IEEE International Conference on Computer Vision (ICCV)*. 2015, pp. 1026–1034 (cit. on p. 76).
- [132] He, K., Zhang, X., Ren, S., and Sun, J. “Identity mappings in deep residual networks”. In: *European Conference on Computer Vision*. Springer. 2016, pp. 630–645 (cit. on p. 33).
- [133] Hempel, C. G. et al. “Aspects of scientific explanation”. In: (1965) (cit. on p. 58).
- [134] Hempel, C. G. and Oppenheim, P. “Studies in the Logic of Explanation”. In: *Philosophy of science* 15.2 (1948), pp. 135–175 (cit. on p. 58).
- [135] Hinton, G. E., Osindero, S., and Teh, Y.-W. “A fast learning algorithm for deep belief nets”. In: *Neural computation* 18.7 (2006), pp. 1527–1554 (cit. on p. 28).

- [136] Hirschberg, J. “Communication and prosody: Functional aspects of prosody”. In: *Speech Communication* 36.1-2 (2002), pp. 31–43 (cit. on pp. 18, 45).
- [137] Hirst, D. and Di Cristo, A. *A survey of intonation systems. ID.(eds.) Intonation Systems. A Survey of Twenty Languages: 1-43*. 1998 (cit. on p. 44).
- [138] Hirst, D., Di Cristo, A., and Espesser, R. “Levels of representation and levels of analysis for the description of intonation systems”. In: *Prosody: Theory and experiment*. Springer, 2000, pp. 51–87 (cit. on pp. 43, 46).
- [139] Hirst, D. and Espesser, R. “Automatic modelling of fundamental frequency using a quadratic spline function.” In: (1993) (cit. on p. 52).
- [140] Hjelmslev, L. and Whitfield, F. J. *Prolegomena to a Theory of Language*. 1953 (cit. on p. 43).
- [141] Hochreiter, S. and Schmidhuber, J. “Long Short-Term Memory”. In: *Neural Computation* 9.8 (1997), pp. 1735–1780. eprint: <https://doi.org/10.1162/neco.1997.9.8.1735>. URL: <https://doi.org/10.1162/neco.1997.9.8.1735> (cit. on pp. 21, 22).
- [142] Holmes, J. N. *Speech synthesis*. Vol. 7. Mills and Boon, 1972 (cit. on p. 17).
- [143] Holmes, W. J. “Copy synthesis of female speech using the JSRU parallel formant synthesiser”. In: *First European Conference on Speech Communication and Technology*. 1989 (cit. on p. 17).
- [144] Holmes W. Holmes J., J. M. “Extension of the Bandwidth of the JSRU Parallel-Formant Synthesizer for High Quality Synthesis of Male and Female Speech”. In: *Proceedings of ICASSP 90*. Vol. 1. 1990, pp. 313–316 (cit. on p. 17).
- [145] Huang, J.-T., Li, J., and Gong, Y. “An analysis of convolutional neural networks for speech recognition”. In: *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2015, pp. 4989–4993 (cit. on p. 31).
- [146] Hubel, D. H. and Wiesel, T. N. “Receptive fields of single neurones in the cat’s striate cortex”. In: *The Journal of Physiology* 148.3 (1959), pp. 574–591. eprint: <https://physoc.onlinelibrary.wiley.com/doi/pdf/10.1113/jphysiol.1959.sp006308>. URL: <https://physoc.onlinelibrary.wiley.com/doi/abs/10.1113/jphysiol.1959.sp006308> (cit. on pp. 25, 26).
- [147] Hubel, D. H. and Wiesel, T. N. “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex”. In: *The Journal of Physiology* 160.1 (1962), pp. 106–154. eprint: <https://physoc.onlinelibrary.wiley.com/doi/pdf/10.1113/jphysiol.1962.sp006837>. URL: <https://physoc.onlinelibrary.wiley.com/doi/abs/10.1113/jphysiol.1962.sp006837> (cit. on pp. 25–27, 30).
- [148] Hullett, P. W., Hamilton, L. S., Mesgarani, N., Schreiner, C. E., and Chang, E. F. “Human superior temporal gyrus organization of spectrotemporal modulation tuning derived from speech stimuli”. In: *Journal of Neuroscience* 36.6 (2016), pp. 2014–2026 (cit. on p. 31).
- [149] Hunt, A. J. and Black, A. W. “Unit selection in a concatenative speech synthesis system using a large speech database”. In: *1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings*. Vol. 1. IEEE. 1996, pp. 373–376 (cit. on p. 52).

- [150] Huynh-Thu, Q. and Ghanbari, M. "Scope of validity of PSNR in image/video quality assessment". In: *Electronics letters* 44.13 (2008), pp. 800–801 (cit. on p. 93).
- [151] Isacenko, A. and Schädlich, H. "Erzeugung kunstlicher deutscher Satzintonationen mit zwei kontrastierenden Tonstufen". In: *Monatsber. Deut. Akad. Wiss. Berlin* 6 (1963) (cit. on p. 14).
- [152] Ishizaka, K. and Flanagan, J. "Synthesis of Voiced Sounds From a Two-Mass Model of the Vocal Cords". In: *Bell System Technical Journal* 51 (July 1972) (cit. on p. 16).
- [153] Jarmołowicz, E., Karpiński, M., Malisz, Z., and Szczyszek, M. "Gesture, prosody and lexicon in task-oriented dialogues: multimedia corpus recording and labelling". In: *Verbal and nonverbal communication behaviours*. Springer, 2007, pp. 99–110 (cit. on p. 47).
- [154] Jassem, W. "A quantitative analysis of standard British-English Nuclear Tones". In: *Journal of Quantitative Linguistics* 3.3 (1996), pp. 229–243 (cit. on pp. 46, 64, 127).
- [155] Jassem, W. *Akcent języka polskiego*. Vol. 31. Zakład Narodowy im. Ossolińskich, 1962 (cit. on p. 64).
- [156] Jassem, W. "Intonacje rdzenne w dialogu angielskim. Analiza akustyczna i statystyczna". In: *Podstawowe założenia fonetyczne i techniczne tłumaczenia różnojęzycznego dialogu w czasie rzeczywistym* (1996), pp. 85–104 (cit. on pp. 46, 64).
- [157] Jassem, W. *Intonation of Conversational English:(educated Southern British)*. 45. Nakł. Wrocławskiego Tow. Naukowego; skł. gł.: Dom Książki, 1952 (cit. on pp. 13, 64).
- [158] Jassem, W. "Polish". In: *Journal of the International Phonetic Association* 33.1 (2003), pp. 103–107 (cit. on p. 64).
- [159] Jilka, M. "Regelbasierte F0 Generierung der Intonationsmuster des Amerikanischen Englisch". In: (1996) (cit. on p. 52).
- [160] Jilka, M., Möhler, G., and Dogil, G. "Rules for the generation of ToBI-based American English intonation". In: *Speech Communication* 28.2 (1999), pp. 83–108 (cit. on p. 52).
- [161] Johnson, M. and Lakoff, G. "Why cognitive linguistics requires embodied realism". In: *Cognitive linguistics* 13.3 (2002), pp. 245–263 (cit. on p. 53).
- [162] Jones, W. E. *Phonetics in linguistics: a book of readings*. 12. Longman Publishing Group, 1973 (cit. on p. 14).
- [163] Joshi, A. V. "Amazon's machine learning toolkit: Sagemaker". In: *Machine Learning and Artificial Intelligence*. Springer, 2020, pp. 233–243 (cit. on p. 75).
- [164] Kalchbrenner, N., Espeholt, L., Simonyan, K., et al. "Neural machine translation in linear time". In: *arXiv preprint arXiv:1610.10099* (2016) (cit. on p. 73).
- [165] Karcevskij, S. "Sur la phonologie de la phrase". In: *Travaux du Cercle linguistique de Prague* 4 (1931), pp. 188–227 (cit. on pp. 13, 43).
- [166] Kawahara, H. *STRAIGHT, exploitation of the other aspect of VOCODER: Perceptually*. 2006 (cit. on pp. 17, 79).
- [167] Kawanami, H., Iwami, Y., Toda, T., Saruwatari, H., and Shikano, K. "GMM-based voice conversion applied to emotional speech synthesis". In: (2003) (cit. on p. 18).

- [168] Kell, C. A., Morillon, B., Kouneiher, F., and Giraud, A.-L. “Lateralization of speech production starts in sensory cortices—a possible sensory origin of cerebral left dominance for speech”. In: *Cerebral Cortex* 21.4 (2011), pp. 932–937 (cit. on p. 28).
- [169] Kernel (image processing). *Kernel (image processing)* — Wikipedia, *The Free Encyclopedia*. [Online; accessed 23.01.2021]. 2021. URL: [https://en.wikipedia.org/wiki/Kernel_\(image_processing\)](https://en.wikipedia.org/wiki/Kernel_(image_processing)) (cit. on p. 24).
- [170] Kim, J., Rohrbach, A., Darrell, T., Canny, J., and Akata, Z. “Textual explanations for self-driving vehicles”. In: *Proceedings of the European conference on computer vision (ECCV)*. 2018, pp. 563–578 (cit. on p. 60).
- [171] Kindermans, P.-J., Schütt, K. T., Alber, M., et al. *Learning how to explain neural networks: PatternNet and PatternAttribution*. 2017. arXiv: 1705.05598 [stat.ML] (cit. on p. 40).
- [172] Kingdon, R. *The groundwork of English intonation*. Longmans, 1959 (cit. on p. 13).
- [173] Kingma, D. P. and Ba, J. “Adam: A method for stochastic optimization”. In: *arXiv preprint arXiv:1412.6980* (2014) (cit. on p. 76).
- [174] Kitcher, P. *Explanatory unification and the causal structure of the world*. University of Minnesota Press, Minneapolis, 1989 (cit. on p. 58).
- [175] Klabbers, E., Stöber, K., Veldhuis, R., Wagner, P., and Breuer, S. “Speech synthesis development made easy: The Bonn open synthesis system”. In: *Seventh European Conference on Speech Communication and Technology*. 2001 (cit. on pp. 17, 64).
- [176] Klatt, D. “Review of text-to-speech conversion for English”. In: *The Journal of the Acoustical Society of America* 82 (Oct. 1987), pp. 737–93 (cit. on pp. 16, 17).
- [177] Klatt, D. H. *Review of selected models of speech perception*. The MIT Press, 1989 (cit. on p. 17).
- [178] Kohlbrenner, M., Bauer, A., Nakajima, S., et al. “Towards best practice in explaining neural network decisions with LRP”. In: *2020 International Joint Conference on Neural Networks (IJCNN)*. IEEE. 2020, pp. 1–7 (cit. on p. 136).
- [179] Krizhevsky, A., Sutskever, I., and Hinton, G. E. “Advances in neural information processing systems”. In: *Neural Information Processing Systems Foundation* 1269 (2012) (cit. on p. 28).
- [180] Krizhevsky, A., Sutskever, I., and Hinton, G. E. “ImageNet Classification with Deep Convolutional Neural Networks”. In: *Commun. ACM* 60.6 (May 2017), pp. 84–90. URL: <https://doi.org/10.1145/3065386> (cit. on pp. 23, 31).
- [181] Królikowski, J. F. *Prozodya polska czyli o Śpiewności i miarach języka Polskiego z przykładami w nótach muzycznych*. nakładem JA Munka, 1821 (cit. on p. 46).
- [182] Kuczmarski, T. “HMM-based speech synthesis applied to Polish”. In: *Speech and Language Technology* 12 (2010), p. 13 (cit. on p. 65).
- [183] Kuczmarski, T. “Overview of HMM-based Speech Synthesis Methods”. In: *Speech and Language Technology* 14 (2014), p. 31 (cit. on p. 18).
- [184] Kumar, D., Taylor, G. W., and Wong, A. *Opening the Black Box of Financial AI with CLEAR-Trade: A CLass-Enhanced Attentive Response Approach for Explaining and Visualizing Deep Learning-Driven Stock Market Prediction*. 2017. arXiv: 1709.01574 [cs.AI] (cit. on pp. 40, 60).

- [185] Lacheret-Dujour, A. *La prosodie du français*. CNRS Ed., 1999 (cit. on p. 41).
- [186] Ladd, D. R. *Intonational Phonology*. Cambridge University Press, 1996 (cit. on pp. 41, 43, 45).
- [187] Ladd, D. R. *Intonational phonology*. Cambridge University Press, 2008 (cit. on p. 28).
- [188] Ladefoged, P. “The revised international phonetic alphabet”. In: *Language* 66.3 (1990), pp. 550–552 (cit. on p. 64).
- [189] Lakoff, G. and Johnson, M. *Metaphors we live by*. University of Chicago press, 1980 (cit. on p. 2).
- [190] Lapuschkin, S., Binder, A., Montavon, G., Müller, K.-R., and Samek, W. “The LRP Toolbox for Artificial Neural Networks”. In: *Journal of Machine Learning Research* 17.1 (Jan. 2016), pp. 3938–3942 (cit. on p. 40).
- [191] Lauritsen, S., Kristensen, M., Olsen, M., et al. “Explainable artificial intelligence model to predict acute critical illness from electronic health records”. In: *Nature Communications* 11 (July 2020) (cit. on p. 40).
- [192] Lawrence, W. “The synthesis of speech from signals which have a low information rate”. In: *Communication Theory* (Jan. 1953), pp. 460–469 (cit. on p. 17).
- [193] Le, T. T., Fu, W., and Moore, J. H. “Scaling tree-based automated machine learning to biomedical big data with a feature set selector”. In: *Bioinformatics* 36.1 (2020), pp. 250–256 (cit. on p. 134).
- [194] Lea, C., Vidal, R., Reiter, A., and Hager, G. D. “Temporal convolutional networks: A unified approach to action segmentation”. In: *European Conference on Computer Vision*. Springer. 2016, pp. 47–54 (cit. on pp. 33, 73).
- [195] Leben, W. R. “Suprasegmental phonology.” PhD thesis. Massachusetts Institute of Technology, 1973 (cit. on p. 45).
- [196] LeCun, Y., Boser, B., Denker, J. S., et al. “Backpropagation Applied to Handwritten Zip Code Recognition”. In: *Neural Computation* 1.4 (1989), pp. 541–551. eprint: <https://doi.org/10.1162/neco.1989.1.4.541>. URL: <https://doi.org/10.1162/neco.1989.1.4.541> (cit. on pp. 23, 24, 27, 31).
- [197] Lee, H., Pham, P., Largman, Y., and Ng, A. “Unsupervised feature learning for audio classification using convolutional deep belief networks”. In: *Advances in neural information processing systems* 22 (2009), pp. 1096–1104 (cit. on p. 31).
- [198] Lee, K.-F. and Hon, H.-W. “Speaker-independent phone recognition using hidden Markov models”. In: *IEEE Transactions on acoustics, speech, and signal processing* 37.11 (1989), pp. 1641–1648 (cit. on pp. 18, 19).
- [199] Lehmann, E. L. and Casella, G. *Theory of point estimation*. Springer Science & Business Media, 1983 (cit. on p. 59).
- [200] Lehmann, E. L. and Casella, G. *Theory of point estimation*. Springer Science & Business Media, 2006 (cit. on p. 93).
- [201] Lemmetty, S. “Review of speech synthesis technology”. In: *Helsinki University of Technology* 320 (1999), pp. 79–90 (cit. on pp. 16–19).
- [202] Levelt, W. J. M. *Speaking: From intention to articulation*. Vol. 1. MIT press, 1993 (cit. on p. 29).

- [203] Li, X. and Zhou, Z. "Speech Command Recognition with Convolutional Neural Network". In: *CS229 Stanford education* (2017) (cit. on p. 31).
- [204] Liberman, M. and Prince, A. "On stress and linguistic rhythm". In: *Linguistic inquiry* 8.2 (1977), pp. 249–336 (cit. on p. 45).
- [205] Liberman, M. Y. "The intonational system of English". PhD thesis. Massachusetts Institute of Technology, 1975 (cit. on p. 45).
- [206] Lieberman, P. and Michaels, S. B. "Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech". In: *The Journal of the Acoustical Society of America* 34.7 (1962), pp. 922–927 (cit. on p. 14).
- [207] Lindsay, G. W. "Convolutional Neural Networks as a Model of the Visual System: Past, Present, and Future". In: *Journal of Cognitive Neuroscience* 0.0 (0). PMID: 32027584, pp. 1–15. eprint: https://doi.org/10.1162/jocn_a_01544. URL: https://doi.org/10.1162/jocn_a_01544 (cit. on p. 28).
- [208] Manassi, M., Sayim, B., and Herzog, M. H. "When crowding of crowding leads to uncrowding". In: *Journal of vision* 13.13 (2013), pp. 10–10 (cit. on p. 27).
- [209] Manassi, M., Sayim, B., and Herzog, M. H. "When crowding of crowding leads to uncrowding". In: *Journal of Vision* 13.13 (2013), pp. 10–10 (cit. on p. 26).
- [210] Mańczak, W. *Enklityki i proklityki w języku polskim*. Komisja Języka Polskiego PAU, 1952 (cit. on p. 46).
- [211] Mańczak, W. *O akcentacji grup dwuwyrzowych*. Komisja Języka Polskiego PAU, 1952 (cit. on p. 46).
- [212] Mańczak, W. "O akcentuacji grup ponaddwuwyrzowych". In: *Język Polski* 32 (1952), pp. 145–156 (cit. on p. 46).
- [213] Markidis, S., Der Chien, S. W., Laure, E., Peng, I. B., and Vetter, J. S. "NVIDIA tensor core programmability, performance & precision". In: *2018 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW)*. IEEE. 2018, pp. 522–531 (cit. on p. 75).
- [214] Martineau, M., Atkinson, P., and McIntosh-Smith, S. "Benchmarking the NVIDIA V100 GPU and tensor cores". In: *European Conference on Parallel Processing*. Springer. 2018, pp. 444–455 (cit. on p. 75).
- [215] Martinez, L. M. and Alonso, J.-M. "Complex Receptive Fields in Primary Visual Cortex". In: *The Neuroscientist* 9.5 (2003). PMID: 14580117, pp. 317–331. eprint: <https://doi.org/10.1177/1073858403252732>. URL: <https://doi.org/10.1177/1073858403252732> (cit. on pp. 25, 26).
- [216] Mathesius, V. "K teorii vetné intonace". In: *Slovo a slovesnost* 3 (1937), pp. 248–249 (cit. on pp. 13, 43).
- [217] Mathew, S. and Varia, J. "Overview of Amazon Web Services". In: *Amazon Whitepapers* (2014) (cit. on p. 75).
- [218] McCullagh, P. et al. "What is a statistical model?" In: *Annals of statistics* 30.5 (2002), pp. 1225–1310 (cit. on p. 59).
- [219] Merkel, D. "Docker: lightweight Linux containers for consistent development and deployment". In: *Linux journal* 2014.239 (2014), p. 2 (cit. on p. 75).

- [220] Mertens, P. “Intonational grouping, boundaries, and syntactic structure in French”. In: *Proceedings of an ESCA Workshop on Prosody*. Lund University, Department of Linguistics, Lund. 1993, pp. 156–159 (cit. on p. 46).
- [221] Mesgarani, N., Cheung, C., Johnson, K., and Chang, E. F. “Phonetic feature encoding in human superior temporal gyrus”. In: *Science* 343.6174 (2014), pp. 1006–1010 (cit. on p. 31).
- [222] Mettas, O. *Etude sur les facteurs ectosémantiques de l’intonation en français*. Centre de Philologie et de Littératures romanes de l’Université de Strasbourg, 1963 (cit. on p. 14).
- [223] Meyer, A., Roelofs, A., and Brehm, L. “Thirty years of Speaking: An introduction to the Special Issue”. In: *Language, Cognition and Neuroscience* 34.9 (2019), pp. 1073–1084. eprint: <https://doi.org/10.1080/23273798.2019.1652763>. URL: <https://doi.org/10.1080/23273798.2019.1652763> (cit. on p. 29).
- [224] Meyer, M., Alter, K., Friederici, A. D., Lohmann, G., and von Cramon, D. Y. “fMRI reveals brain regions mediating slow prosodic modulations in spoken sentences”. In: *Human brain mapping* 17.2 (2002), pp. 73–88 (cit. on p. 29).
- [225] Mitchell, R. L. C., Elliott, R., Barry, M., Cruttenden, A., and Woodruff, P. W. R. “The neural response to emotional prosody, as revealed by functional magnetic resonance imaging”. In: *Neuropsychologia* 41.10 (2003), pp. 1410–1421 (cit. on p. 29).
- [226] Mixdorff, H. “A novel approach to the fully automatic extraction of Fujisaki model parameters”. In: *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 00CH37100)*. Vol. 3. IEEE. 2000, pp. 1281–1284 (cit. on p. 135).
- [227] Mixdorff, H. *An integrated approach to modeling German prosody*. Web Universitätsverlag, 2002 (cit. on pp. 43, 52).
- [228] Möbius, B. “Components of a quantitative model of German intonation”. In: *Proceedings of ICPhS*. Vol. 95. 1995, pp. 108–115 (cit. on pp. 43, 136).
- [229] Möbius, B., Pätzold, M., and Hess, W. “Analysis and synthesis of German F0 contours by means of Fujisaki’s model”. In: *Speech Communication* 13.1-2 (1993), pp. 53–61 (cit. on p. 43).
- [230] Möbius, B. and Van Santen, J. P. “A quantitative model of F0 generation and alignment”. In: *Intonation*. Springer, 2000, pp. 269–288 (cit. on p. 52).
- [231] Möhler, G. “Describing intonation with a parametric model”. In: *Fifth International Conference on Spoken Language Processing*. 1998 (cit. on p. 52).
- [232] Moneta, J. *Enchiridion Polonicum oder Polnisches Handbuch*. 1720, p. 422 (cit. on p. 46).
- [233] Montavon, G., Lapuschkin, S., Binder, A., Samek, W., and Müller, K.-R. “Explaining nonlinear classification decisions with deep Taylor decomposition”. In: *Pattern Recognition* 65 (2017), pp. 211–222 (cit. on pp. 38–40).
- [234] Montavon, G., Samek, W., and Müller, K.-R. “Methods for interpreting and understanding deep neural networks”. In: *Digital Signal Processing* 73 (2018), pp. 1–15 (cit. on pp. 8, 37, 40).
- [235] Nair, V. and Hinton, G. E. “Rectified linear units improve restricted boltzmann machines”. In: *ICML*. 2010 (cit. on p. 75).

- [236] Nowaczyński, T. *O prozody i harmonii języka polskiego [...]*. w Drukarni JK Mci i Rzeczypospolitey u XX. Scholarum Piarum, 1781 (cit. on p. 46).
- [237] Ohala, J. J. and Jaeger, J. J. *Experimental phonology*. Academic Press Orlando, F., 1986 (cit. on p. 30).
- [238] Öhman, S. *Word and sentence intonation: A quantitative model*. Speech Transmission Laboratory, Department of Speech Communication, Royal Institute of Technnology, 1967 (cit. on pp. 14, 43).
- [239] Oord, A. van den, Dieleman, S., Zen, H., et al. *WaveNet: A Generative Model for Raw Audio*. 2016. arXiv: 1609.03499 [cs.SD] (cit. on pp. 3, 17, 31–34, 56, 59, 73, 74, 79, 117).
- [240] Ostendorf, M., Veilleux, N., Hendrix, M., and Macannuco, D. “Linking speech and language processing through prosody”. In: *The Journal of the Acoustical Society of America* 95.5 (1994), pp. 2947–2947 (cit. on p. 45).
- [241] Palaz, D., Doss, M. M., and Collobert, R. “Convolutional neural networks-based continuous speech recognition using raw speech signal”. In: *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2015, pp. 4295–4299 (cit. on p. 31).
- [242] Palmer, H. E. *English intonation with systematic exercises*. W. Heffer, 1924 (cit. on p. 13).
- [243] Pezoa, F., Reutter, J. L., Suarez, F., Ugarte, M., and Vrgoč, D. “Foundations of JSON schema”. In: *Proceedings of the 25th International Conference on World Wide Web*. 2016, pp. 263–273 (cit. on p. 82).
- [244] Pierre, R. and Martin, P. *Prolegomenes al etude des structures intonatives*. M. Didier, 1970 (cit. on p. 41).
- [245] Pierrehumbert, J. *The phonology and phonetics of English intonation; PhD dissertation*. 1980 (cit. on pp. 14, 28, 44, 64).
- [246] Pike, K. L. “The Intonation of American English”. In: (1945) (cit. on p. 14).
- [247] Ping, W., Peng, K., and Chen, J. “Clarinet: Parallel wave generation in end-to-end text-to-speech”. In: *arXiv preprint arXiv:1807.07281* (2018) (cit. on pp. 79, 80).
- [248] Ping, W., Peng, K., Gibiansky, A., et al. “Deep voice 3: Scaling text-to-speech with convolutional sequence learning”. In: *arXiv preprint arXiv:1710.07654* (2017) (cit. on pp. 35, 56).
- [249] Post, B. and Alter, K. “Neural correlates of categorical linguistic and gradient paralinguistic intonation.” In: *ICPhS*. 2015 (cit. on p. 29).
- [250] Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., and Fong, C. “The use of prosody in syntactic disambiguation”. In: *the Journal of the Acoustical Society of America* 90.6 (1991), pp. 2956–2970 (cit. on p. 45).
- [251] *Prosodic Phonology*. Foris, 1986 (cit. on p. 45).
- [252] Rabiner, L. R. “A tutorial on hidden Markov models and selected applications in speech recognition”. In: *Proceedings of the IEEE* 77.2 (1989), pp. 257–286 (cit. on p. 18).

- [253] Ranzato, M., Huang, F. J., Boureau, Y.-L., and LeCun, Y. “Unsupervised learning of invariant feature hierarchies with applications to object recognition”. In: *2007 IEEE conference on computer vision and pattern recognition*. IEEE, 2007, pp. 1–8 (cit. on p. 28).
- [254] Rashno, E., Akbari, A., and Nasersharif, B. *A Convolutional Neural Network model based on Neutrosophy for Noisy Speech Recognition*. Infinite Study, 2019 (cit. on p. 31).
- [255] Remy, P. *Temporal Convolutional Networks for Keras*. <https://github.com/philipperemy/keras-tcn>. 2020 (cit. on p. 73).
- [256] Riesenhuber, M. and Poggio, T. “Hierarchical models of object recognition in cortex”. In: *Nature Neuroscience* 2.11 (1999), pp. 1019–1025 (cit. on p. 31).
- [257] Robin, D. A., Tranel, D., and Damasio, H. “Auditory perception of temporal and spectral events in patients with focal left and right cerebral lesions”. In: *Brain and language* 39.4 (1990), pp. 539–555 (cit. on p. 29).
- [258] Rosen, G. “Dynamic analog speech synthesizer”. In: *The Journal of the Acoustical Society of America* 30.3 (1958), pp. 201–209 (cit. on p. 17).
- [259] Rossi, M. “Contribution à l’étude des faits prosodiques dans un parler de l’Italie du Nord”. In: *Langage et Comportement* 1 (1965), pp. 5–30 (cit. on p. 13).
- [260] A. Botinis, ed. *Intonation: Past, Present, Future*. Dordrecht: Springer Netherlands, 2000, pp. 13–52. URL: https://doi.org/10.1007/978-94-011-4317-2_2 (cit. on pp. 13–15, 41–44).
- [261] Rossi, M., Di Cristo, A., Hirst, D., Martin, P., and Nishinuma, Y. “L’intonation: de l’acoustique à la sémantique.” In: (1981) (cit. on pp. 41, 46).
- [262] Rumelhart, D., Hinton, G. E., and Williams, R. J. “Learning representations by back-propagating errors”. In: *Nature* 323 (1986), pp. 533–536 (cit. on pp. 21, 36).
- [263] Salmon, W. C. *Scientific explanation and the causal structure of the world*. Princeton University Press, 1984 (cit. on p. 58).
- [264] Samek, W., Montavon, G., Binder, A., Lapuschkin, S., and Müller, K.-R. “Interpreting the predictions of complex ML models by layer-wise relevance propagation”. In: *arXiv preprint arXiv:1611.08191* (2016) (cit. on pp. 8, 37, 38, 40).
- [265] Samek, W. and Müller, K.-R. “Towards explainable artificial intelligence”. In: *Explainable AI: interpreting, explaining and visualizing deep learning*. Springer, 2019, pp. 5–22 (cit. on p. 4).
- [266] Samuelsson, J., Skoglund, J., and Lindén, J. “Controlling spectral dynamics in LPC quantization for perceptual enhancement”. In: *Conference Record of the Thirty-First Asilomar Conference on Signals, Systems and Computers (Cat. No. 97CB36136)*. Vol. 2. IEEE, 1997, pp. 1066–1070 (cit. on p. 93).
- [267] Schroeder, M. “A brief history of synthetic speech”. In: *Speech Communication* 13 (Oct. 1993), pp. 231–237 (cit. on pp. 16, 17, 19).
- [268] Schubiger, M. *English intonation, its form and function*. M. Niemeyer Verlag, 1958 (cit. on p. 13).
- [269] Selkirk, E. O. *The relation between sound and structure*. 1984 (cit. on pp. 41, 43, 45).

- [270] Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., and Poggio, T. “Robust Object Recognition with Cortex-Like Mechanisms”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29.3 (2007), pp. 411–426 (cit. on p. 31).
- [271] Shamma, S. “Spectro-Temporal Receptive Fields”. In: *Encyclopedia of Computational Neuroscience*. Ed. by D. Jaeger and R. Jung. New York, NY: Springer New York, 2013, pp. 1–6. URL: https://doi.org/10.1007/978-1-4614-7320-6_437-1 (cit. on p. 31).
- [272] Shaumyan, S. K. “Abstraction in phonology and semantics”. In: *Papers in Slavic Philology* (1977), p. 241 (cit. on p. 30).
- [273] Shaw, J. “The Realistic Empiricism of Mach, James, and Russell: Neutral Monism Reconceived”. In: *Dialogue* -1 (Nov. 2015), pp. 1–3 (cit. on p. 14).
- [274] Shen, J., Pang, R., Weiss, R. J., et al. “Natural TTS Synthesis by Conditioning Wavenet on Mel Spectrogram Predictions”. In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2018), pp. 4779–4783 (cit. on pp. 34, 56).
- [275] Shensa, M. J. “The discrete wavelet transform: wedding the a trous and Mallat algorithms”. In: *IEEE Transactions on signal processing* 40.10 (1992), pp. 2464–2482 (cit. on p. 32).
- [276] Shickel, B. and Rashidi, P. *Sequential Interpretability: Methods, Applications, and Future Direction for Understanding Deep Learning Models in the Context of Sequential Data*. 2020. arXiv: 2004.12524 [cs.LG] (cit. on p. 40).
- [277] Siddiqui, S. A., Mercier, D., Munir, M., Dengel, A., and Ahmed, S. “TSViz: Demystification of Deep Learning Models for Time-Series Analysis”. In: *IEEE Access* (2019), pp. 1–1 (cit. on pp. 40, 78).
- [278] Silverman, K., Beckman, M., Pitrelli, J., et al. “ToBI: A standard for labeling English prosody”. In: *Second international conference on spoken language processing*. 1992 (cit. on p. 65).
- [279] Simonyan, K., Vedaldi, A., and Zisserman, A. “Deep inside convolutional networks: Visualising image classification models and saliency maps”. In: *arXiv preprint arXiv:1312.6034* (2013) (cit. on p. 37).
- [280] Simonyan, K. and Zisserman, A. “Very deep convolutional networks for large-scale image recognition”. In: *arXiv preprint arXiv:1409.1556* (2014) (cit. on p. 28).
- [281] Smilkov, D., Thorat, N., Kim, B., Viégas, F., and Wattenberg, M. “Smoothgrad: removing noise by adding noise”. In: *arXiv preprint arXiv:1706.03825* (2017) (cit. on p. 39).
- [282] Smith, H. L. and Trager, G. L. *An outline of English structure*. American Council of Learned Societies, Washington, DC, 1957 (cit. on pp. 14, 45).
- [283] *Speech Signal Processing Toolkit (SPTK) Version 3.11*. <http://sp-tk.sourceforge.net>. accessed on 02 Feb 2021 (cit. on p. 70).
- [284] Springenberg, J. T., Dosovitskiy, A., Brox, T., and Riedmiller, M. A. “Striving for Simplicity: The All Convolutional Net”. In: *CoRR abs/1412.6806* (2015) (cit. on p. 40).
- [285] Srivastava, N. “Improving neural networks with dropout”. In: *University of Toronto* 182.566 (2013), p. 7 (cit. on p. 75).

- [286] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”. In: *Journal of Machine Learning Research* 15.1 (Jan. 2014), pp. 1929–1958 (cit. on p. 24).
- [287] Steffen-Batogowa, M. *Struktura akcentowa języka polskiego*. Wydawn. Nauk. PWN, 2000 (cit. on p. 46).
- [288] Steffen-Batogowa, M. *Struktura przebiegu melodii polskiego języka ogólnego*. Sorus, 1996 (cit. on p. 46).
- [289] Sturm, I., Lapuschkin, S., Samek, W., and Müller, K.-R. “Interpretable deep neural networks for single-trial EEG classification”. In: *Journal of Neuroscience Methods* 274 (2016), pp. 141–145. URL: <http://www.sciencedirect.com/science/article/pii/S0165027016302333> (cit. on p. 40).
- [290] Sundararajan, M., Taly, A., and Yan, Q. “Axiomatic Attribution for Deep Networks”. In: *ArXiv abs/1703.01365* (2017) (cit. on p. 40).
- [291] Syrdal, A., Stylianou, Y., Garrison, L., Conkie, A., and Schroeter, J. “TD-PSOLA versus harmonic plus noise model in diphone based speech synthesis”. In: *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP’98 (Cat. No. 98CH36181)*. Vol. 1. IEEE. 1998, pp. 273–276 (cit. on p. 79).
- [292] t’Hart, J., Collier, R., and Cohen, A. *A perceptual study of intonation: an experimental-phonetic approach to speech melody*. Cambridge University Press, 1990 (cit. on p. 44).
- [293] Talkin, D. and Kleijn, W. B. “A robust algorithm for pitch tracking (RAPT)”. In: *Speech coding and synthesis* 495 (1995), p. 518 (cit. on p. 70).
- [294] Tamm, M.-O., Muhammad, Y., and Muhammad, N. “Classification of Vowels from Imagined Speech with Convolutional Neural Networks”. In: *Computers* 9.2 (2020), p. 46 (cit. on p. 35).
- [295] Tang, C., Hamilton, L., and Chang, E. “Intonational speech prosody encoding in the human auditory cortex”. In: *Science* 357.6353 (2017), pp. 797–801 (cit. on p. 28).
- [296] Tatham, M. “Cognitive Phonetics”. In: *Advances in Speech, Hearing and Language Processing* 1 (1990), pp. 193–218 (cit. on pp. 15, 30).
- [297] Taylor, P., Black, A. W., and Caley, R. “The architecture of the Festival speech synthesis system”. In: *The Third ESCA/COCOSDA Workshop (ETRW) on Speech Synthesis*. 1998 (cit. on pp. 17, 52, 65).
- [298] Tazebay, M. V. and Akansu, A. N. “Adaptive subband transforms in time-frequency excisers for DSSS communications systems”. In: *IEEE Transactions on Signal Processing* 43.11 (1995), pp. 2776–2782 (cit. on p. 32).
- [299] Tazebay, M. V. and Akansu, A. N. “Progressive optimality in hierarchical filter banks”. In: *Proceedings of 1st International Conference on Image Processing*. Vol. 1. IEEE. 1994, pp. 825–829 (cit. on p. 32).
- [300] Teranishi, R. and Umeda, N. “Use of pronouncing dictionary in speech synthesis experiments”. In: *Reports of the Sixth International Congress on Acoustics*. Vol. 2. 1968, pp. 155–158 (cit. on p. 17).
- [301] Tian, B., Kuśmierk, P., and Rauschecker, J. P. “Analogues of simple and complex cells in rhesus monkey auditory cortex”. In: *Proceedings of the National Academy of Sciences* 110.19 (2013), pp. 7892–7897 (cit. on p. 30).

- [302] Tokuda, K., Kobayashi, T., Imai, S., and Chiba, T. “Spectral estimation of speech by mel-generalized cepstral analysis”. In: *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)* 76.2 (1993), pp. 30–43 (cit. on p. 73).
- [303] Tokuda, K., Yoshimura, T., Masuko, T., Kobayashi, T., and Kitamura, T. “Speech parameter generation algorithms for HMM-based speech synthesis”. In: *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 00CH37100)*. Vol. 3. IEEE. 2000, pp. 1315–1318 (cit. on p. 21).
- [304] Tokuda, K., Zen, H., and Black, A. W. *An HMM-Based Speech Synthesis System Applied To English*. 2002 (cit. on pp. 17, 19).
- [305] Tourville, J. A., Reilly, K. J., and Guenther, F. H. “Neural mechanisms underlying auditory feedback control of speech”. In: *Neuroimage* 39.3 (2008), pp. 1429–1443 (cit. on p. 28).
- [306] Trager, G. L. and Smith, H. L. *An Outline of English Structure (Norman, OK)*. 1951 (cit. on pp. 44, 45).
- [307] Trubetzkoy, N. S. “Grundzüge der Phonologie”. In: *Travaux du Cercle linguistique de Prague* 7 (1939), p. 271 (cit. on p. 41).
- [308] Tynjanov, J. and Jakobson, R. “Problemy izučenija literatury i jazyka”. In: *Novyj lef* 12 (1928), pp. 35–37 (cit. on p. 43).
- [309] Uldall, E. “Attitudinal meanings conveyed by intonation contours”. In: *Language and Speech* 3.4 (1960), pp. 223–234 (cit. on p. 14).
- [310] UPC, A. B., Höge, H., Tropf, H. S., et al. *Deliverable no.: D8 Title: TTS Baselines and specifications*. <http://suendermann.com/su/pdf/tcstar2005b.pdf>. accessed on 13 Apr 2008 (cit. on p. 62).
- [311] Valbret, H., Moulines, E., and Tubach, J.-P. “Voice transformation using PSOLA technique”. In: *Speech communication* 11.2-3 (1992), pp. 175–187 (cit. on p. 79).
- [312] Van den Oord, A., Kalchbrenner, N., Espeholt, L., Vinyals, O., Graves, A., et al. “Conditional image generation with pixelcnn decoders”. In: *Advances in neural information processing systems* 29 (2016), pp. 4790–4798 (cit. on pp. 3, 31, 33, 60).
- [313] Van den Oord, A., Li, Y., Babuschkin, I., et al. “Parallel Wavenet: Fast high-fidelity speech synthesis”. In: *International conference on machine learning*. PMLR. 2018, pp. 3918–3926 (cit. on pp. 79, 80).
- [314] Van Fraassen, B. C. et al. *The scientific image*. Oxford University Press, 1980 (cit. on p. 58).
- [315] Van Rossum, G. and Drake, F. L. *Python 3 Reference Manual*. Scotts Valley: CreateSpace, 2009 (cit. on p. 73).
- [316] Van Santen, J., Kain, A., Klabbers, E., and Mishra, T. “Synthesis of prosody using multi-level unit sequences”. In: *Speech Communication* 46.3-4 (2005), pp. 365–375 (cit. on p. 43).
- [317] Wagner, A. “A comprehensive model of intonation for application in speech synthesis”. In: *Unpublished PhD thesis, Adam Mickiewicz University, Poznan, Poland* (2008) (cit. on pp. 48, 52).
- [318] Wagner, A., Bachan, J., Klessa, K., and Demenko, G. “Przegląd wybranych aspektów analizy prozodii mowy spontanicznej na potrzeby technologii mowy”. In: *Prace Filologiczne* 66 (2015), pp. 271–298 (cit. on p. 49).

- [319] Wang, X., Takaki, S., and Yamagishi, J. “Neural source-filter waveform models for statistical parametric speech synthesis”. In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 28 (2019), pp. 402–415 (cit. on pp. 8, 79–81, 117, 120).
- [320] Wang, Y., Skerry-Ryan, R. J., Stanton, D., et al. “Tacotron: Towards end-to-end speech synthesis”. In: *arXiv preprint arXiv:1703.10135* (2017) (cit. on pp. 34, 56).
- [321] Wells, J., Barry, W., Grice, M., Fourcin, A., and Gibbon, D. “Standard computer-compatible transcription”. In: *Esprit project 2589 (SAM), Doc. no. SAM-UCL 37* (1992) (cit. on p. 64).
- [322] Wells, J. C. et al. “SAMPA computer readable phonetic alphabet”. In: *Handbook of standards and resources for spoken language systems* 4 (1997), pp. 684–732 (cit. on p. 64).
- [323] Wells, R. S. “The pitch phonemes of English”. In: *Language* (1945), pp. 27–39 (cit. on pp. 14, 45).
- [324] Wildgruber, D., Riecker, A., Hertrich, I., et al. “Identification of emotional intonation evaluated by fMRI”. In: *Neuroimage* 24.4 (2005), pp. 1233–1241 (cit. on p. 29).
- [325] Wittgenstein, L. “Logisch-philosophische abhandlung”. In: *Annalen der Naturphilosophie* 14 (1921), pp. 185–262 (cit. on p. 82).
- [326] Wodarz, H.-W. “Über vergleichende satzmelodische Untersuchungen”. In: *Phonetica* 5.2 (1960), pp. 75–98 (cit. on p. 13).
- [327] Wode, H. “Englische Satzintonation (Part 1 of 4)”. In: *Phonetica* 15.3-4 (1966), pp. 129–151 (cit. on p. 13).
- [328] Yamagishi, J., Zen, H., Toda, T., and Tokuda, K. *Speaker-independent HMM-based speech synthesis system: HTS-2007 system for the Blizzard Challenge 2007*. 2007 (cit. on p. 19).
- [329] Ye, C., Evanusa, M., He, H., et al. “Network Deconvolution”. In: *CoRR abs/1905.11926* (2019). arXiv: 1905.11926. URL: <http://arxiv.org/abs/1905.11926> (cit. on p. 40).
- [330] Young, S., Evermann, G., Gales, M., et al. *The HTK book*. Vol. 3. 175. University of Cambridge, Department of Engineering Cambridge, 2002, p. 12 (cit. on pp. 65, 67).
- [331] Young, S. J. and Young, S. J. “The HTK hidden Markov model toolkit: Design and philosophy”. In: (1993) (cit. on p. 65).
- [332] Zeiler, M. D. and Fergus, R. “Visualizing and understanding convolutional networks”. In: *European Conference on Computer Vision*. Springer. 2014, pp. 818–833 (cit. on pp. 28, 36, 40).
- [333] Zen, H., Senior, A., and Schuster, M. “Statistical parametric speech synthesis using deep neural networks”. In: *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. 2013, pp. 7962–7966 (cit. on pp. 17, 19–21).
- [334] Zen, H. *Statistical parametric speech synthesis: from HMM to LSTM-RNN*. Lecture given at RTTH Summer School on Speech Technology, Barcelona, Spain. 2015 (cit. on p. 19).
- [335] Zen, H., Tokuda, K., and Black, A. W. “Statistical parametric speech synthesis”. In: *Speech Communication* 51.11 (2009), pp. 1039–1064 (cit. on pp. 20, 65).

- [336] Zen, H., Tokuda, K., Masuko, T., Kobayashi, T., and Kitamura, T. “Hidden semi-Markov model based speech synthesis”. In: *Eighth International Conference on Spoken Language Processing*. 2004 (cit. on p. 18).
- [337] Zintgraf, L. M., Cohen, T. S., Adel, T., and Welling, M. “Visualizing deep neural network decisions: Prediction difference analysis”. In: *arXiv preprint arXiv:1702.04595* (2017) (cit. on p. 60).

List of Figures

2.1	An overview of the basic HMM-based speech synthesis system. Adopted from the HTS Slides [121] (released under the Creative Commons Attribution 3.0 license)	18
2.2	General DNN-based speech synthesis system. Adopted from Zen et al. [333].	20
2.3	Long Short Term Memory gate scheme. Adopted from Hochreiter and Schmidhuber [141].	22
2.4	Example of a 2-dimensional matrix convolution.	23
2.5	Examples of convoluting and image with different convolution kernels. Adopted from the Wikipedia [169].	24
2.6	Convolutional Neural Network (LeNet-5). Adopted from LeCun et al. [196].	24
2.7	Simple receptive fields. Adopted from Hubel and Wiesel [147].	26
2.8	Three different types of complex receptive fields. Adopted from Hubel and Wiesel [147].	27
2.9	Hierarchical, feedforward visual processing in human brain. Adopted from Manassi et al. [208].	27
2.10	A blueprint for the speaker - model of speech production. Adopted from Levelt [202] and A. Meyer et al. [223].	29
2.11	Dilated causal convolutions. Adopted from the original WaveNet paper [239].	32
2.12	Residual and skip connections from a stack of k gated convolutional layers. Adopted from the original WaveNet paper [239].	33
2.13	Google WaveNet evaluation results as compared with Google's best concatenative and parametric systems. Adopted from Oord et al. [239].	34
2.14	Tacotron 2 architecture. Adopted from Shen et al. [274] (©2018 IEEE).	34
2.15	Learned convolutional filter visualization example. Adopted from Zeiler and Fergus [332].	36
2.16	Results of sensitivity-based and relevance-based explainability methods. Based on Samek et al. [264]	38

2.17	Computational flow of deep Taylor decomposition. Adopted from Montavon (2017) [233].	39
2.18	Acoustic realizations of the 9 different accents. Adopted from Demenko [70].	48
3.1	Google's Deep Dream example.	57
3.2	Extracted F_0 values for "Słyszałam odgłos zbliżającego się pociągu" (<i>I heard the sound of an approaching train</i>).	72
3.3	Architecture of the Residual Block used in the current network architecture.	74
3.4	Three types of neural waveform models in training stage. Adopted from X. Wang et al. [319] (©[2019] IEEE).	80
3.5	Diagram of harmonic-plus-noise NSF (hn-NSF) model. Adopted from [319] (©[2019] IEEE).	81
3.6	Infrastructure of the web experiment framework.	83
3.7	Perceptual evaluation experiment trial scheme.	84
4.1	Result of F_0 prediction for "Lokatorzy znaleźli się w podbramkowej sytuacji i musieli się wyprowadzić" (<i>The tenants found themselves in a difficult situation and had to move out</i>).	87
4.2	Result of F_0 prediction for "Powodzenie nie jest gwarantowane" (<i>Success is not guaranteed</i>).	88
4.3	Result of F_0 prediction for "Gaduła była bardzo nieznośna" (<i>Gabby was very annoying.</i>).	89
4.4	Result of F_0 prediction for "Może przyniosą też gorzałę" (<i>Maybe they will bring booze too.</i>).	89
4.5	Result of F_0 prediction for "To jest ważna godzina dla nas wszystkich" (<i>This is an important hour for all of us</i>).	90
4.6	Result of F_0 prediction for "Myślę, że chleb razowy będzie najlepszy" (<i>I think that a wholemeal bread will be the best</i>).	90
4.7	Result of F_0 prediction for "Słyszałam odgłos zbliżającego się pociągu" (<i>I heard the sound of an approaching train</i>).	91
4.8	Result of F_0 prediction for "Czy to był łatwy dobór?" (<i>Was it an easy choice?</i>).	91
4.9	Result of F_0 prediction for "To Majka" (<i>This is Majka</i>).	92
4.10	Result of F_0 prediction for "Na czym polega kandyzacja?" (<i>How does candying work?</i>).	92
4.11	Mean Opinion Score-based evaluation results.	94

4.12	Mean Opinion Score-based evaluation mean response times for synthetic and natural stimuli.	94
4.13	Mean Opinion Score-based evaluation total numbers of specific scores for natural stimuli.	95
4.14	Mean Opinion Score-based evaluation total numbers of specific scores for synthetic stimuli.	95
4.15	Mean Opinion Score-based evaluation parameters correlation matrix. .	96
4.16	ABX experiment results. Mean score per listener (%).	97
4.17	ABX experiment results. Number of individuals with n -correct answers out of 10.	97
4.18	ABX experiment mean response times for synthetic and natural stimuli.	98
4.19	ABX experiment number of correct and incorrect answers in case of natural and synthetic stimuli.	98
4.20	ABX experiment parameters correlation matrix.	99
4.21	Fundamental frequency predictions for "Słyszałam odgłos zbliżającego się pociągu" (<i>I heard the sound of an approaching train</i>) aligned with feature relevance heatmap.	100
4.22	Fundamental frequency predictions for "Waluta jeny - w języku greckim - jest cenna" (<i>The currency yen, in Greek, is valuable</i>) aligned with relevance heatmap for general feature groups.	102
4.23	Fundamental frequency predictions for "Słyszałam odgłos zbliżającego się pociągu" (<i>I heard the sound of an approaching train</i>) aligned with relevance heatmap for general feature groups.	103
4.24	Fundamental frequency predictions for "Tego z pewnością żaden nie zrobi" (<i>Certainly none will do that</i>) aligned with relevance heatmap for general feature groups.	104
4.25	Fundamental frequency predictions for "Musisz dojrzeć do tego, by to zrozumieć" (<i>You have to grow up to understand it</i>) aligned with relevance heatmap for features grouped by the level of utterance.	105
4.26	Fundamental frequency predictions for "Przepłynęłam na grzbiecie siedemnaście długości basenu" (<i>I swam seventeen pool lengths on my back.</i>) aligned with relevance heatmap for features grouped by the type of relation.	105
4.27	Fundamental frequency predictions for "Ciocia pracuje w urzędzie państwowym" (<i>Aunt works in a government office</i>) aligned with relevance heatmap for features grouped by the type of relation and the level of utterance.	106
4.28	Relevance sum across the whole test set.	107

4.29	Mean relevance of features based on regular sum.	107
4.30	Absolute mean relevance of features (based on absolute sum).	108
4.31	Mean relevance of features calculated for positive-only values.	108
4.32	Mean relevance of features calculated for negative-only values.	109
4.33	Regular mean relevance per feature group. The y -axis scaling due to the high relevance of V/UV renders comparatively flat plots for other features.	110
4.34	Regular mean relevance per feature group with the most relevant feature (V/UV) excluded.	111
4.35	Regular mean relevance per feature group with the most relevant feature (V/UV) excluded. Medium granularity of feature groups.	112
4.36	Regular mean relevance per feature group with the most relevant feature (V/UV) excluded. High granularity of feature groups.	113
4.37	Absolute mean relevance per feature group with the most relevant feature (V/UV) excluded.	114
4.38	Positive relevance-only mean per feature group with the most relevant feature (V/UV) excluded.	115
4.39	Negative relevance-only mean per feature group with the most relevant feature (V/UV) excluded.	116

List of Tables

2.1	Pre-ictal accent types	47
2.2	Ictal accent types	47
3.1	Stress and accent labels used in the original Polish BOSS speech corpus	62
3.2	Prosodic phrase boundary labels used in the original Polish BOSS speech corpus	63
3.3	An example of the intermediate Festvox-based label structure for "Smutek malował się na jej twarzy" (<i>Sadness was showing on her face</i>).	66
3.4	Tone accent and phrase boundary labels conversion scheme.	67
3.5	An example fragment of the HTK label file for the sentence "Smutek malował się na jej twarzy".	68
3.6	List of segmental HTK HED question types used for feature extraction .	69
3.7	List of non-segmental HTK HED question types used for feature extraction.	71
4.1	Mean MOS scores in the test dataset for $\log F_0$ predictions.	93

List of Listings

3.1	An example BLF label file for the sentence "Smutek malował się na jej twarzy" (<i>Sadness was showing on her face</i>) taken from the original Polish BOSS corpus.	64
A.1	A comma-separated listing of feature indexes and their codenames including the HTK HED matching patterns used for extraction from the full-context labels.	171
B.1	Structure of features groups as defined in the Python source code of the current work.	212

Listing of Features

This section contains a raw listing (Listing A.1) of the features used for training the intonation model, F_0 inference and input feature analysis. Each numbered line in the listing starts with a numeric index of the feature. The same indexes appear in some of the figures throughout this work and in the project repository, where the individual relevance of all input features is presented. The index is followed by a comma after which a HTK keyword `QS` appears. The keyword is used internally by some of the feature extraction scripts employed in this work and denotes a *question* – a special regular expression-like mask used for pattern matching on the full-context HTK labels. After that a feature name follows. The names of features start with a code that denotes the quintphone context to which the feature pertains. `LL` stands for the leftmost context, `L` for the direct left context, `C` for the central (current) segment, `R` for the direct right context and `RR` for the rightmost context of the quintphone. The code is followed by a dash after which a symbolic feature name appears. After the name a list of matching masks used to extract the current feature is given in curly braces. The masks are used to match parts of the full-context HTK labels and if any of the expressions returns `True` than a `1` is returned for the whole feature, and `0` otherwise. In this way it is easy to obtain a vector of one-hot encoded features ready for use in neural network training, as most neural nets expect the input feature vector to contain values in the 0–1 or -1–1 range.

The lines 0–82 contain features related to the phonetic and phonological identity of the leftmost segment in the quintphone. This includes the vowel/consonant distinction, segment type, place of articulation, voicing, and specific phoneme identity.

Lines 83–164, 165–246, 247–326 and 327–405 contain that same set of features but for the remaining quintphone contexts; the direct left, central, direct right and rightmost contexts.

The features in lines 406–436 represent the number of segments found in the current syllable to the right (`Fw`) and to the left (`Bw`) of the current segment. Both the equality (`==`) and inequality (`<=`) variants of the quantitative features are included in this and all other similar cases where length or distance is measured.

Lines 437–462 contain features related to various information the most immediate left neighbour of the current syllable. In addition to simple binary information about the presence or absence of stress and accent, also features representing the presence of specific ToBI tones and syllable length measured in segments are included.

Lines 463–490 contain that same information but for the current syllable.

Additionally, in lines 491–602 features representing the forward and backward position of the current syllable in the current word and phrase are given.

Features that represent the number of stressed and accented syllables before and after the current syllable in the current phrase are listed in lines 603–680.

Lines 681–776, in turn, are related to the distance from the previous and next stressed and accented syllables.

Lines 777–792 contain additional information about the phonetic/phonological identity of the syllable nucleus.

Features included in lines 794–819 contain information similar to these provided for the immediate left neighbour of the current syllable, but for the immediate right neighbour, i.e.: presence or absence of stress and accent, information about the specific ToBI tone and syllable length measured in segments.

Starting from line 820, word-level features are listed.

Lines 820–834 contain features that express the length of the immediate left neighbour of the current word, expressed in number of syllables.

The lengths of the current word are included in lines 835–849.

Additionally, information about the forward and backward position of the current word in the current phrase are covered by features in lines 850–904.

Similarly to some features included for the current syllable, lines 905–942 contain features related to the number of content (accented) words before and after the current word in the current phrase.

Features in lines 943–968 encode the number of words from the current word to the previous and next content word.

Lines 969–983 contain features related to the length of the right neighbour of the current word expressed in number of syllables.

Starting from line 984, phrase-related features are listed.

Lines 984–1209, contain features related to the length of the immediate right and left neighbours, as well as, the current phrase, expressed in number of syllables and words. Additionally, the forward and backward position in the current utterance was included for the current phrase.

Finally, lines 1210–1296 contain features related to the length of the whole utterance expressed in number of syllables, words and phrases.

The last feature in line 1297 represents the presence of voicing (voiced/unvoiced) at the current position. This feature was appended manually for reference as it is not included in the HTK question script but extracted separately.

```
1 0, QS LL-Vowel{i^*, y^*, e^*, a^*, o^*, u^*, schwa^*}
2 1, QS LL-Consonant{gs^*, p^*, b^*, t^*, d^*, k^*, g^*, ki^*, gi^*, f^*, v^*, s
   ^*, si^*, z^*, zi^*, sz^*, rz^*, x^*, c^*, dz^*, cz^*, drz^*, ci^*, dzi^*, m
   ^*, n^*, ni^*, ng^*, l^*, r^*, w^*, ww^*, j^*, jj^*}
3 2, QS LL-Stop{gs^*, p^*, b^*, t^*, d^*, k^*, g^*}
4 3, QS LL-Nasal{ww^*, jj^*, m^*, n^*, ni^*, ng^*}
5 4, QS LL-Fricative{f^*, v^*, s^*, si^*, z^*, zi^*, sz^*, rz^*, x^*}
6 5, QS LL-Front{e^*, i^*, y^*, f^*, v^*, p^*, b^*, m^*, w^*, ww^*}
7 6, QS LL-Central{schwa^*, a^*, t^*, d^*, s^*, si^*, z^*, zi^*, n^*, r^*, l^*, t
   ^*, d^*, sz^*, rz^*, cz^*, drz^*, c^*, dz^*, ci^*, dzi^*}
8 7, QS LL-Back{o^*, u^*, k^*, g^*, ki^*, gi^*, ng^*, x^*, gs^*}
9 8, QS LL-Front_Vowel{e^*, i^*, y^*}
10 9, QS LL-Central_Vowel{a^*, schwa^*}
11 10, QS LL-Back_Vowel{o^*, u^*}
12 11, QS LL-High_Vowel{i^*, y^*, u^*}
13 12, QS LL-Medium_Vowel{e^*, o^*}
14 13, QS LL-Low_Vowel{a^*}
15 14, QS LL-Rounded_Vowel{o^*, u^*}
16 15, QS LL-Unrounded_Vowel{a^*, e^*, i^*, y^*}
17 16, QS LL-IVowel{i^*}
18 17, QS LL-EVowel{e^*}
19 18, QS LL-AVowel{a^*}
20 19, QS LL-OVowel{o^*}
21 20, QS LL-UVowel{u^*}
22 21, QS LL-YVowel{y^*}
23 22, QS LL-SCHWAVowel{schwa^*}
24 23, QS LL-Unvoiced_Consonant{gs^*, p^*, t^*, k^*, ki^*, f^*, v^*, s^*, sz^*,
   x^*, c^*, cz^*, ci^*}
25 24, QS LL-Voiced_Consonant{b^*, d^*, g^*, gi^*, v^*, z^*, zi^*, rz^*, dz^*,
   drz^*, dzi^*, m^*, n^*, ni^*, ng^*, l^*, r^*, w^*, ww^*, j^*, jj^*}
26 25, QS LL-Front_Consonant{f^*, v^*, f^*, p^*, b^*, m^*, w^*, ww^*}
27 26, QS LL-Central_Consonant{t^*, d^*, s^*, si^*, z^*, zi^*, n^*, r^*, l^*, t
   ^*, d^*, sz^*, rz^*, cz^*, drz^*, c^*, dz^*, ci^*, dzi^*}
```

```

28 27, QS LL-Back_Consonant{gs^*, k^*, g^*, ki^*, gi^*, ng^*, x^*}
29 28, QS LL-Fortis_Consonant{gs^*, cz^*, f^*, k^*, p^*, s^*, sz^*, t^*, ci^*, c
    ^*, ki^*}
30 29, QS LL-Lenis_Consonant{drz^*, v^*, g^*, b^*, rz^*, z^*, d^*, dzi^*, dz^*,
    gi^*, zi^*}
31 30, QS LL-Neigther_F_or_L{m^*, n^*, ni^*, ng^*, l^*, r^*, w^*, ww^*, j^*, jj
    ^*}
32 31, QS LL-Voiced_Stop{b^*, d^*, g^*}
33 32, QS LL-Unvoiced_Stop{p^*, t^*, k^*, gs^*}
34 33, QS LL-Front_Stop{b^*, p^*}
35 34, QS LL-Central_Stop{d^*, t^*}
36 35, QS LL-Back_Stop{g^*, k^*, gs^*}
37 36, QS LL-Voiced_Fricative{v^*, z^*, zi^*, rz^*}
38 37, QS LL-Unvoiced_Fricative{f^*, s^*, si^*, sz^*, x^*}
39 38, QS LL-Front_Fricative{f^*, v^*}
40 39, QS LL-Affricate_Consonant{dz^*, drz^*, dzi^*, c^*, cz^*, ci^*}
41 40, QS LL-silences{pau^*}
42 41, QS LL-schwa{schwa^*}
43 42, QS LL-a{a^*}
44 43, QS LL-e{e^*}
45 44, QS LL-i{i^*}
46 45, QS LL-y{y^*}
47 46, QS LL-o{o^*}
48 47, QS LL-u{u^*}
49 48, QS LL-p{p^*}
50 49, QS LL-b{b^*}
51 50, QS LL-t{t^*}
52 51, QS LL-d{d^*}
53 52, QS LL-k{k^*}
54 53, QS LL-ki{ki^*}
55 54, QS LL-g{g^*}
56 55, QS LL-gi{gi^*}
57 56, QS LL-f{f^*}
58 57, QS LL-v{v^*}
59 58, QS LL-s{s^*}
60 59, QS LL-si{si^*}
61 60, QS LL-z{z^*}
62 61, QS LL-zi{zi^*}
63 62, QS LL-sz{sz^*}
64 63, QS LL-rz{rz^*}
65 64, QS LL-x{x^*}
66 65, QS LL-c{c^*}
67 66, QS LL-dz{dz^*}
68 67, QS LL-cz{cz^*}
69 68, QS LL-drz{drz^*}
70 69, QS LL-ci{ci^*}
71 70, QS LL-dzi{dzi^*}

```

72 71, QS LL-m{m^*}
73 72, QS LL-n{n^*}
74 73, QS LL-ni{ni^*}
75 74, QS LL-ng{ng^*}
76 75, QS LL-l{l^*}
77 76, QS LL-r{r^*}
78 77, QS LL-w{w^*}
79 78, QS LL-ww{ww^*}
80 79, QS LL-j{j^*}
81 80, QS LL-jj{jj^*}
82 81, QS LL-gs{gs^*}
83 82, QS L-Vowel{*^schwa-, *^i-, *^y-, *^e-, *^a-, *^o-, *^u-}
84 83, QS L-Consonant{*^gs-, *^p-, *^b-, *^t-, *^d-, *^k-, *^g-, *^ki-, *^gi
-, *^f-, *^v-, *^s-, *^si-, *^z-, *^zi-, *^sz-, *^rz-, *^x-, *^c-, *^dz-, *^
cz-, *^drz-, *^ci-, *^dzi-, *^m-, *^n-, *^ni-, *^ng-, *^l-, *^r-, *^w-, *^
ww-, *^j-, *^jj-}
85 84, QS L-Stop{*^p-, *^b-, *^t-, *^d-, *^k-, *^g-, *^gs-}
86 85, QS L-Nasal{*^ww-, *^jj-, *^m-, *^n-, *^ni-, *^ng-}
87 86, QS L-Fricative{*^f-, *^v-, *^s-, *^si-, *^z-, *^zi-, *^sz-, *^rz-, *^x-}
88 87, QS L-Front{*^e-, *^i-, *^y-, *^f-, *^v-, *^p-, *^b-, *^m-, *^w-, *^ww-}
89 88, QS L-Central{*^schwa-, *^a-, *^t-, *^d-, *^s-, *^si-, *^z-, *^zi-, *^n
-, *^r-, *^l-, *^t-, *^d-, *^sz-, *^rz-, *^cz-, *^drz-, *^c-, *^dz-, *^ci
-, *^dzi-}
90 89, QS L-Back{*^o-, *^u-, *^k-, *^g-, *^ki-, *^gi-, *^ng-, *^x-, *^gs-}
91 90, QS L-Front_Vowel{*^e-, *^i-, *^y-}
92 91, QS L-Central_Vowel{*^a-, *^schwa-}
93 92, QS L-Back_Vowel{*^o-, *^u-}
94 93, QS L-High_Vowel{*^i-, *^y-, *^u-}
95 94, QS L-Medium_Vowel{*^e-, *^o-}
96 95, QS L-Low_Vowel{*^a-}
97 96, QS L-Rounded_Vowel{*^o-, *^u-}
98 97, QS L-Unrounded_Vowel{*^a-, *^e-, *^i-, *^y-}
99 98, QS L-IVowel{*^i-}
100 99, QS L-EVowel{*^e-}
101 100, QS L-AVowel{*^a-}
102 101, QS L-OVowel{*^o-}
103 102, QS L-UVowel{*^u-}
104 103, QS L-YVowel{*^y-}
105 104, QS LL-SCHWAvowel{*^schwa-}
106 105, QS L-Unvoiced_Consonant{*^gs-, *^p-, *^t-, *^k-, *^ki-, *^f-, *^v-, *^
s-, *^sz-, *^x-, *^c-, *^cz-, *^ci-}
107 106, QS L-Voiced_Consonant{*^b-, *^d-, *^g-, *^gi-, *^v-, *^z-, *^zi-, *^rz
-, *^dz-, *^drz-, *^dzi-, *^m-, *^n-, *^ni-, *^ng-, *^l-, *^r-, *^w-, *^ww
-, *^j-, *^jj-}
108 107, QS L-Front_Consonant{*^f-, *^v-, *^f-, *^p-, *^b-, *^m-, *^w-, *^ww-}

109 108, QS L-Central_Consonant{*^t-,*^d-,*^s-,*^si-,*^z-,*^zi-,*^n-,*^r-
-,*^l-,*^t-,*^d-,*^sz-,*^rz-,*^cz-,*^drz-,*^c-,*^dz-,*^ci-,*^dzi-
-}
110 109, QS L-Back_Consonant{*^gs-,*^k-,*^g-,*^ki-,*^gi-,*^ng-,*^x-}
111 110, QS L-Fortis_Consonant{*^gs-,*^cz-,*^f-,*^k-,*^p-,*^s-,*^sz-,*^t-
-,*^ci-,*^c-,*^ki-}
112 111, QS L-Lenis_Consonant{*^drz-,*^v-,*^g-,*^b-,*^rz-,*^z-,*^d-,*^
dzi-,*^dz-,*^gi-,*^zi-}
113 112, QS L-Neigther_F_or_L{*^m-,*^n-,*^ni-,*^ng-,*^l-,*^r-,*^w-,*^ww-
-,*^j-,*^jj-}
114 113, QS L-Voiced_Stop{*^b-,*^d-,*^g-}
115 114, QS L-Unvoiced_Stop{*^gs-,*^p-,*^t-,*^k-}
116 115, QS L-Front_Stop{*^b-,*^p-}
117 116, QS L-Central_Stop{*^d-,*^t-}
118 117, QS L-Back_Stop{*^gs-,*^g-,*^k-}
119 118, QS L-Voiced_Fricative{*^v-,*^z-,*^zi-,*^rz-}
120 119, QS L-Unvoiced_Fricative{*^f-,*^s-,*^si-,*^sz-,*^x-}
121 120, QS L-Front_Fricative{*^f-,*^v-}
122 121, QS L-Affricate_Consonant{*^dz-,*^drz-,*^dzi-,*^c-,*^cz-,*^ci-}
123 122, QS L-silences{*^pau-}
124 123, QS L-schwa{*^schwa-}
125 124, QS L-a{*^a-}
126 125, QS L-e{*^e-}
127 126, QS L-i{*^i-}
128 127, QS L-y{*^y-}
129 128, QS L-o{*^o-}
130 129, QS L-u{*^u-}
131 130, QS L-p{*^p-}
132 131, QS L-b{*^b-}
133 132, QS L-t{*^t-}
134 133, QS L-d{*^d-}
135 134, QS L-k{*^k-}
136 135, QS L-ki{*^ki-}
137 136, QS L-g{*^g-}
138 137, QS L-gi{*^gi-}
139 138, QS L-f{*^f-}
140 139, QS L-v{*^v-}
141 140, QS L-s{*^s-}
142 141, QS L-si{*^si-}
143 142, QS L-z{*^z-}
144 143, QS L-zi{*^zi-}
145 144, QS L-sz{*^sz-}
146 145, QS L-rz{*^rz-}
147 146, QS L-x{*^x-}
148 147, QS L-c{*^c-}
149 148, QS L-dz{*^dz-}
150 149, QS L-cz{*^cz-}

151 150, QS L-drz{*^drz-}
152 151, QS L-ci{*^ci-}
153 152, QS L-dzi{*^dzi-}
154 153, QS L-m{*^m-}
155 154, QS L-n{*^n-}
156 155, QS L-ni{*^ni-}
157 156, QS L-ng{*^ng-}
158 157, QS L-l{*^l-}
159 158, QS L-r{*^r-}
160 159, QS L-w{*^w-}
161 160, QS L-ww{*^ww-}
162 161, QS L-j{*^j-}
163 162, QS L-jj{*^jj-}
164 163, QS L-gs{*^gs-}
165 164, QS C-Vowel{*schwa+,*i+,*y+,*e+,*a+,*o+,*u+}
166 165, QS C-Consonant{*gs+,*p+,*b+,*t+,*d+,*k+,*g+,*ki
+,*gi+,*f+,*v+,*s+,*si+,*z+,*zi+,*sz+,*rz+,*
x+,*c+,*dz+,*cz+,*drz+,*ci+,*dzi+,*m+,*n+,*ni
+,*ng+,*l+,*r+,*w+,*ww+,*j+,*jj+}
167 166, QS C-Stop{*p+,*b+,*t+,*d+,*k+,*g+,*gs+}
168 167, QS C-Nasal{*ww+,*jj+,*m+,*n+,*ni+,*ng+}
169 168, QS C-Fricative{*f+,*v+,*s+,*si+,*z+,*zi+,*sz+,*
rz+,*x+}
170 169, QS C-Front{*e+,*i+,*y+,*f+,*v+,*p+,*b+,*m+,*w
+,*ww+}
171 170, QS C-Central{*schwa+,*a+,*t+,*d+,*s+,*si+,*z+,*
zi+,*n+,*r+,*l+,*t+,*d+,*sz+,*rz+,*cz+,*drz
+,*c+,*dz+,*ci+,*dzi+}
172 171, QS C-Back{*o+,*u+,*k+,*g+,*ki+,*gi+,*ng+,*x+,*
gs+}
173 172, QS C-Front_Vowel{*e+,*i+,*y+}
174 173, QS C-Central_Vowel{*a+,*schwa+}
175 174, QS C-Back_Vowel{*o+,*u+}
176 175, QS C-High_Vowel{*i+,*y+,*u+}
177 176, QS C-Medium_Vowel{*e+,*o+}
178 177, QS C-Low_Vowel{*a+}
179 178, QS C-Rounded_Vowel{*o+,*u+}
180 179, QS C-Unrounded_Vowel{*a+,*e+,*i+,*y+}
181 180, QS C-IVowel{*i+}
182 181, QS C-EVowel{*e+}
183 182, QS C-AVowel{*a+}
184 183, QS C-OVowel{*o+}
185 184, QS C-UVowel{*u+}
186 185, QS C-YVowel{*y+}
187 186, QS C-SCHWAVowel{*schwa+}
188 187, QS C-Unvoiced_Consonant{*p+,*t+,*k+,*ki+,*f+,*v+,*
s+,*sz+,*x+,*c+,*cz+,*ci+,*gs+}

189 188, QS C-Voiced_Consonant{*b+*,*d+*,*g+*,*gi+*,*v+*,*z+*,*zi+*,*rz+*,*dz+*,*drz+*,*dzi+*,*m+*,*n+*,*ni+*,*ng+*,*l+*,*r+*,*w+*,*ww+*,*j+*,*jj+*}

190 189, QS C-Front_Consonant{*f+*,*v+*,*f+*,*p+*,*b+*,*m+*,*w+*,*ww+*}

191 190, QS C-Central_Consonant{*t+*,*d+*,*s+*,*si+*,*z+*,*zi+*,*n+*,*r+*,*l+*,*t+*,*d+*,*sz+*,*rz+*,*cz+*,*drz+*,*c+*,*dz+*,*ci+*,*dzi+*}

192 191, QS C-Back_Consonant{*k+*,*g+*,*ki+*,*gi+*,*ng+*,*x+*,*gs+*}

193 192, QS C-Fortis_Consonant{*cz+*,*f+*,*k+*,*p+*,*s+*,*sz+*,*t+*,*ci+*,*c+*,*ki+*,*gs+*}

194 193, QS C-Lenis_Consonant{*drz+*,*v+*,*g+*,*b+*,*rz+*,*z+*,*d+*,*dzi+*,*dz+*,*gi+*,*zi+*}

195 194, QS C-Neither_F_or_L{*m+*,*n+*,*ni+*,*ng+*,*l+*,*r+*,*w+*,*ww+*,*j+*,*jj+*}

196 195, QS C-Voiced_Stop{*b+*,*d+*,*g+*}

197 196, QS C-Unvoiced_Stop{*p+*,*t+*,*k+*,*gs+*}

198 197, QS C-Front_Stop{*b+*,*p+*}

199 198, QS C-Central_Stop{*d+*,*t+*}

200 199, QS C-Back_Stop{*g+*,*k+*,*gs+*}

201 200, QS C-Voiced_Fricative{*v+*,*z+*,*zi+*,*rz+*}

202 201, QS C-Unvoiced_Fricative{*f+*,*s+*,*si+*,*sz+*,*x+*}

203 202, QS C-Front_Fricative{*f+*,*v+*}

204 203, QS C-Affricate_Consonant{*dz+*,*drz+*,*dzi+*,*c+*,*cz+*,*ci+*}

205 204, QS C-silences{*pau+*}

206 205, QS C-schwa{*schwa+*}

207 206, QS C-a{*a+*}

208 207, QS C-e{*e+*}

209 208, QS C-i{*i+*}

210 209, QS C-y{*y+*}

211 210, QS C-o{*o+*}

212 211, QS C-u{*u+*}

213 212, QS C-p{*p+*}

214 213, QS C-b{*b+*}

215 214, QS C-t{*t+*}

216 215, QS C-d{*d+*}

217 216, QS C-k{*k+*}

218 217, QS C-ki{*ki+*}

219 218, QS C-g{*g+*}

220 219, QS C-gi{*gi+*}

221 220, QS C-f{*f+*}

222 221, QS C-v{*v+*}

223 222, QS C-s{*s+*}

224 223, QS C-si{*si+*}

225 224, QS C-z{*z+*}

226 225, QS C-zi{* -zi+*}
227 226, QS C-sz{* -sz+*}
228 227, QS C-rz{* -rz+*}
229 228, QS C-x{* -x+*}
230 229, QS C-c{* -c+*}
231 230, QS C-dz{* -dz+*}
232 231, QS C-cz{* -cz+*}
233 232, QS C-drz{* -drz+*}
234 233, QS C-ci{* -ci+*}
235 234, QS C-dzi{* -dzi+*}
236 235, QS C-m{* -m+*}
237 236, QS C-n{* -n+*}
238 237, QS C-ni{* -ni+*}
239 238, QS C-ng{* -ng+*}
240 239, QS C-l{* -l+*}
241 240, QS C-r{* -r+*}
242 241, QS C-w{* -w+*}
243 242, QS C-ww{* -ww+*}
244 243, QS C-j{* -j+*}
245 244, QS C-jj{* -jj+*}
246 245, QS C-gs{* -gs+*}
247 246, QS R-Vowel{*+i=*,*+y=*,*+e=*,*+a=*,*+o=*,*+u=*,*+schwa=*}
248 247, QS R-Consonant{*+gs=*,*+p=*,*+b=*,*+t=*,*+d=*,*+k=*,*+g=*,*+ki
=*,*+gi=*,*+f=*,*+v=*,*+s=*,*+si=*,*+z=*,*+zi=*,*+sz=*,*+rz=*,*+
x=*,*+c=*,*+dz=*,*+cz=*,*+drz=*,*+ci=*,*+dzi=*,*+m=*,*+n=*,*+ni
=*,*+ng=*,*+l=*,*+r=*,*+w=*,*+ww=*,*+j=*,*+jj=*}
249 248, QS R-Stop{*+p=*,*+b=*,*+t=*,*+d=*,*+k=*,*+g=*,*+gs=*}
250 249, QS R-Nasal{*+ww=*,*+jj=*,*+m=*,*+n=*,*+ni=*,*+ng=*}
251 250, QS R-Fricative{*+f=*,*+v=*,*+s=*,*+si=*,*+z=*,*+zi=*,*+sz=*,*+
rz=*,*+x=*}
252 251, QS R-Front{*+e=*,*+i=*,*+y=*,*+f=*,*+v=*,*+p=*,*+b=*,*+m=*,*+w
=*,*+ww=*}
253 252, QS R-Central{*+schwa=*,*+a=*,*+t=*,*+d=*,*+s=*,*+si=*,*+z=*,*+
zi=*,*+n=*,*+r=*,*+l=*,*+t=*,*+d=*,*+sz=*,*+rz=*,*+cz=*,*+drz
=*,*+c=*,*+dz=*,*+ci=*,*+dzi=*}
254 253, QS R-Back{*+o=*,*+u=*,*+k=*,*+g=*,*+ki=*,*+gi=*,*+ng=*,*+x=*,*+
gs=*}
255 254, QS R-Front_Vowel{*+e=*,*+i=*,*+y=*}
256 255, QS R-Central_Vowel{*+a=*,*+schwa=*}
257 256, QS R-Back_Vowel{*+o=*,*+u=*}
258 257, QS R-High_Vowel{*+i=*,*+y=*,*+u=*}
259 258, QS R-Medium_Vowel{*+e=*,*+o=*}
260 259, QS R-Low_Vowel{*+a=*}
261 260, QS R-Rounded_Vowel{*+o=*,*+u=*}
262 261, QS R-Unrounded_Vowel{*+a=*,*+e=*,*+i=*,*+y=*}
263 262, QS R-IVowel{*+i=*}
264 263, QS R-0Vowel{*+o=*}

```

265 264, QS R-UVowel{**u=*}
266 265, QS R-YVowel{**y=*}
267 266, QS R-SCHWAVowel{**schwa=*}
268 267, QS R-Unvoiced_Consonant{**p=*,**t=*,**k=*,**ki=*,**f=*,**v=*,**
    s=*,**sz=*,**x=*,**c=*,**cz=*,**ci=*,**gs=*}
269 268, QS R-Voiced_Consonant{**b=*,**d=*,**g=*,**gi=*,**v=*,**z=*,**zi
    =*,**rz=*,**dz=*,**drz=*,**dzi=*,**m=*,**n=*,**ni=*,**ng=*,**l
    =*,**r=*,**w=*,**ww=*,**j=*,**jj=*}
270 269, QS R-Front_Consonant{**f=*,**v=*,**f=*,**p=*,**b=*,**m=*,**w
    =*,**ww=*}
271 270, QS R-Central_Consonant{**t=*,**d=*,**s=*,**si=*,**z=*,**zi=*,**
    n=*,**r=*,**l=*,**t=*,**d=*,**sz=*,**rz=*,**cz=*,**drz=*,**c
    =*,**dz=*,**ci=*,**dzi=*}
272 271, QS R-Back_Consonant{**k=*,**g=*,**ki=*,**gi=*,**ng=*,**x=*,**gs
    =*}
273 272, QS R-Fortis_Consonant{**cz=*,**f=*,**k=*,**p=*,**s=*,**sz=*,**t
    =*,**ci=*,**c=*,**ki=*,**gs=*}
274 273, QS R-Lenis_Consonant{**drz=*,**v=*,**g=*,**b=*,**rz=*,**z=*,**d
    =*,**dzi=*,**dz=*,**gi=*,**zi=*}
275 274, QS R-Neigther_F_or_L{**m=*,**n=*,**ni=*,**ng=*,**l=*,**r=*,**w
    =*,**ww=*,**j=*,**jj=*}
276 275, QS R-Voiced_Stop{**b=*,**d=*,**g=*}
277 276, QS R-Unvoiced_Stop{**p=*,**t=*,**k=*,**gs=*}
278 277, QS R-Front_Stop{**b=*,**p=*}
279 278, QS R-Central_Stop{**d=*,**t=*}
280 279, QS R-Back_Stop{**g=*,**k=*,**gs=*}
281 280, QS R-Voiced_Fricative{**v=*,**z=*,**zi=*,**rz=*}
282 281, QS R-Unvoiced_Fricative{**f=*,**s=*,**si=*,**sz=*,**x=*}
283 282, QS R-Front_Fricative{**f=*,**v=*}
284 283, QS R-Affricate_Consonant{**dz=*,**drz=*,**dzi=*,**c=*,**cz=*,**
    ci=*}
285 284, QS R-silences{**pau=*}
286 285, QS R-schwa{**schwa=*}
287 286, QS R-a{**a=*}
288 287, QS R-e{**e=*}
289 288, QS R-i{**i=*}
290 289, QS R-y{**y=*}
291 290, QS R-o{**o=*}
292 291, QS R-u{**u=*}
293 292, QS R-p{**p=*}
294 293, QS R-b{**b=*}
295 294, QS R-t{**t=*}
296 295, QS R-d{**d=*}
297 296, QS R-k{**k=*}
298 297, QS R-ki{**ki=*}
299 298, QS R-g{**g=*}
300 299, QS R-gi{**gi=*}

```


301 300, QS R-f{**f=*}
302 301, QS R-v{**v=*}
303 302, QS R-s{**s=*}
304 303, QS R-si{**si=*}
305 304, QS R-z{**z=*}
306 305, QS R-zi{**zi=*}
307 306, QS R-sz{**sz=*}
308 307, QS R-rz{**rz=*}
309 308, QS R-x{**x=*}
310 309, QS R-c{**c=*}
311 310, QS R-dz{**dz=*}
312 311, QS R-cz{**cz=*}
313 312, QS R-drz{**drz=*}
314 313, QS R-ci{**ci=*}
315 314, QS R-dzi{**dzi=*}
316 315, QS R-m{**m=*}
317 316, QS R-n{**n=*}
318 317, QS R-ni{**ni=*}
319 318, QS R-ng{**ng=*}
320 319, QS R-l{**l=*}
321 320, QS R-r{**r=*}
322 321, QS R-w{**w=*}
323 322, QS R-ww{**ww=*}
324 323, QS R-j{**j=*}
325 324, QS R-jj{**jj=*}
326 325, QS R-gs{**gs=*}
327 326, QS RR-Vowel{*=i@*,*=y@*,*=e@*,*=a@*,*=o@*,*=u@*,*=schwa@*}
328 327, QS RR-Consonant{*=gs@*,*=p@*,*=b@*,*=t@*,*=d@*,*=k@*,*=g@*,*=
ki@*,*=gi@*,*=f@*,*=v@*,*=s@*,*=si@*,*=z@*,*=zi@*,*=sz@*,*=rz@
,=x@*,*=c@*,*=dz@*,*=cz@*,*=drz@*,*=ci@*,*=dzi@*,*=m@*,*=n@
,=ni@*,*=ng@*,*=l@*,*=r@*,*=w@*,*=ww@*,*=j@*,*=jj@*}
329 328, QS RR-Stop{*=gs@*,*=p@*,*=b@*,*=t@*,*=d@*,*=k@*,*=g@*}
330 329, QS RR-Nasal{*=ww@*,*=jj@*,*=m@*,*=n@*,*=ni@*,*=ng@*}
331 330, QS RR-Fricative{*=f@*,*=v@*,*=s@*,*=si@*,*=z@*,*=zi@*,*=sz@*,*=
rz@*,*=x@*}
332 331, QS RR-Front{*=e@*,*=i@*,*=y@*,*=f@*,*=v@*,*=p@*,*=b@*,*=m@*,*=
w@*,*=ww@*}
333 332, QS RR-Central{*=schwa@*,*=a@*,*=t@*,*=d@*,*=s@*,*=si@*,*=z@*,*=
zi@*,*=n@*,*=r@*,*=l@*,*=t@*,*=d@*,*=sz@*,*=rz@*,*=cz@*,*=drz@
,=c@*,*=dz@*,*=ci@*,*=dzi@*}
334 333, QS RR-Back{*=o@*,*=u@*,*=k@*,*=g@*,*=ki@*,*=gi@*,*=ng@*,*=x@
,=gs@*}
335 334, QS RR-Front_Vowel{*=e@*,*=i@*,*=y@*}
336 335, QS RR-Central_Vowel{*=a@*,*=schwa@*}
337 336, QS RR-Back_Vowel{*=o@*,*=u@*}
338 337, QS RR-High_Vowel{*=i@*,*=y@*,*=u@*}
339 338, QS RR-Medium_Vowel{*=e@*,*=o@*}

```

340 339, QS RR-Low_Vowel{*=a@*}
341 340, QS RR-Rounded_Vowel{*=o@*,*=u@*}
342 341, QS RR-Unrounded_Vowel{*=a@*,*=e@*,*=i@*,*=y@*}
343 342, QS RR-IVowel{*=i@*}
344 343, QS RR-OVowel{*=o@*}
345 344, QS RR-UVowel{*=u@*}
346 345, QS RR-YVowel{*=y@*}
347 346, QS RR-SCHWAVowel{*=schwa@*}
348 347, QS RR-Unvoiced_Consonant{*=p@*,*=t@*,*=k@*,*=ki@*,*=f@*,*=v@
    *,*=s@*,*=sz@*,*=x@*,*=c@*,*=cz@*,*=ci@*,*=gs@*}
349 348, QS RR-Voiced_Consonant{*=b@*,*=d@*,*=g@*,*=gi@*,*=v@*,*=z@*,*=
    zi@*,*=rz@*,*=dz@*,*=drz@*,*=dzi@*,*=m@*,*=n@*,*=ni@*,*=ng@*,*=
    l@*,*=r@*,*=w@*,*=ww@*,*=j@*,*=jj@*}
350 349, QS RR-Front_Consonant{*=f@*,*=v@*,*=f@*,*=p@*,*=b@*,*=m@*,*=w@
    *,*=ww@*}
351 350, QS RR-Central_Consonant{*=t@*,*=d@*,*=s@*,*=si@*,*=z@*,*=zi@
    *,*=n@*,*=r@*,*=l@*,*=t@*,*=d@*,*=sz@*,*=rz@*,*=cz@*,*=drz@*,*=
    c@*,*=dz@*,*=ci@*,*=dzi@*}
352 351, QS RR-Back_Consonant{*=k@*,*=g@*,*=ki@*,*=gi@*,*=ng@*,*=x@*,*=
    gs@*}
353 352, QS RR-Fortis_Consonant{*=cz@*,*=f@*,*=k@*,*=p@*,*=s@*,*=sz@*,*=
    t@*,*=ci@*,*=c@*,*=ki@*,*=gs@*}
354 353, QS RR-Lenis_Consonant{*=drz@*,*=v@*,*=g@*,*=b@*,*=rz@*,*=z@*,*=
    d@*,*=dzi@*,*=dz@*,*=gi@*,*=zi@*}
355 354, QS RR-Neigther_F_or_L{*=m@*,*=n@*,*=ni@*,*=ng@*,*=l@*,*=r@*,*=
    w@*,*=ww@*,*=j@*,*=jj@*}
356 355, QS RR-Voiced_Stop{*=b@*,*=d@*,*=g@*}
357 356, QS RR-Unvoiced_Stop{*=p@*,*=t@*,*=k@*,*=gs@*}
358 357, QS RR-Front_Stop{*=b@*,*=p@*}
359 358, QS RR-Central_Stop{*=d@*,*=t@*}
360 359, QS RR-Back_Stop{*=g@*,*=k@*,*=gs@*}
361 360, QS RR-Voiced_Fricative{*=v@*,*=z@*,*=zi@*,*=rz@*}
362 361, QS RR-Unvoiced_Fricative{*=f@*,*=s@*,*=si@*,*=sz@*,*=x@*}
363 362, QS RR-Front_Fricative{*=f@*,*=v@*}
364 363, QS RR-Affricate_Consonant{*=dz@*,*=drz@*,*=dzi@*,*=c@*,*=cz@
    *,*=ci@*}
365 364, QS RR-silences{*=pau@*}
366 365, QS RR-a{*=a@*}
367 366, QS RR-e{*=e@*}
368 367, QS RR-i{*=i@*}
369 368, QS RR-y{*=y@*}
370 369, QS RR-o{*=o@*}
371 370, QS RR-u{*=u@*}
372 371, QS RR-p{*=p@*}
373 372, QS RR-b{*=b@*}
374 373, QS RR-t{*=t@*}
375 374, QS RR-d{*=d@*}

```

```

376 375, QS RR-k{*k@*}
377 376, QS RR-ki{*ki@*}
378 377, QS RR-g{*g@*}
379 378, QS RR-gi{*gi@*}
380 379, QS RR-f{*f@*}
381 380, QS RR-v{*v@*}
382 381, QS RR-s{*s@*}
383 382, QS RR-si{*si@*}
384 383, QS RR-z{*z@*}
385 384, QS RR-zi{*zi@*}
386 385, QS RR-sz{*sz@*}
387 386, QS RR-rz{*rz@*}
388 387, QS RR-x{*x@*}
389 388, QS RR-c{*c@*}
390 389, QS RR-dz{*dz@*}
391 390, QS RR-cz{*cz@*}
392 391, QS RR-drz{*drz@*}
393 392, QS RR-ci{*ci@*}
394 393, QS RR-dzi{*dzi@*}
395 394, QS RR-m{*m@*}
396 395, QS RR-n{*n@*}
397 396, QS RR-ni{*ni@*}
398 397, QS RR-ng{*ng@*}
399 398, QS RR-l{*l@*}
400 399, QS RR-r{*r@*}
401 400, QS RR-w{*w@*}
402 401, QS RR-ww{*ww@*}
403 402, QS RR-j{*j@*}
404 403, QS RR-jj{*jj@*}
405 404, QS RR-gs{*gs@*}
406 405, QS Seg_Fw==x{*@x_*}
407 406, QS Seg_Fw==1{*@1_*}
408 407, QS Seg_Fw==2{*@2_*}
409 408, QS Seg_Fw==3{*@3_*}
410 409, QS Seg_Fw==4{*@4_*}
411 410, QS Seg_Fw==5{*@5_*}
412 411, QS Seg_Fw==6{*@6_*}
413 412, QS Seg_Fw==7{*@7_*}
414 413, QS Seg_Fw<=1{*@x_*,*@1_*}
415 414, QS Seg_Fw<=2{*@x_*,*@1_*,*@2_*}
416 415, QS Seg_Fw<=3{*@x_*,*@1_*,*@2_*,*@3_*}
417 416, QS Seg_Fw<=4{*@x_*,*@1_*,*@2_*,*@3_*,*@4_*}
418 417, QS Seg_Fw<=5{*@x_*,*@1_*,*@2_*,*@3_*,*@4_*,*@5_*}
419 418, QS Seg_Fw<=6{*@x_*,*@1_*,*@2_*,*@3_*,*@4_*,*@5_*,*@6_*}
420 419, QS Seg_Fw<=7{*@x_*,*@1_*,*@2_*,*@3_*,*@4_*,*@5_*,*@6_*,*@7_*}
421 420, QS Seg_Bw==x{*_x/A:*}
422 421, QS Seg_Bw==1{*_1/A:*}

```

```

423 422, QS Seg_Bw==2{*_2/A:*}
424 423, QS Seg_Bw==3{*_3/A:*}
425 424, QS Seg_Bw==4{*_4/A:*}
426 425, QS Seg_Bw==5{*_5/A:*}
427 426, QS Seg_Bw==6{*_6/A:*}
428 427, QS Seg_Bw==7{*_7/A:*}
429 428, QS Seg_Bw<=0{*_x/A:*,*_0/A:*}
430 429, QS Seg_Bw<=1{*_x/A:*,*_0/A:*,*_1/A:*}
431 430, QS Seg_Bw<=2{*_x/A:*,*_0/A:*,*_1/A:*,*_2/A:*}
432 431, QS Seg_Bw<=3{*_x/A:*,*_0/A:*,*_1/A:*,*_2/A:*,*_3/A:*}
433 432, QS Seg_Bw<=4{*_x/A:*,*_0/A:*,*_1/A:*,*_2/A:*,*_3/A:*,*_4/A:*}
434 433, QS Seg_Bw<=5{*_x/A:*,*_0/A:*,*_1/A:*,*_2/A:*,*_3/A:*,*_4/A:*,*_
    _5/A:*}
435 434, QS Seg_Bw<=6{*_x/A:*,*_0/A:*,*_1/A:*,*_2/A:*,*_3/A:*,*_4/A:*,*_
    _5/A:*,*_6/A:*}
436 435, QS Seg_Bw<=7{*_x/A:*,*_0/A:*,*_1/A:*,*_2/A:*,*_3/A:*,*_4/A:*,*_
    _5/A:*,*_6/A:*,*_7/A:*}
437 436, QS L-Syl_Stress==1{* /A:1_*}
438 437, QS L-Syl_Stress==0{* /A:0_*}
439 438, QS L-Syl_Accent==1{* _1_*}
440 439, QS L-Syl_Accent==0{* _0_*}
441 440, QS L-Syl_TOBI_Accent-H{* /K:H?/L:*}
442 441, QS L-Syl_TOBI_Accent-L{* /K:L?/L:*}
443 442, QS L-Syl_TOBI_Accent-L*+H{* /K:L?+H*}
444 443, QS L-Syl_TOBI_Accent-L+H{* /K:L+H*}
445 444, QS L-Syl_TOBI_Accent-0{* /K:0*}
446 445, QS L-Syl_TOBI_Accent-NONE{* /K:NONE*}
447 446, QS L-Syl_TOBI_Accent-x{* /K:x*}
448 447, QS L-Syl_Num-Segs==0{* _0/B:*}
449 448, QS L-Syl_Num-Segs==1{* _1/B:*}
450 449, QS L-Syl_Num-Segs==2{* _2/B:*}
451 450, QS L-Syl_Num-Segs==3{* _3/B:*}
452 451, QS L-Syl_Num-Segs==4{* _4/B:*}
453 452, QS L-Syl_Num-Segs==5{* _5/B:*}
454 453, QS L-Syl_Num-Segs==6{* _6/B:*}
455 454, QS L-Syl_Num-Segs==7{* _7/B:*}
456 455, QS L-Syl_Num-Segs<=1{* _0/B:*,*_1/B:*}
457 456, QS L-Syl_Num-Segs<=2{* _0/B:*,*_1/B:*,*_2/B:*}
458 457, QS L-Syl_Num-Segs<=3{* _0/B:*,*_1/B:*,*_2/B:*,*_3/B:*}
459 458, QS L-Syl_Num-Segs<=4{* _0/B:*,*_1/B:*,*_2/B:*,*_3/B:*,*_4/B:*}
460 459, QS L-Syl_Num-Segs<=5{* _0/B:*,*_1/B:*,*_2/B:*,*_3/B:*,*_4/B:*,*_
    _5/B:*}
461 460, QS L-Syl_Num-Segs<=6{* _0/B:*,*_1/B:*,*_2/B:*,*_3/B:*,*_4/B:*,*_
    _5/B:*,*_6/B:*}
462 461, QS L-Syl_Num-Segs<=7{* _0/B:*,*_1/B:*,*_2/B:*,*_3/B:*,*_4/B:*,*_
    _5/B:*,*_6/B:*,*_7/B:*}
463 462, QS C-Syl_Stress==1{* /B:1-*}

```

```

464 463, QS C-Syl_Stress==0{* /B:0-*}
465 464, QS C-Syl_Stress==x{* /B:x-*}
466 465, QS C-Syl_Accent==1{* -1-*}
467 466, QS C-Syl_Accent==0{* -0-*}
468 467, QS C-Syl_Accent==x{* -x-*}
469 468, QS C-Syl_TOBI_Accent-H{* /L:H?/M:*}
470 469, QS C-Syl_TOBI_Accent-L{* /L:L?/M:*}
471 470, QS C-Syl_TOBI_Accent-L*+H{* /L:L?+H*}
472 471, QS C-Syl_TOBI_Accent-L+H{* /L:L+H*}
473 472, QS C-Syl_TOBI_Accent-0{* /L:0*}
474 473, QS C-Syl_TOBI_Accent-NONE{* /L:NONE*}
475 474, QS C-Syl_TOBI_Accent-x{* /L:x*}
476 475, QS C-Syl_Num-Segs==x{* -x@*}
477 476, QS C-Syl_Num-Segs==1{* -1@*}
478 477, QS C-Syl_Num-Segs==2{* -2@*}
479 478, QS C-Syl_Num-Segs==3{* -3@*}
480 479, QS C-Syl_Num-Segs==4{* -4@*}
481 480, QS C-Syl_Num-Segs==5{* -5@*}
482 481, QS C-Syl_Num-Segs==6{* -6@*}
483 482, QS C-Syl_Num-Segs==7{* -7@*}
484 483, QS C-Syl_Num-Segs<=1{* -x@*,* -1@*}
485 484, QS C-Syl_Num-Segs<=2{* -x@*,* -1@*,* -2@*}
486 485, QS C-Syl_Num-Segs<=3{* -x@*,* -1@*,* -2@*,* -3@*}
487 486, QS C-Syl_Num-Segs<=4{* -x@*,* -1@*,* -2@*,* -3@*,* -4@*}
488 487, QS C-Syl_Num-Segs<=5{* -x@*,* -1@*,* -2@*,* -3@*,* -4@*,* -5@*}
489 488, QS C-Syl_Num-Segs<=6{* -x@*,* -1@*,* -2@*,* -3@*,* -4@*,* -5@*,* -6@*}
490 489, QS C-Syl_Num-Segs<=7{* -x@*,* -1@*,* -2@*,* -3@*,* -4@*,* -5@*,* -6@
    *,* -7@*}
491 490, QS Pos_C-Syl_in_C-Word(Fw)==x{* @x-*}
492 491, QS Pos_C-Syl_in_C-Word(Fw)==1{* @1-*}
493 492, QS Pos_C-Syl_in_C-Word(Fw)==2{* @2-*}
494 493, QS Pos_C-Syl_in_C-Word(Fw)==3{* @3-*}
495 494, QS Pos_C-Syl_in_C-Word(Fw)==4{* @4-*}
496 495, QS Pos_C-Syl_in_C-Word(Fw)==5{* @5-*}
497 496, QS Pos_C-Syl_in_C-Word(Fw)==6{* @6-*}
498 497, QS Pos_C-Syl_in_C-Word(Fw)==7{* @7-*}
499 498, QS Pos_C-Syl_in_C-Word(Fw)<=1{* @x-*,* @1-*}
500 499, QS Pos_C-Syl_in_C-Word(Fw)<=2{* @x-*,* @1-*,* @2-*}
501 500, QS Pos_C-Syl_in_C-Word(Fw)<=3{* @x-*,* @1-*,* @2-*,* @3-*}
502 501, QS Pos_C-Syl_in_C-Word(Fw)<=4{* @x-*,* @1-*,* @2-*,* @3-*,* @4-*}
503 502, QS Pos_C-Syl_in_C-Word(Fw)<=5{* @x-*,* @1-*,* @2-*,* @3-*,* @4-*,* @5
    -*}
504 503, QS Pos_C-Syl_in_C-Word(Fw)<=6{* @x-*,* @1-*,* @2-*,* @3-*,* @4-*,* @5
    -*,* @6-*}
505 504, QS Pos_C-Syl_in_C-Word(Fw)<=7{* @x-*,* @1-*,* @2-*,* @3-*,* @4-*,* @5
    -*,* @6-*,* @7-*}
506 505, QS Pos_C-Syl_in_C-Word(Bw)==x{* -x&*}

```

```

507 506, QS Pos_C-Syl_in_C-Word(Bw)==1{* -1&*}
508 507, QS Pos_C-Syl_in_C-Word(Bw)==2{* -2&*}
509 508, QS Pos_C-Syl_in_C-Word(Bw)==3{* -3&*}
510 509, QS Pos_C-Syl_in_C-Word(Bw)==4{* -4&*}
511 510, QS Pos_C-Syl_in_C-Word(Bw)==5{* -5&*}
512 511, QS Pos_C-Syl_in_C-Word(Bw)==6{* -6&*}
513 512, QS Pos_C-Syl_in_C-Word(Bw)==7{* -7&*}
514 513, QS Pos_C-Syl_in_C-Word(Bw)<=1{* -x&*, *-1&*}
515 514, QS Pos_C-Syl_in_C-Word(Bw)<=2{* -x&*, *-1&*, *-2&*}
516 515, QS Pos_C-Syl_in_C-Word(Bw)<=3{* -x&*, *-1&*, *-2&*, *-3&*}
517 516, QS Pos_C-Syl_in_C-Word(Bw)<=4{* -x&*, *-1&*, *-2&*, *-3&*, *-4&*}
518 517, QS Pos_C-Syl_in_C-Word(Bw)<=5{* -x
    &*, *-1&*, *-2&*, *-3&*, *-4&*, *-5&*}
519 518, QS Pos_C-Syl_in_C-Word(Bw)<=6{* -x
    &*, *-1&*, *-2&*, *-3&*, *-4&*, *-5&*, *-6&*}
520 519, QS Pos_C-Syl_in_C-Word(Bw)<=7{* -x
    &*, *-1&*, *-2&*, *-3&*, *-4&*, *-5&*, *-6&*, *-7&*}
521 520, QS Pos_C-Syl_in_C-Phrase(Fw)==x{*&x-*}
522 521, QS Pos_C-Syl_in_C-Phrase(Fw)==1{*&1-*}
523 522, QS Pos_C-Syl_in_C-Phrase(Fw)==2{*&2-*}
524 523, QS Pos_C-Syl_in_C-Phrase(Fw)==3{*&3-*}
525 524, QS Pos_C-Syl_in_C-Phrase(Fw)==4{*&4-*}
526 525, QS Pos_C-Syl_in_C-Phrase(Fw)==5{*&5-*}
527 526, QS Pos_C-Syl_in_C-Phrase(Fw)==6{*&6-*}
528 527, QS Pos_C-Syl_in_C-Phrase(Fw)==7{*&7-*}
529 528, QS Pos_C-Syl_in_C-Phrase(Fw)==8{*&8-*}
530 529, QS Pos_C-Syl_in_C-Phrase(Fw)==9{*&9-*}
531 530, QS Pos_C-Syl_in_C-Phrase(Fw)==10{*&10-*}
532 531, QS Pos_C-Syl_in_C-Phrase(Fw)==11{*&11-*}
533 532, QS Pos_C-Syl_in_C-Phrase(Fw)==12{*&12-*}
534 533, QS Pos_C-Syl_in_C-Phrase(Fw)==13{*&13-*}
535 534, QS Pos_C-Syl_in_C-Phrase(Fw)==14{*&14-*}
536 535, QS Pos_C-Syl_in_C-Phrase(Fw)==15{*&15-*}
537 536, QS Pos_C-Syl_in_C-Phrase(Fw)==16{*&16-*}
538 537, QS Pos_C-Syl_in_C-Phrase(Fw)==17{*&17-*}
539 538, QS Pos_C-Syl_in_C-Phrase(Fw)==18{*&18-*}
540 539, QS Pos_C-Syl_in_C-Phrase(Fw)==19{*&19-*}
541 540, QS Pos_C-Syl_in_C-Phrase(Fw)==20{*&20-*}
542 541, QS Pos_C-Syl_in_C-Phrase(Fw)<=1{*&x-*, *&0-*, *&1-*}
543 542, QS Pos_C-Syl_in_C-Phrase(Fw)<=2{*&x-*, *&0-*, *&1-*, *&2-*}
544 543, QS Pos_C-Syl_in_C-Phrase(Fw)<=3{*&x-*, *&0-*, *&1-*, *&2-*, *&3-*}
545 544, QS Pos_C-Syl_in_C-Phrase(Fw)<=4{*&x
    -*, *&0-*, *&1-*, *&2-*, *&3-*, *&4-*}
546 545, QS Pos_C-Syl_in_C-Phrase(Fw)<=5{*&x
    -*, *&0-*, *&1-*, *&2-*, *&3-*, *&4-*, *&5-*}
547 546, QS Pos_C-Syl_in_C-Phrase(Fw)<=6{*&x
    -*, *&0-*, *&1-*, *&2-*, *&3-*, *&4-*, *&5-*, *&6-*}

```

```

548 547, QS Pos_C-Syl_in_C-Phrase(Fw) <=7{*&x
    -*, *&0-*, *&1-*, *&2-*, *&3-*, *&4-*, *&5-*, *&6-*, *&7-*}
549 548, QS Pos_C-Syl_in_C-Phrase(Fw) <=8{*&x
    -*, *&0-*, *&1-*, *&2-*, *&3-*, *&4-*, *&5-*, *&6-*, *&7-*, *&8-*}
550 549, QS Pos_C-Syl_in_C-Phrase(Fw) <=9{*&?-*}
551 550, QS Pos_C-Syl_in_C-Phrase(Fw) <=10{*&?-*, *&10-*}
552 551, QS Pos_C-Syl_in_C-Phrase(Fw) <=11{*&?-*, *&10-*, *&11-*}
553 552, QS Pos_C-Syl_in_C-Phrase(Fw) <=12{*&?-*, *&10-*, *&11-*, *&12-*}
554 553, QS Pos_C-Syl_in_C-Phrase(Fw)
    <=13{*&?-*, *&10-*, *&11-*, *&12-*, *&13-*}
555 554, QS Pos_C-Syl_in_C-Phrase(Fw)
    <=14{*&?-*, *&10-*, *&11-*, *&12-*, *&13-*, *&14-*}
556 555, QS Pos_C-Syl_in_C-Phrase(Fw)
    <=15{*&?-*, *&10-*, *&11-*, *&12-*, *&13-*, *&14-*, *&15-*}
557 556, QS Pos_C-Syl_in_C-Phrase(Fw)
    <=16{*&?-*, *&10-*, *&11-*, *&12-*, *&13-*, *&14-*, *&15-*, *&16-*}
558 557, QS Pos_C-Syl_in_C-Phrase(Fw)
    <=17{*&?-*, *&10-*, *&11-*, *&12-*, *&13-*, *&14-*, *&15-*, *&16-*, *&17-*}

559 558, QS Pos_C-Syl_in_C-Phrase(Fw)
    <=18{*&?-*, *&10-*, *&11-*, *&12-*, *&13-*, *&14-*, *&15-*, *&16-*, *&17-*, *&18-*}

560 559, QS Pos_C-Syl_in_C-Phrase(Fw) <=19{*&?-*, *&1?-*}
561 560, QS Pos_C-Syl_in_C-Phrase(Fw) <=20{*&?-*, *&1?-*, *&20-*}
562 561, QS Pos_C-Syl_in_C-Phrase(Bw) ==x{* -x#*}
563 562, QS Pos_C-Syl_in_C-Phrase(Bw) ==1{* -1#*}
564 563, QS Pos_C-Syl_in_C-Phrase(Bw) ==2{* -2#*}
565 564, QS Pos_C-Syl_in_C-Phrase(Bw) ==3{* -3#*}
566 565, QS Pos_C-Syl_in_C-Phrase(Bw) ==4{* -4#*}
567 566, QS Pos_C-Syl_in_C-Phrase(Bw) ==5{* -5#*}
568 567, QS Pos_C-Syl_in_C-Phrase(Bw) ==6{* -6#*}
569 568, QS Pos_C-Syl_in_C-Phrase(Bw) ==7{* -7#*}
570 569, QS Pos_C-Syl_in_C-Phrase(Bw) ==8{* -8#*}
571 570, QS Pos_C-Syl_in_C-Phrase(Bw) ==9{* -9#*}
572 571, QS Pos_C-Syl_in_C-Phrase(Bw) ==10{* -10#*}
573 572, QS Pos_C-Syl_in_C-Phrase(Bw) ==11{* -11#*}
574 573, QS Pos_C-Syl_in_C-Phrase(Bw) ==12{* -12#*}
575 574, QS Pos_C-Syl_in_C-Phrase(Bw) ==13{* -13#*}
576 575, QS Pos_C-Syl_in_C-Phrase(Bw) ==14{* -14#*}
577 576, QS Pos_C-Syl_in_C-Phrase(Bw) ==15{* -15#*}
578 577, QS Pos_C-Syl_in_C-Phrase(Bw) ==16{* -16#*}
579 578, QS Pos_C-Syl_in_C-Phrase(Bw) ==17{* -17#*}
580 579, QS Pos_C-Syl_in_C-Phrase(Bw) ==18{* -18#*}
581 580, QS Pos_C-Syl_in_C-Phrase(Bw) ==19{* -19#*}
582 581, QS Pos_C-Syl_in_C-Phrase(Bw) ==20{* -20#*}
583 582, QS Pos_C-Syl_in_C-Phrase(Bw) <=1{* -x#*, *-1#*}
584 583, QS Pos_C-Syl_in_C-Phrase(Bw) <=2{* -x#*, *-1#*, *-2#*}

```

```

585 584, QS Pos_C-Syl_in_C-Phrase(Bw) <=3{*x##, *-1##, *-2##, *-3##}
586 585, QS Pos_C-Syl_in_C-Phrase(Bw) <=4{*x##, *-1##, *-2##, *-3##, *-4##}
587 586, QS Pos_C-Syl_in_C-Phrase(Bw) <=5{*x
    ##, *-1##, *-2##, *-3##, *-4##, *-5##}
588 587, QS Pos_C-Syl_in_C-Phrase(Bw) <=6{*x
    ##, *-1##, *-2##, *-3##, *-4##, *-5##, *-6##}
589 588, QS Pos_C-Syl_in_C-Phrase(Bw) <=7{*x
    ##, *-1##, *-2##, *-3##, *-4##, *-5##, *-6##, *-7##}
590 589, QS Pos_C-Syl_in_C-Phrase(Bw) <=8{*x
    ##, *-1##, *-2##, *-3##, *-4##, *-5##, *-6##, *-7##, *-8##}
591 590, QS Pos_C-Syl_in_C-Phrase(Bw) <=9{*?##}
592 591, QS Pos_C-Syl_in_C-Phrase(Bw) <=10{*?##, *-10##}
593 592, QS Pos_C-Syl_in_C-Phrase(Bw) <=11{*?##, *-10##, *-11##}
594 593, QS Pos_C-Syl_in_C-Phrase(Bw) <=12{*?##, *-10##, *-11##, *-12##}
595 594, QS Pos_C-Syl_in_C-Phrase(Bw)
    <=13{*?##, *-10##, *-11##, *-12##, *-13##}
596 595, QS Pos_C-Syl_in_C-Phrase(Bw)
    <=14{*?##, *-10##, *-11##, *-12##, *-13##, *-14##}
597 596, QS Pos_C-Syl_in_C-Phrase(Bw)
    <=15{*?##, *-10##, *-11##, *-12##, *-13##, *-14##, *-15##}
598 597, QS Pos_C-Syl_in_C-Phrase(Bw)
    <=16{*?##, *-10##, *-11##, *-12##, *-13##, *-14##, *-15##, *-16##}
599 598, QS Pos_C-Syl_in_C-Phrase(Bw)
    <=17{*?##, *-10##, *-11##, *-12##, *-13##, *-14##, *-15##, *-16##, *-17##}

600 599, QS Pos_C-Syl_in_C-Phrase(Bw)
    <=18{*?##, *-10##, *-11##, *-12##, *-13##, *-14##, *-15##, *-16##, *-17##, *-18##}

601 600, QS Pos_C-Syl_in_C-Phrase(Bw) <=19{*?##, *-1?##}
602 601, QS Pos_C-Syl_in_C-Phrase(Bw) <=20{*?##, *-1?##, *-20##}
603 602, QS Num-StressedSyl_before_C-Syl_in_C-Phrase==x{*#x-*}
604 603, QS Num-StressedSyl_before_C-Syl_in_C-Phrase==1{*#1-*}
605 604, QS Num-StressedSyl_before_C-Syl_in_C-Phrase==2{*#2-*}
606 605, QS Num-StressedSyl_before_C-Syl_in_C-Phrase==3{*#3-*}
607 606, QS Num-StressedSyl_before_C-Syl_in_C-Phrase==4{*#4-*}
608 607, QS Num-StressedSyl_before_C-Syl_in_C-Phrase==5{*#5-*}
609 608, QS Num-StressedSyl_before_C-Syl_in_C-Phrase==6{*#6-*}
610 609, QS Num-StressedSyl_before_C-Syl_in_C-Phrase==7{*#7-*}
611 610, QS Num-StressedSyl_before_C-Syl_in_C-Phrase==8{*#8-*}
612 611, QS Num-StressedSyl_before_C-Syl_in_C-Phrase==9{*#9-*}
613 612, QS Num-StressedSyl_before_C-Syl_in_C-Phrase==10{*#10-*}
614 613, QS Num-StressedSyl_before_C-Syl_in_C-Phrase==11{*#11-*}
615 614, QS Num-StressedSyl_before_C-Syl_in_C-Phrase==12{*#12-*}
616 615, QS Num-StressedSyl_before_C-Syl_in_C-Phrase <=1{*#x-*, *#1-*}
617 616, QS Num-StressedSyl_before_C-Syl_in_C-Phrase <=2{*#x
    -*, *#1-*, *#2-*}

```



```

618 617, QS Num-StressedSyl_before_C-Syl_in_C-Phrase<=3{##x
    -*,##1-*,##2-*,##3-*}
619 618, QS Num-StressedSyl_before_C-Syl_in_C-Phrase<=4{##x
    -*,##1-*,##2-*,##3-*,##4-*}
620 619, QS Num-StressedSyl_before_C-Syl_in_C-Phrase<=5{##x
    -*,##1-*,##2-*,##3-*,##4-*,##5-*}
621 620, QS Num-StressedSyl_before_C-Syl_in_C-Phrase<=6{##x
    -*,##1-*,##2-*,##3-*,##4-*,##5-*,##6-*}
622 621, QS Num-StressedSyl_before_C-Syl_in_C-Phrase<=7{##x
    -*,##1-*,##2-*,##3-*,##4-*,##5-*,##6-*,##7-*}
623 622, QS Num-StressedSyl_before_C-Syl_in_C-Phrase<=8{##x
    -*,##1-*,##2-*,##3-*,##4-*,##5-*,##6-*,##7-*,##8-*}
624 623, QS Num-StressedSyl_before_C-Syl_in_C-Phrase<=9{##?-*}
625 624, QS Num-StressedSyl_before_C-Syl_in_C-Phrase<=10{##?-*,##10-*}
626 625, QS Num-StressedSyl_before_C-Syl_in_C-Phrase
    <=11{##?-*,##10-*,##11-*}
627 626, QS Num-StressedSyl_before_C-Syl_in_C-Phrase
    <=12{##?-*,##10-*,##11-*,##12-*}
628 627, QS Num-StressedSyl_after_C-Syl_in_C-Phrase==x{*-x$*}
629 628, QS Num-StressedSyl_after_C-Syl_in_C-Phrase==1{*-1$*}
630 629, QS Num-StressedSyl_after_C-Syl_in_C-Phrase==2{*-2$*}
631 630, QS Num-StressedSyl_after_C-Syl_in_C-Phrase==3{*-3$*}
632 631, QS Num-StressedSyl_after_C-Syl_in_C-Phrase==4{*-4$*}
633 632, QS Num-StressedSyl_after_C-Syl_in_C-Phrase==5{*-5$*}
634 633, QS Num-StressedSyl_after_C-Syl_in_C-Phrase==6{*-6$*}
635 634, QS Num-StressedSyl_after_C-Syl_in_C-Phrase==7{*-7$*}
636 635, QS Num-StressedSyl_after_C-Syl_in_C-Phrase==8{*-8$*}
637 636, QS Num-StressedSyl_after_C-Syl_in_C-Phrase==9{*-9$*}
638 637, QS Num-StressedSyl_after_C-Syl_in_C-Phrase==10{*-10$*}
639 638, QS Num-StressedSyl_after_C-Syl_in_C-Phrase==11{*-11$*}
640 639, QS Num-StressedSyl_after_C-Syl_in_C-Phrase==12{*-12$*}
641 640, QS Num-StressedSyl_after_C-Syl_in_C-Phrase<=1{*-x$,*-1$*}
642 641, QS Num-StressedSyl_after_C-Syl_in_C-Phrase<=2{*-x$,*-1$,*-2$
    *}
643 642, QS Num-StressedSyl_after_C-Syl_in_C-Phrase<=3{*-x$,*-1$,*-2$
    *,*-3$*}
644 643, QS Num-StressedSyl_after_C-Syl_in_C-Phrase<=4{*-x$,*-1$,*-2$
    *,*-3$,*-4$*}
645 644, QS Num-StressedSyl_after_C-Syl_in_C-Phrase<=5{*-x$,*-1$,*-2$
    *,*-3$,*-4$,*-5$*}
646 645, QS Num-StressedSyl_after_C-Syl_in_C-Phrase<=6{*-x$,*-1$,*-2$
    *,*-3$,*-4$,*-5$,*-6$*}
647 646, QS Num-StressedSyl_after_C-Syl_in_C-Phrase<=7{*-x$,*-1$,*-2$
    *,*-3$,*-4$,*-5$,*-6$,*-7$*}
648 647, QS Num-StressedSyl_after_C-Syl_in_C-Phrase<=8{*-x$,*-1$,*-2$
    *,*-3$,*-4$,*-5$,*-6$,*-7$,*-8$*}
649 648, QS Num-StressedSyl_after_C-Syl_in_C-Phrase<=9{*-?*$*}

```

```

650 649, QS Num-StressedSyl_after_C-Syl_in_C-Phrase<=10{*~?*$*,*-10$*}
651 650, QS Num-StressedSyl_after_C-Syl_in_C-Phrase<=11{*~?*$*,*-10$
    *,*-11$*}
652 651, QS Num-StressedSyl_after_C-Syl_in_C-Phrase<=12{*~?*$*,*-10$
    *,*-11$*,*-12$*}
653 652, QS Num-AccentedSyl_before_C-Syl_in_C-Phrase==x{*x-*}
654 653, QS Num-AccentedSyl_before_C-Syl_in_C-Phrase==1{*1-*}
655 654, QS Num-AccentedSyl_before_C-Syl_in_C-Phrase==2{*2-*}
656 655, QS Num-AccentedSyl_before_C-Syl_in_C-Phrase==3{*3-*}
657 656, QS Num-AccentedSyl_before_C-Syl_in_C-Phrase==4{*4-*}
658 657, QS Num-AccentedSyl_before_C-Syl_in_C-Phrase==5{*5-*}
659 658, QS Num-AccentedSyl_before_C-Syl_in_C-Phrase==6{*6-*}
660 659, QS Num-AccentedSyl_before_C-Syl_in_C-Phrase<=1{*x-*,*1-*}
661 660, QS Num-AccentedSyl_before_C-Syl_in_C-Phrase<=2{*x-*,*1-*,*2
    -*}
662 661, QS Num-AccentedSyl_before_C-Syl_in_C-Phrase<=3{*x-*,*1-*,*2
    -*,*3-*}
663 662, QS Num-AccentedSyl_before_C-Syl_in_C-Phrase<=4{*x-*,*1-*,*2
    -*,*3-*,*4-*}
664 663, QS Num-AccentedSyl_before_C-Syl_in_C-Phrase<=5{*x-*,*1-*,*2
    -*,*3-*,*4-*,*5-*}
665 664, QS Num-AccentedSyl_before_C-Syl_in_C-Phrase<=6{*x-*,*1-*,*2
    -*,*3-*,*4-*,*5-*,*6-*}
666 665, QS Num-AccentedSyl_after_C-Syl_in_C-Phrase==x{*~x!*}
667 666, QS Num-AccentedSyl_after_C-Syl_in_C-Phrase==1{*~1!*}
668 667, QS Num-AccentedSyl_after_C-Syl_in_C-Phrase==2{*~2!*}
669 668, QS Num-AccentedSyl_after_C-Syl_in_C-Phrase==3{*~3!*}
670 669, QS Num-AccentedSyl_after_C-Syl_in_C-Phrase==4{*~4!*}
671 670, QS Num-AccentedSyl_after_C-Syl_in_C-Phrase==5{*~5!*}
672 671, QS Num-AccentedSyl_after_C-Syl_in_C-Phrase==6{*~6!*}
673 672, QS Num-AccentedSyl_after_C-Syl_in_C-Phrase==7{*~7!*}
674 673, QS Num-AccentedSyl_after_C-Syl_in_C-Phrase<=1{*~x!*,*~1!*}
675 674, QS Num-AccentedSyl_after_C-Syl_in_C-Phrase<=2{*~x
    !*,*~1!*,*~2!*}
676 675, QS Num-AccentedSyl_after_C-Syl_in_C-Phrase<=3{*~x
    !*,*~1!*,*~2!*,*~3!*}
677 676, QS Num-AccentedSyl_after_C-Syl_in_C-Phrase<=4{*~x
    !*,*~1!*,*~2!*,*~3!*,*~4!*}
678 677, QS Num-AccentedSyl_after_C-Syl_in_C-Phrase<=5{*~x
    !*,*~1!*,*~2!*,*~3!*,*~4!*,*~5!*}
679 678, QS Num-AccentedSyl_after_C-Syl_in_C-Phrase<=6{*~x
    !*,*~1!*,*~2!*,*~3!*,*~4!*,*~5!*,*~6!*}
680 679, QS Num-AccentedSyl_after_C-Syl_in_C-Phrase<=7{*~x
    !*,*~1!*,*~2!*,*~3!*,*~4!*,*~5!*,*~6!*,*~7!*}
681 680, QS Num-Syl_from_prev-StressedSyl==x{*!x-*}
682 681, QS Num-Syl_from_prev-StressedSyl==0{*!0-*}
683 682, QS Num-Syl_from_prev-StressedSyl==1{*!1-*}

```

```

684 683, QS Num-Syl_from_prev-StressedSyl==2{#!2-*}
685 684, QS Num-Syl_from_prev-StressedSyl==3{#!3-*}
686 685, QS Num-Syl_from_prev-StressedSyl==4{#!4-*}
687 686, QS Num-Syl_from_prev-StressedSyl==5{#!5-*}
688 687, QS Num-Syl_from_prev-StressedSyl <=0{#!x-*,*!0-*}
689 688, QS Num-Syl_from_prev-StressedSyl <=1{#!x-*,*!0-*,*!1-*}
690 689, QS Num-Syl_from_prev-StressedSyl <=2{#!x-*,*!0-*,*!1-*,*!2-*}
691 690, QS Num-Syl_from_prev-StressedSyl <=3{#!x
    -*,*!0-*,*!1-*,*!2-*,*!3-*}
692 691, QS Num-Syl_from_prev-StressedSyl <=4{#!x
    -*,*!0-*,*!1-*,*!2-*,*!3-*,*!4-*}
693 692, QS Num-Syl_from_prev-StressedSyl <=5{#!x
    -*,*!0-*,*!1-*,*!2-*,*!3-*,*!4-*,*!5-*}
694 693, QS Num-Syl_from_next-StressedSyl==x{*x-*}
695 694, QS Num-Syl_from_next-StressedSyl==0{*0-*}
696 695, QS Num-Syl_from_next-StressedSyl==1{*1-*}
697 696, QS Num-Syl_from_next-StressedSyl==2{*2-*}
698 697, QS Num-Syl_from_next-StressedSyl==3{*3-*}
699 698, QS Num-Syl_from_next-StressedSyl==4{*4-*}
700 699, QS Num-Syl_from_next-StressedSyl==5{*5-*}
701 700, QS Num-Syl_from_next-StressedSyl <=0{*x-*,*0-*}
702 701, QS Num-Syl_from_next-StressedSyl <=1{*x-*,*0-*,*1-*}
703 702, QS Num-Syl_from_next-StressedSyl <=2{*x-*,*0-*,*1-*,*2-*}
704 703, QS Num-Syl_from_next-StressedSyl <=3{*x
    ;*,*0-*,*1-*,*2-*,*3-*}
705 704, QS Num-Syl_from_next-StressedSyl <=4{*x
    ;*,*0-*,*1-*,*2-*,*3-*,*4-*}
706 705, QS Num-Syl_from_next-StressedSyl <=5{*x
    ;*,*0-*,*1-*,*2-*,*3-*,*4-*,*5-*}
707 706, QS Num-Syl_from_prev-AccentedSyl==x{*x-*}
708 707, QS Num-Syl_from_prev-AccentedSyl==0{*0-*}
709 708, QS Num-Syl_from_prev-AccentedSyl==1{*1-*}
710 709, QS Num-Syl_from_prev-AccentedSyl==2{*2-*}
711 710, QS Num-Syl_from_prev-AccentedSyl==3{*3-*}
712 711, QS Num-Syl_from_prev-AccentedSyl==4{*4-*}
713 712, QS Num-Syl_from_prev-AccentedSyl==5{*5-*}
714 713, QS Num-Syl_from_prev-AccentedSyl==6{*6-*}
715 714, QS Num-Syl_from_prev-AccentedSyl==7{*7-*}
716 715, QS Num-Syl_from_prev-AccentedSyl==8{*8-*}
717 716, QS Num-Syl_from_prev-AccentedSyl==9{*9-*}
718 717, QS Num-Syl_from_prev-AccentedSyl==10{*10-*}
719 718, QS Num-Syl_from_prev-AccentedSyl==11{*11-*}
720 719, QS Num-Syl_from_prev-AccentedSyl==12{*12-*}
721 720, QS Num-Syl_from_prev-AccentedSyl==13{*13-*}
722 721, QS Num-Syl_from_prev-AccentedSyl==14{*14-*}
723 722, QS Num-Syl_from_prev-AccentedSyl==15{*15-*}
724 723, QS Num-Syl_from_prev-AccentedSyl==16{*16-*}

```

```

725 724, QS Num-Syl_from_prev-AccentedSyl <=0{*;x-*,*;0-*}
726 725, QS Num-Syl_from_prev-AccentedSyl <=1{*;x-*,*;0-*,*;1-*}
727 726, QS Num-Syl_from_prev-AccentedSyl <=2{*;x-*,*;0-*,*;1-*,*;2-*}
728 727, QS Num-Syl_from_prev-AccentedSyl <=3{*;x
    -*,*;0-*,*;1-*,*;2-*,*;3-*}
729 728, QS Num-Syl_from_prev-AccentedSyl <=4{*;x
    -*,*;0-*,*;1-*,*;2-*,*;3-*,*;4-*}
730 729, QS Num-Syl_from_prev-AccentedSyl <=5{*;x
    -*,*;0-*,*;1-*,*;2-*,*;3-*,*;4-*,*;5-*}
731 730, QS Num-Syl_from_prev-AccentedSyl <=6{*;x
    -*,*;0-*,*;1-*,*;2-*,*;3-*,*;4-*,*;5-*,*;6-*}
732 731, QS Num-Syl_from_prev-AccentedSyl <=7{*;x
    -*,*;0-*,*;1-*,*;2-*,*;3-*,*;4-*,*;5-*,*;6-*,*;7-*}
733 732, QS Num-Syl_from_prev-AccentedSyl <=8{*;x
    -*,*;0-*,*;1-*,*;2-*,*;3-*,*;4-*,*;5-*,*;6-*,*;7-*,*;8-*}
734 733, QS Num-Syl_from_prev-AccentedSyl <=9{*;?-*}
735 734, QS Num-Syl_from_prev-AccentedSyl <=10{*;?-*,*;10-*}
736 735, QS Num-Syl_from_prev-AccentedSyl <=11{*;?-*,*;10-*,*;11-*}
737 736, QS Num-Syl_from_prev-AccentedSyl
    <=12{*;?-*,*;10-*,*;11-*,*;12-*}
738 737, QS Num-Syl_from_prev-AccentedSyl
    <=13{*;?-*,*;10-*,*;11-*,*;12-*,*;13-*}
739 738, QS Num-Syl_from_prev-AccentedSyl
    <=14{*;?-*,*;10-*,*;11-*,*;12-*,*;13-*,*;14-*}
740 739, QS Num-Syl_from_prev-AccentedSyl
    <=15{*;?-*,*;10-*,*;11-*,*;12-*,*;13-*,*;14-*,*;15-*}
741 740, QS Num-Syl_from_prev-AccentedSyl
    <=16{*;?-*,*;10-*,*;11-*,*;12-*,*;13-*,*;14-*,*;15-*,*;16-*}
742 741, QS Num-Syl_from_next-AccentedSyl==x{* -x|*}
743 742, QS Num-Syl_from_next-AccentedSyl==0{* -0|*}
744 743, QS Num-Syl_from_next-AccentedSyl==1{* -1|*}
745 744, QS Num-Syl_from_next-AccentedSyl==2{* -2|*}
746 745, QS Num-Syl_from_next-AccentedSyl==3{* -3|*}
747 746, QS Num-Syl_from_next-AccentedSyl==4{* -4|*}
748 747, QS Num-Syl_from_next-AccentedSyl==5{* -5|*}
749 748, QS Num-Syl_from_next-AccentedSyl==6{* -6|*}
750 749, QS Num-Syl_from_next-AccentedSyl==7{* -7|*}
751 750, QS Num-Syl_from_next-AccentedSyl==8{* -8|*}
752 751, QS Num-Syl_from_next-AccentedSyl==9{* -9|*}
753 752, QS Num-Syl_from_next-AccentedSyl==10{* -10|*}
754 753, QS Num-Syl_from_next-AccentedSyl==11{* -11|*}
755 754, QS Num-Syl_from_next-AccentedSyl==12{* -12|*}
756 755, QS Num-Syl_from_next-AccentedSyl==13{* -13|*}
757 756, QS Num-Syl_from_next-AccentedSyl==14{* -14|*}
758 757, QS Num-Syl_from_next-AccentedSyl==15{* -15|*}
759 758, QS Num-Syl_from_next-AccentedSyl==16{* -16|*}
760 759, QS Num-Syl_from_next-AccentedSyl <=0{* -x|*,*-0|*}

```

```

761 760, QS Num-Syl_from_next-AccentedSyl <=1{* -x|* ,*-0|* ,*-1|*}
762 761, QS Num-Syl_from_next-AccentedSyl <=2{* -x|* ,*-0|* ,*-1|* ,*-2|*}
763 762, QS Num-Syl_from_next-AccentedSyl <=3{* -x
    |* ,*-0|* ,*-1|* ,*-2|* ,*-3|*}
764 763, QS Num-Syl_from_next-AccentedSyl <=4{* -x
    |* ,*-0|* ,*-1|* ,*-2|* ,*-3|* ,*-4|*}
765 764, QS Num-Syl_from_next-AccentedSyl <=5{* -x
    |* ,*-0|* ,*-1|* ,*-2|* ,*-3|* ,*-4|* ,*-5|*}
766 765, QS Num-Syl_from_next-AccentedSyl <=6{* -x
    |* ,*-0|* ,*-1|* ,*-2|* ,*-3|* ,*-4|* ,*-5|* ,*-6|*}
767 766, QS Num-Syl_from_next-AccentedSyl <=7{* -x
    |* ,*-0|* ,*-1|* ,*-2|* ,*-3|* ,*-4|* ,*-5|* ,*-6|* ,*-7|*}
768 767, QS Num-Syl_from_next-AccentedSyl <=8{* -x
    |* ,*-0|* ,*-1|* ,*-2|* ,*-3|* ,*-4|* ,*-5|* ,*-6|* ,*-7|* ,*-8|*}
769 768, QS Num-Syl_from_next-AccentedSyl <=9{* -?|*}
770 769, QS Num-Syl_from_next-AccentedSyl <=10{* -?|* ,*-10|*}
771 770, QS Num-Syl_from_next-AccentedSyl <=11{* -?|* ,*-10|* ,*-11|*}
772 771, QS Num-Syl_from_next-AccentedSyl
    <=12{* -?|* ,*-10|* ,*-11|* ,*-12|*}
773 772, QS Num-Syl_from_next-AccentedSyl
    <=13{* -?|* ,*-10|* ,*-11|* ,*-12|* ,*-13|*}
774 773, QS Num-Syl_from_next-AccentedSyl
    <=14{* -?|* ,*-10|* ,*-11|* ,*-12|* ,*-13|* ,*-14|*}
775 774, QS Num-Syl_from_next-AccentedSyl
    <=15{* -?|* ,*-10|* ,*-11|* ,*-12|* ,*-13|* ,*-14|* ,*-15|*}
776 775, QS Num-Syl_from_next-AccentedSyl
    <=16{* -?|* ,*-10|* ,*-11|* ,*-12|* ,*-13|* ,*-14|* ,*-15|* ,*-16|*}
777 776, QS C-Syl_Vowel==x{*|x/C:*}
778 777, QS C-Syl_Vowel{*|i/C:* ,*|y/C:* ,*|e/C:* ,*|a/C:* ,*|o/C:* ,*|u/C
    :* ,*|schwa/C:*}
779 778, QS C-Syl_Front_Vowel{*|e/C:* ,*|i/C:* ,*|y/C:*}
780 779, QS C-Syl_Central_Vowel{*|a/C:* ,*|schwa/C:*}
781 780, QS C-Syl_Back_Vowel{*|o/C:* ,*|u/C:*}
782 781, QS C-Syl_High_Vowel{*|i/C:* ,*|y/C:* ,*|u/C:*}
783 782, QS C-Syl_Medium_Vowel{*|e/C:* ,*|o/C:* ,*|schwa/C:*}
784 783, QS C-Syl_Low_Vowel{*|a/C:*}
785 784, QS C-Syl_Rounded_Vowel{*|o/C:* ,*|u/C:*}
786 785, QS C-Syl_Unrounded_Vowel{*|a/C:* ,*|e/C:* ,*|i/C:* ,*|y/C:*}
787 786, QS C-Syl_IVowel{*|i/C:*}
788 787, QS C-Syl_EVowel{*|e/C:*}
789 788, QS C-Syl_AVowel{*|a/C:*}
790 789, QS C-Syl_OVowel{*|o/C:*}
791 790, QS C-Syl_UVowel{*|u/C:*}
792 791, QS C-Syl_YVowel{*|y/C:*}
793 792, QS C-Syl_SCHWAVowel{*|schwa/C:*}
794 793, QS R-Syl_Stress==1{* /C:1+*}
795 794, QS R-Syl_Stress==0{* /C:0+*}

```

```

796 795, QS R-Syl_Accent==1{**1+*}
797 796, QS R-Syl_Accent==0{**0+*}
798 797, QS R-Syl_TOBI_Accent-H{* /M:H?}
799 798, QS R-Syl_TOBI_Accent-L{* /M:L?}
800 799, QS R-Syl_TOBI_Accent-L**H{* /M:L?+H}
801 800, QS R-Syl_TOBI_Accent-L+H{* /M:L+H?}
802 801, QS R-Syl_TOBI_Accent-0{* /M:0}
803 802, QS R-Syl_TOBI_Accent-NONE{* /M:NONE}
804 803, QS R-Syl_TOBI_Accent-x{* /M:x}
805 804, QS R-Syl_Num-Segs==0{**0/D:*}
806 805, QS R-Syl_Num-Segs==1{**1/D:*}
807 806, QS R-Syl_Num-Segs==2{**2/D:*}
808 807, QS R-Syl_Num-Segs==3{**3/D:*}
809 808, QS R-Syl_Num-Segs==4{**4/D:*}
810 809, QS R-Syl_Num-Segs==5{**5/D:*}
811 810, QS R-Syl_Num-Segs==6{**6/D:*}
812 811, QS R-Syl_Num-Segs==7{**7/D:*}
813 812, QS R-Syl_Num-Segs<=1{**0/D:*,**1/D:*}
814 813, QS R-Syl_Num-Segs<=2{**0/D:*,**1/D:*,**2/D:*}
815 814, QS R-Syl_Num-Segs<=3{**0/D:*,**1/D:*,**2/D:*,**3/D:*}
816 815, QS R-Syl_Num-Segs<=4{**0/D:*,**1/D:*,**2/D:*,**3/D:*,**4/D:*}
817 816, QS R-Syl_Num-Segs<=5{**0/D:*,**1/D:*,**2/D:*,**3/D:*,**4/D
:*,**5/D:*}
818 817, QS R-Syl_Num-Segs<=6{**0/D:*,**1/D:*,**2/D:*,**3/D:*,**4/D
:*,**5/D:*,**6/D:*}
819 818, QS R-Syl_Num-Segs<=7{**0/D:*,**1/D:*,**2/D:*,**3/D:*,**4/D
:*,**5/D:*,**6/D:*,**7/D:*}
820 819, QS L-Word_Num-Syls==0{* _0/E:*}
821 820, QS L-Word_Num-Syls==1{* _1/E:*}
822 821, QS L-Word_Num-Syls==2{* _2/E:*}
823 822, QS L-Word_Num-Syls==3{* _3/E:*}
824 823, QS L-Word_Num-Syls==4{* _4/E:*}
825 824, QS L-Word_Num-Syls==5{* _5/E:*}
826 825, QS L-Word_Num-Syls==6{* _6/E:*}
827 826, QS L-Word_Num-Syls==7{* _7/E:*}
828 827, QS L-Word_Num-Syls<=1{* _0/E:*,*_1/E:*}
829 828, QS L-Word_Num-Syls<=2{* _0/E:*,*_1/E:*,*_2/E:*}
830 829, QS L-Word_Num-Syls<=3{* _0/E:*,*_1/E:*,*_2/E:*,*_3/E:*}
831 830, QS L-Word_Num-Syls<=4{* _0/E:*,*_1/E:*,*_2/E:*,*_3/E:*,*_4/E:*}
832 831, QS L-Word_Num-Syls<=5{* _0/E:*,*_1/E:*,*_2/E:*,*_3/E:*,*_4/E:*,*
_5/E:*}
833 832, QS L-Word_Num-Syls<=6{* _0/E:*,*_1/E:*,*_2/E:*,*_3/E:*,*_4/E:*,*
_5/E:*,*_6/E:*}
834 833, QS L-Word_Num-Syls<=7{* _0/E:*,*_1/E:*,*_2/E:*,*_3/E:*,*_4/E:*,*
_5/E:*,*_6/E:*,*_7/E:*}
835 834, QS C-Word_Num-Syls==x{**x@*}
836 835, QS C-Word_Num-Syls==1{**1@*}

```

```

837 836, QS C-Word_Num-Syls==2{*+2@*}
838 837, QS C-Word_Num-Syls==3{*+3@*}
839 838, QS C-Word_Num-Syls==4{*+4@*}
840 839, QS C-Word_Num-Syls==5{*+5@*}
841 840, QS C-Word_Num-Syls==6{*+6@*}
842 841, QS C-Word_Num-Syls==7{*+7@*}
843 842, QS C-Word_Num-Syls<=1{*+x@*,**+1@*}
844 843, QS C-Word_Num-Syls<=2{*+x@*,**+1@*,**+2@*}
845 844, QS C-Word_Num-Syls<=3{*+x@*,**+1@*,**+2@*,**+3@*}
846 845, QS C-Word_Num-Syls<=4{*+x@*,**+1@*,**+2@*,**+3@*,**+4@*}
847 846, QS C-Word_Num-Syls<=5{*+x@*,**+1@*,**+2@*,**+3@*,**+4@*,**+5@*}
848 847, QS C-Word_Num-Syls<=6{*+x@*,**+1@*,**+2@*,**+3@*,**+4@*,**+5@*,**+6@
    *}
849 848, QS C-Word_Num-Syls<=7{*+x@*,**+1@*,**+2@*,**+3@*,**+4@*,**+5@*,**+6@
    *,**+7@*}
850 849, QS Pos_C-Word_in_C-Phrase(Fw)==x{*@x+*}
851 850, QS Pos_C-Word_in_C-Phrase(Fw)==1{*@1+*}
852 851, QS Pos_C-Word_in_C-Phrase(Fw)==2{*@2+*}
853 852, QS Pos_C-Word_in_C-Phrase(Fw)==3{*@3+*}
854 853, QS Pos_C-Word_in_C-Phrase(Fw)==4{*@4+*}
855 854, QS Pos_C-Word_in_C-Phrase(Fw)==5{*@5+*}
856 855, QS Pos_C-Word_in_C-Phrase(Fw)==6{*@6+*}
857 856, QS Pos_C-Word_in_C-Phrase(Fw)==7{*@7+*}
858 857, QS Pos_C-Word_in_C-Phrase(Fw)==8{*@8+*}
859 858, QS Pos_C-Word_in_C-Phrase(Fw)==9{*@9+*}
860 859, QS Pos_C-Word_in_C-Phrase(Fw)==10{*@10+*}
861 860, QS Pos_C-Word_in_C-Phrase(Fw)==11{*@11+*}
862 861, QS Pos_C-Word_in_C-Phrase(Fw)==12{*@12+*}
863 862, QS Pos_C-Word_in_C-Phrase(Fw)==13{*@13+*}
864 863, QS Pos_C-Word_in_C-Phrase(Fw)<=1{*@x+*,*@1+*}
865 864, QS Pos_C-Word_in_C-Phrase(Fw)<=2{*@x+*,*@1+*,*@2+*}
866 865, QS Pos_C-Word_in_C-Phrase(Fw)<=3{*@x+*,*@1+*,*@2+*,*@3+*}
867 866, QS Pos_C-Word_in_C-Phrase(Fw)<=4{*@x+*,*@1+*,*@2+*,*@3+*,*@4+*}
868 867, QS Pos_C-Word_in_C-Phrase(Fw)<=5{*@x+*,*@1+*,*@2+*,*@3+*,*@4
    +*,*@5+*}
869 868, QS Pos_C-Word_in_C-Phrase(Fw)<=6{*@x+*,*@1+*,*@2+*,*@3+*,*@4
    +*,*@5+*,*@6+*}
870 869, QS Pos_C-Word_in_C-Phrase(Fw)<=7{*@x+*,*@1+*,*@2+*,*@3+*,*@4
    +*,*@5+*,*@6+*,*@7+*}
871 870, QS Pos_C-Word_in_C-Phrase(Fw)<=8{*@x+*,*@1+*,*@2+*,*@3+*,*@4
    +*,*@5+*,*@6+*,*@7+*,*@8+*}
872 871, QS Pos_C-Word_in_C-Phrase(Fw)<=9{*@?+*}
873 872, QS Pos_C-Word_in_C-Phrase(Fw)<=10{*@?+*,*@10+*}
874 873, QS Pos_C-Word_in_C-Phrase(Fw)<=11{*@?+*,*@10+*,*@11+*}
875 874, QS Pos_C-Word_in_C-Phrase(Fw)<=12{*@?+*,*@10+*,*@11+*,*@12+*}
876 875, QS Pos_C-Word_in_C-Phrase(Fw)<=13{*@?+*,*@10+*,*@11+*,*@12+*,*
    @13+*}

```



```

877 876, QS Pos_C-Word_in_C-Phrase(Bw)==x{**x**}
878 877, QS Pos_C-Word_in_C-Phrase(Bw)==0{**0**}
879 878, QS Pos_C-Word_in_C-Phrase(Bw)==1{**1**}
880 879, QS Pos_C-Word_in_C-Phrase(Bw)==2{**2**}
881 880, QS Pos_C-Word_in_C-Phrase(Bw)==3{**3**}
882 881, QS Pos_C-Word_in_C-Phrase(Bw)==4{**4**}
883 882, QS Pos_C-Word_in_C-Phrase(Bw)==5{**5**}
884 883, QS Pos_C-Word_in_C-Phrase(Bw)==6{**6**}
885 884, QS Pos_C-Word_in_C-Phrase(Bw)==7{**7**}
886 885, QS Pos_C-Word_in_C-Phrase(Bw)==8{**8**}
887 886, QS Pos_C-Word_in_C-Phrase(Bw)==9{**9**}
888 887, QS Pos_C-Word_in_C-Phrase(Bw)==10{**10**}
889 888, QS Pos_C-Word_in_C-Phrase(Bw)==11{**11**}
890 889, QS Pos_C-Word_in_C-Phrase(Bw)==12{**12**}
891 890, QS Pos_C-Word_in_C-Phrase(Bw)==13{**13**}
892 891, QS Pos_C-Word_in_C-Phrase(Bw)<=1{**x**,**0**,**1**}
893 892, QS Pos_C-Word_in_C-Phrase(Bw)<=2{**x**,**0**,**1**,**2**}
894 893, QS Pos_C-Word_in_C-Phrase(Bw)<=3{**x**,**0**,**1**,**2**,**3**}
895 894, QS Pos_C-Word_in_C-Phrase(Bw)<=4{**x
    **,**0**,**1**,**2**,**3**,**4**}
896 895, QS Pos_C-Word_in_C-Phrase(Bw)<=5{**x
    **,**0**,**1**,**2**,**3**,**4**,**5**}
897 896, QS Pos_C-Word_in_C-Phrase(Bw)<=6{**x
    **,**0**,**1**,**2**,**3**,**4**,**5**,**6**}
898 897, QS Pos_C-Word_in_C-Phrase(Bw)<=7{**x
    **,**0**,**1**,**2**,**3**,**4**,**5**,**6**,**7**}
899 898, QS Pos_C-Word_in_C-Phrase(Bw)<=8{**x
    **,**0**,**1**,**2**,**3**,**4**,**5**,**6**,**7**,**8**}
900 899, QS Pos_C-Word_in_C-Phrase(Bw)<=9{**?***}
901 900, QS Pos_C-Word_in_C-Phrase(Bw)<=10{**?***,**10**}
902 901, QS Pos_C-Word_in_C-Phrase(Bw)<=11{**?***,**10**,**11**}
903 902, QS Pos_C-Word_in_C-Phrase(Bw)<=12{**?***,**10**,**11**,**12**}
904 903, QS Pos_C-Word_in_C-Phrase(Bw)
    <=13{**?***,**10**,**11**,**12**,**13**}
905 904, QS Num-ContWord_before_C-Word_in_C-Phrase==x{**x**}
906 905, QS Num-ContWord_before_C-Word_in_C-Phrase==1{**1**}
907 906, QS Num-ContWord_before_C-Word_in_C-Phrase==2{**2**}
908 907, QS Num-ContWord_before_C-Word_in_C-Phrase==3{**3**}
909 908, QS Num-ContWord_before_C-Word_in_C-Phrase==4{**4**}
910 909, QS Num-ContWord_before_C-Word_in_C-Phrase==5{**5**}
911 910, QS Num-ContWord_before_C-Word_in_C-Phrase==6{**6**}
912 911, QS Num-ContWord_before_C-Word_in_C-Phrase==7{**7**}
913 912, QS Num-ContWord_before_C-Word_in_C-Phrase==8{**8**}
914 913, QS Num-ContWord_before_C-Word_in_C-Phrase==9{**9**}
915 914, QS Num-ContWord_before_C-Word_in_C-Phrase<=1{**x**,**1**}
916 915, QS Num-ContWord_before_C-Word_in_C-Phrase<=2{**x**,**1**,**2**}

```



```

917 916, QS Num-ContWord_before_C-Word_in_C-Phrase <=3{*&x
    +*, *&1+*, *&2+*, *&3+*}
918 917, QS Num-ContWord_before_C-Word_in_C-Phrase <=4{*&x
    +*, *&1+*, *&2+*, *&3+*, *&4+*}
919 918, QS Num-ContWord_before_C-Word_in_C-Phrase <=5{*&x
    +*, *&1+*, *&2+*, *&3+*, *&4+*, *&5+*}
920 919, QS Num-ContWord_before_C-Word_in_C-Phrase <=6{*&x
    +*, *&1+*, *&2+*, *&3+*, *&4+*, *&5+*, *&6+*}
921 920, QS Num-ContWord_before_C-Word_in_C-Phrase <=7{*&x
    +*, *&1+*, *&2+*, *&3+*, *&4+*, *&5+*, *&6+*, *&7+*}
922 921, QS Num-ContWord_before_C-Word_in_C-Phrase <=8{*&x
    +*, *&1+*, *&2+*, *&3+*, *&4+*, *&5+*, *&6+*, *&7+*, *&8+*}
923 922, QS Num-ContWord_before_C-Word_in_C-Phrase <=9{*&?+*}
924 923, QS Num-ContWord_after_C-Word_in_C-Phrase ==x{*+x#*}
925 924, QS Num-ContWord_after_C-Word_in_C-Phrase ==0{*+0#*}
926 925, QS Num-ContWord_after_C-Word_in_C-Phrase ==1{*+1#*}
927 926, QS Num-ContWord_after_C-Word_in_C-Phrase ==2{*+2#*}
928 927, QS Num-ContWord_after_C-Word_in_C-Phrase ==3{*+3#*}
929 928, QS Num-ContWord_after_C-Word_in_C-Phrase ==4{*+4#*}
930 929, QS Num-ContWord_after_C-Word_in_C-Phrase ==5{*+5#*}
931 930, QS Num-ContWord_after_C-Word_in_C-Phrase ==6{*+6#*}
932 931, QS Num-ContWord_after_C-Word_in_C-Phrase ==7{*+7#*}
933 932, QS Num-ContWord_after_C-Word_in_C-Phrase ==8{*+8#*}
934 933, QS Num-ContWord_after_C-Word_in_C-Phrase <=0{*+x#*, *+0#*}
935 934, QS Num-ContWord_after_C-Word_in_C-Phrase <=1{*+x#*, *+0#*, *+1#*}
936 935, QS Num-ContWord_after_C-Word_in_C-Phrase <=2{*+x
    #*, *+0#*, *+1#*, *+2#*}
937 936, QS Num-ContWord_after_C-Word_in_C-Phrase <=3{*+x
    #*, *+0#*, *+1#*, *+2#*, *+3#*}
938 937, QS Num-ContWord_after_C-Word_in_C-Phrase <=4{*+x
    #*, *+0#*, *+1#*, *+2#*, *+3#*, *+4#*}
939 938, QS Num-ContWord_after_C-Word_in_C-Phrase <=5{*+x
    #*, *+0#*, *+1#*, *+2#*, *+3#*, *+4#*, *+5#*}
940 939, QS Num-ContWord_after_C-Word_in_C-Phrase <=6{*+x
    #*, *+0#*, *+1#*, *+2#*, *+3#*, *+4#*, *+5#*, *+6#*}
941 940, QS Num-ContWord_after_C-Word_in_C-Phrase <=7{*+x
    #*, *+0#*, *+1#*, *+2#*, *+3#*, *+4#*, *+5#*, *+6#*, *+7#*}
942 941, QS Num-ContWord_after_C-Word_in_C-Phrase <=8{*+x
    #*, *+0#*, *+1#*, *+2#*, *+3#*, *+4#*, *+5#*, *+6#*, *+7#*, *+8#*}
943 942, QS Num-Words_from_prev-ContWord ==x{##x+*}
944 943, QS Num-Words_from_prev-ContWord ==0{##0+*}
945 944, QS Num-Words_from_prev-ContWord ==1{##1+*}
946 945, QS Num-Words_from_prev-ContWord ==2{##2+*}
947 946, QS Num-Words_from_prev-ContWord ==3{##3+*}
948 947, QS Num-Words_from_prev-ContWord ==4{##4+*}
949 948, QS Num-Words_from_prev-ContWord ==5{##5+*}
950 949, QS Num-Words_from_prev-ContWord <=0{##x+*, ##0+*}

```

```

951 950, QS Num-Words_from_prev-ContWord<=1{##x+*,##0+*,##1+*}
952 951, QS Num-Words_from_prev-ContWord<=2{##x+*,##0+*,##1+*,##2+*}
953 952, QS Num-Words_from_prev-ContWord<=3{##x
    +*,##0+*,##1+*,##2+*,##3+*}
954 953, QS Num-Words_from_prev-ContWord<=4{##x
    +*,##0+*,##1+*,##2+*,##3+*,##4+*}
955 954, QS Num-Words_from_prev-ContWord<=5{##x
    +*,##0+*,##1+*,##2+*,##3+*,##4+*,##5+*}
956 955, QS Num-Words_from_next-ContWord==x{**x/F:*}
957 956, QS Num-Words_from_next-ContWord==0{**0/F:*}
958 957, QS Num-Words_from_next-ContWord==1{**1/F:*}
959 958, QS Num-Words_from_next-ContWord==2{**2/F:*}
960 959, QS Num-Words_from_next-ContWord==3{**3/F:*}
961 960, QS Num-Words_from_next-ContWord==4{**4/F:*}
962 961, QS Num-Words_from_next-ContWord==5{**5/F:*}
963 962, QS Num-Words_from_next-ContWord<=0{**x/F:*,**0/F:*}
964 963, QS Num-Words_from_next-ContWord<=1{**x/F:*,**0/F:*,**1/F:*}
965 964, QS Num-Words_from_next-ContWord<=2{**x/F:*,**0/F:*,**1/F:*,**2/
    F:*}
966 965, QS Num-Words_from_next-ContWord<=3{**x/F:*,**0/F:*,**1/F:*,**2/
    F:*,**3/F:*}
967 966, QS Num-Words_from_next-ContWord<=4{**x/F:*,**0/F:*,**1/F:*,**2/
    F:*,**3/F:*,**4/F:*}
968 967, QS Num-Words_from_next-ContWord<=5{**x/F:*,**0/F:*,**1/F:*,**2/
    F:*,**3/F:*,**4/F:*,**5/F:*}
969 968, QS R-Word_Num-Syls==0{*_0/G:*}
970 969, QS R-Word_Num-Syls==1{*_1/G:*}
971 970, QS R-Word_Num-Syls==2{*_2/G:*}
972 971, QS R-Word_Num-Syls==3{*_3/G:*}
973 972, QS R-Word_Num-Syls==4{*_4/G:*}
974 973, QS R-Word_Num-Syls==5{*_5/G:*}
975 974, QS R-Word_Num-Syls==6{*_6/G:*}
976 975, QS R-Word_Num-Syls==7{*_7/G:*}
977 976, QS R-Word_Num-Syls<=1{*_0/G:*,*_1/G:*}
978 977, QS R-Word_Num-Syls<=2{*_0/G:*,*_1/G:*,*_2/G:*}
979 978, QS R-Word_Num-Syls<=3{*_0/G:*,*_1/G:*,*_2/G:*,*_3/G:*}
980 979, QS R-Word_Num-Syls<=4{*_0/G:*,*_1/G:*,*_2/G:*,*_3/G:*,*_4/G:*}
981 980, QS R-Word_Num-Syls<=5{*_0/G:*,*_1/G:*,*_2/G:*,*_3/G:*,*_4/G:*,*
    _5/G:*}
982 981, QS R-Word_Num-Syls<=6{*_0/G:*,*_1/G:*,*_2/G:*,*_3/G:*,*_4/G:*,*
    _5/G:*,*_6/G:*}
983 982, QS R-Word_Num-Syls<=7{*_0/G:*,*_1/G:*,*_2/G:*,*_3/G:*,*_4/G:*,*
    _5/G:*,*_6/G:*,*_7/G:*}
984 983, QS L-Phrase_Num-Syls==0{*/G:0_*}
985 984, QS L-Phrase_Num-Syls==1{*/G:1_*}
986 985, QS L-Phrase_Num-Syls==2{*/G:2_*}
987 986, QS L-Phrase_Num-Syls==3{*/G:3_*}

```

```

988 987, QS L-Phrase_Num-Syls==4{*/G:4_*}
989 988, QS L-Phrase_Num-Syls==5{*/G:5_*}
990 989, QS L-Phrase_Num-Syls==6{*/G:6_*}
991 990, QS L-Phrase_Num-Syls==7{*/G:7_*}
992 991, QS L-Phrase_Num-Syls==8{*/G:8_*}
993 992, QS L-Phrase_Num-Syls==9{*/G:9_*}
994 993, QS L-Phrase_Num-Syls==10{*/G:10_*}
995 994, QS L-Phrase_Num-Syls==11{*/G:11_*}
996 995, QS L-Phrase_Num-Syls==12{*/G:12_*}
997 996, QS L-Phrase_Num-Syls==13{*/G:13_*}
998 997, QS L-Phrase_Num-Syls==14{*/G:14_*}
999 998, QS L-Phrase_Num-Syls==15{*/G:15_*}
1000 999, QS L-Phrase_Num-Syls==16{*/G:16_*}
1001 1000, QS L-Phrase_Num-Syls==17{*/G:17_*}
1002 1001, QS L-Phrase_Num-Syls==18{*/G:18_*}
1003 1002, QS L-Phrase_Num-Syls==19{*/G:19_*}
1004 1003, QS L-Phrase_Num-Syls==20{*/G:20_*}
1005 1004, QS L-Phrase_Num-Syls<=1{*/G:0_*,*/G:1_*}
1006 1005, QS L-Phrase_Num-Syls<=2{*/G:0_*,*/G:1_*,*/G:2_*}
1007 1006, QS L-Phrase_Num-Syls<=3{*/G:0_*,*/G:1_*,*/G:2_*,*/G:3_*}
1008 1007, QS L-Phrase_Num-Syls<=4{*/G:0_*,*/G:1_*,*/G:2_*,*/G:3_*,*/G:4_
    *}
1009 1008, QS L-Phrase_Num-Syls<=5{*/G:0_*,*/G:1_*,*/G:2_*,*/G:3_*,*/G:4_
    *,*/G:5_*}
1010 1009, QS L-Phrase_Num-Syls<=6{*/G:0_*,*/G:1_*,*/G:2_*,*/G:3_*,*/G:4_
    *,*/G:5_*,*/G:6_*}
1011 1010, QS L-Phrase_Num-Syls<=7{*/G:0_*,*/G:1_*,*/G:2_*,*/G:3_*,*/G:4_
    *,*/G:5_*,*/G:6_*,*/G:7_*}
1012 1011, QS L-Phrase_Num-Syls<=8{*/G:0_*,*/G:1_*,*/G:2_*,*/G:3_*,*/G:4_
    *,*/G:5_*,*/G:6_*,*/G:7_*,*/G:8_*}
1013 1012, QS L-Phrase_Num-Syls<=9{*/G:?_*}
1014 1013, QS L-Phrase_Num-Syls<=10{*/G:?_*,*/G:10_*}
1015 1014, QS L-Phrase_Num-Syls<=11{*/G:?_*,*/G:10_*,*/G:11_*}
1016 1015, QS L-Phrase_Num-Syls<=12{*/G:?_*,*/G:10_*,*/G:11_*,*/G:12_*}
1017 1016, QS L-Phrase_Num-Syls<=13{*/G:?_*,*/G:10_*,*/G:11_*,*/G:12_*,*/
    G:13_*}
1018 1017, QS L-Phrase_Num-Syls<=14{*/G:?_*,*/G:10_*,*/G:11_*,*/G:12_*,*/
    G:13_*,*/G:14_*}
1019 1018, QS L-Phrase_Num-Syls<=15{*/G:?_*,*/G:10_*,*/G:11_*,*/G:12_*,*/
    G:13_*,*/G:14_*,*/G:15_*}
1020 1019, QS L-Phrase_Num-Syls<=16{*/G:?_*,*/G:10_*,*/G:11_*,*/G:12_*,*/
    G:13_*,*/G:14_*,*/G:15_*,*/G:16_*}
1021 1020, QS L-Phrase_Num-Syls<=17{*/G:?_*,*/G:10_*,*/G:11_*,*/G:12_*,*/
    G:13_*,*/G:14_*,*/G:15_*,*/G:16_*,*/G:17_*}
1022 1021, QS L-Phrase_Num-Syls<=18{*/G:?_*,*/G:10_*,*/G:11_*,*/G:12_*,*/
    G:13_*,*/G:14_*,*/G:15_*,*/G:16_*,*/G:17_*,*/G:18_*}
1023 1022, QS L-Phrase_Num-Syls<=19{*/G:?_*,*/G:1?_*}

```

```

1024 1023, QS L-Phrase_Num-Syls <=20{* /G: ?_* ,*/G: 1?_* ,*/G: 20_*}
1025 1024, QS L-Phrase_Num-Words ==0{* _0/H:*}
1026 1025, QS L-Phrase_Num-Words ==1{* _1/H:*}
1027 1026, QS L-Phrase_Num-Words ==2{* _2/H:*}
1028 1027, QS L-Phrase_Num-Words ==3{* _3/H:*}
1029 1028, QS L-Phrase_Num-Words ==4{* _4/H:*}
1030 1029, QS L-Phrase_Num-Words ==5{* _5/H:*}
1031 1030, QS L-Phrase_Num-Words ==6{* _6/H:*}
1032 1031, QS L-Phrase_Num-Words ==7{* _7/H:*}
1033 1032, QS L-Phrase_Num-Words ==8{* _8/H:*}
1034 1033, QS L-Phrase_Num-Words ==9{* _9/H:*}
1035 1034, QS L-Phrase_Num-Words ==10{* _10/H:*}
1036 1035, QS L-Phrase_Num-Words ==11{* _11/H:*}
1037 1036, QS L-Phrase_Num-Words ==12{* _12/H:*}
1038 1037, QS L-Phrase_Num-Words ==13{* _13/H:*}
1039 1038, QS L-Phrase_Num-Words <=1{* _0/H:* ,*_1/H:*}
1040 1039, QS L-Phrase_Num-Words <=2{* _0/H:* ,*_1/H:* ,*_2/H:*}
1041 1040, QS L-Phrase_Num-Words <=3{* _0/H:* ,*_1/H:* ,*_2/H:* ,*_3/H:*}
1042 1041, QS L-Phrase_Num-Words <=4{* _0/H:* ,*_1/H:* ,*_2/H:* ,*_3/H:* ,*_4/H
    :*}
1043 1042, QS L-Phrase_Num-Words <=5{* _0/H:* ,*_1/H:* ,*_2/H:* ,*_3/H:* ,*_4/H
    :* ,*_5/H:*}
1044 1043, QS L-Phrase_Num-Words <=6{* _0/H:* ,*_1/H:* ,*_2/H:* ,*_3/H:* ,*_4/H
    :* ,*_5/H:* ,*_6/H:*}
1045 1044, QS L-Phrase_Num-Words <=7{* _0/H:* ,*_1/H:* ,*_2/H:* ,*_3/H:* ,*_4/H
    :* ,*_5/H:* ,*_6/H:* ,*_7/H:*}
1046 1045, QS L-Phrase_Num-Words <=8{* _0/H:* ,*_1/H:* ,*_2/H:* ,*_3/H:* ,*_4/H
    :* ,*_5/H:* ,*_6/H:* ,*_7/H:* ,*_8/H:*}
1047 1046, QS L-Phrase_Num-Words <=9{* _?/H:*}
1048 1047, QS L-Phrase_Num-Words <=10{* _?/H:* ,*_10/H:*}
1049 1048, QS L-Phrase_Num-Words <=11{* _?/H:* ,*_10/H:* ,*_11/H:*}
1050 1049, QS L-Phrase_Num-Words <=12{* _?/H:* ,*_10/H:* ,*_11/H:* ,*_12/H:*}
1051 1050, QS L-Phrase_Num-Words <=13{* _?/H:* ,*_10/H:* ,*_11/H:* ,*_12/H:* ,*_
    _13/H:*}
1052 1051, QS C-Phrase_Num-Syls ==x{* /H: x=*}
1053 1052, QS C-Phrase_Num-Syls ==0{* /H: 0=*}
1054 1053, QS C-Phrase_Num-Syls ==1{* /H: 1=*}
1055 1054, QS C-Phrase_Num-Syls ==2{* /H: 2=*}
1056 1055, QS C-Phrase_Num-Syls ==3{* /H: 3=*}
1057 1056, QS C-Phrase_Num-Syls ==4{* /H: 4=*}
1058 1057, QS C-Phrase_Num-Syls ==5{* /H: 5=*}
1059 1058, QS C-Phrase_Num-Syls ==6{* /H: 6=*}
1060 1059, QS C-Phrase_Num-Syls ==7{* /H: 7=*}
1061 1060, QS C-Phrase_Num-Syls ==8{* /H: 8=*}
1062 1061, QS C-Phrase_Num-Syls ==9{* /H: 9=*}
1063 1062, QS C-Phrase_Num-Syls ==10{* /H: 10=*}
1064 1063, QS C-Phrase_Num-Syls ==11{* /H: 11=*}

```

1065 1064, QS C-Phrase_Num-Syls==12{* /H:12=*}
1066 1065, QS C-Phrase_Num-Syls==13{* /H:13=*}
1067 1066, QS C-Phrase_Num-Syls==14{* /H:14=*}
1068 1067, QS C-Phrase_Num-Syls==15{* /H:15=*}
1069 1068, QS C-Phrase_Num-Syls==16{* /H:16=*}
1070 1069, QS C-Phrase_Num-Syls==17{* /H:17=*}
1071 1070, QS C-Phrase_Num-Syls==18{* /H:18=*}
1072 1071, QS C-Phrase_Num-Syls==19{* /H:19=*}
1073 1072, QS C-Phrase_Num-Syls==20{* /H:20=*}
1074 1073, QS C-Phrase_Num-Syls<=0{* /H:x=*,*/H:0=*}
1075 1074, QS C-Phrase_Num-Syls<=1{* /H:x=*,*/H:0=*,*/H:1=*}
1076 1075, QS C-Phrase_Num-Syls<=2{* /H:x=*,*/H:0=*,*/H:1=*,*/H:2=*}
1077 1076, QS C-Phrase_Num-Syls<=3{* /H:x=*,*/H:0=*,*/H:1=*,*/H:2=*,*/H
:3=*}
1078 1077, QS C-Phrase_Num-Syls<=4{* /H:x=*,*/H:0=*,*/H:1=*,*/H:2=*,*/H
:3=*,*/H:4=*}
1079 1078, QS C-Phrase_Num-Syls<=5{* /H:x=*,*/H:0=*,*/H:1=*,*/H:2=*,*/H
:3=*,*/H:4=*,*/H:5=*}
1080 1079, QS C-Phrase_Num-Syls<=6{* /H:x=*,*/H:0=*,*/H:1=*,*/H:2=*,*/H
:3=*,*/H:4=*,*/H:5=*,*/H:6=*}
1081 1080, QS C-Phrase_Num-Syls<=7{* /H:x=*,*/H:0=*,*/H:1=*,*/H:2=*,*/H
:3=*,*/H:4=*,*/H:5=*,*/H:6=*,*/H:7=*}
1082 1081, QS C-Phrase_Num-Syls<=8{* /H:x=*,*/H:0=*,*/H:1=*,*/H:2=*,*/H
:3=*,*/H:4=*,*/H:5=*,*/H:6=*,*/H:7=*,*/H:8=*}
1083 1082, QS C-Phrase_Num-Syls<=9{* /H:?=*}
1084 1083, QS C-Phrase_Num-Syls<=10{* /H:?=*,*/H:10=*}
1085 1084, QS C-Phrase_Num-Syls<=11{* /H:?=*,*/H:10=*,*/H:11=*}
1086 1085, QS C-Phrase_Num-Syls<=12{* /H:?=*,*/H:10=*,*/H:11=*,*/H:12=*}
1087 1086, QS C-Phrase_Num-Syls<=13{* /H:?=*,*/H:10=*,*/H:11=*,*/H:12=*,*/
H:13=*}
1088 1087, QS C-Phrase_Num-Syls<=14{* /H:?=*,*/H:10=*,*/H:11=*,*/H:12=*,*/
H:13=*,*/H:14=*}
1089 1088, QS C-Phrase_Num-Syls<=15{* /H:?=*,*/H:10=*,*/H:11=*,*/H:12=*,*/
H:13=*,*/H:14=*,*/H:15=*}
1090 1089, QS C-Phrase_Num-Syls<=16{* /H:?=*,*/H:10=*,*/H:11=*,*/H:12=*,*/
H:13=*,*/H:14=*,*/H:15=*,*/H:16=*}
1091 1090, QS C-Phrase_Num-Syls<=17{* /H:?=*,*/H:10=*,*/H:11=*,*/H:12=*,*/
H:13=*,*/H:14=*,*/H:15=*,*/H:16=*,*/H:17=*}
1092 1091, QS C-Phrase_Num-Syls<=18{* /H:?=*,*/H:10=*,*/H:11=*,*/H:12=*,*/
H:13=*,*/H:14=*,*/H:15=*,*/H:16=*,*/H:17=*,*/H:18=*}
1093 1092, QS C-Phrase_Num-Syls<=19{* /H:?=*,*/H:1?=*}
1094 1093, QS C-Phrase_Num-Syls<=20{* /H:?=*,*/H:1?=?,*/H:20=*}
1095 1094, QS C-Phrase_Num-Words==x{* =x@*}
1096 1095, QS C-Phrase_Num-Words==0{* =0@*}
1097 1096, QS C-Phrase_Num-Words==1{* =1@*}
1098 1097, QS C-Phrase_Num-Words==2{* =2@*}
1099 1098, QS C-Phrase_Num-Words==3{* =3@*}

```

1100 1099, QS C-Phrase_Num-Words==4{*=4@*}
1101 1100, QS C-Phrase_Num-Words==5{*=5@*}
1102 1101, QS C-Phrase_Num-Words==6{*=6@*}
1103 1102, QS C-Phrase_Num-Words==7{*=7@*}
1104 1103, QS C-Phrase_Num-Words==8{*=8@*}
1105 1104, QS C-Phrase_Num-Words==9{*=9@*}
1106 1105, QS C-Phrase_Num-Words==10{*=10@*}
1107 1106, QS C-Phrase_Num-Words==11{*=11@*}
1108 1107, QS C-Phrase_Num-Words==12{*=12@*}
1109 1108, QS C-Phrase_Num-Words==13{*=13@*}
1110 1109, QS C-Phrase_Num-Words<=0{*=x@*,*=0@*}
1111 1110, QS C-Phrase_Num-Words<=1{*=x@*,*=0@*,*=1@*}
1112 1111, QS C-Phrase_Num-Words<=2{*=x@*,*=0@*,*=1@*,*=2@*}
1113 1112, QS C-Phrase_Num-Words<=3{*=x@*,*=0@*,*=1@*,*=2@*,*=3@*}
1114 1113, QS C-Phrase_Num-Words<=4{*=x@*,*=0@*,*=1@*,*=2@*,*=3@*,*=4@*}
1115 1114, QS C-Phrase_Num-Words<=5{*=x@*,*=0@*,*=1@*,*=2@*,*=3@*,*=4@
    *,*=5@*}
1116 1115, QS C-Phrase_Num-Words<=6{*=x@*,*=0@*,*=1@*,*=2@*,*=3@*,*=4@
    *,*=5@*,*=6@*}
1117 1116, QS C-Phrase_Num-Words<=7{*=x@*,*=0@*,*=1@*,*=2@*,*=3@*,*=4@
    *,*=5@*,*=6@*,*=7@*}
1118 1117, QS C-Phrase_Num-Words<=8{*=x@*,*=0@*,*=1@*,*=2@*,*=3@*,*=4@
    *,*=5@*,*=6@*,*=7@*,*=8@*}
1119 1118, QS C-Phrase_Num-Words<=9{*=x@*,*=0@*,*=1@*,*=2@*,*=3@*,*=4@
    *,*=5@*,*=6@*,*=7@*,*=8@*,*=9@*}
1120 1119, QS C-Phrase_Num-Words<=10{*=x@*,*=0@*,*=1@*,*=2@*,*=3@*,*=4@
    *,*=5@*,*=6@*,*=7@*,*=8@*,*=9@*,*=10@*}
1121 1120, QS C-Phrase_Num-Words<=11{*=x@*,*=0@*,*=1@*,*=2@*,*=3@*,*=4@
    *,*=5@*,*=6@*,*=7@*,*=8@*,*=9@*,*=10@*,*=11@*}
1122 1121, QS C-Phrase_Num-Words<=12{*=x@*,*=0@*,*=1@*,*=2@*,*=3@*,*=4@
    *,*=5@*,*=6@*,*=7@*,*=8@*,*=9@*,*=10@*,*=11@*,*=12@*}
1123 1122, QS C-Phrase_Num-Words<=13{*=x@*,*=0@*,*=1@*,*=2@*,*=3@*,*=4@
    *,*=5@*,*=6@*,*=7@*,*=8@*,*=9@*,*=10@*,*=11@*,*=12@*,*=13@*}
1124 1123, QS Pos_C-Phrase_in_Utterance(Fw)==1{*@1=*}
1125 1124, QS Pos_C-Phrase_in_Utterance(Fw)==2{*@2=*}
1126 1125, QS Pos_C-Phrase_in_Utterance(Fw)==3{*@3=*}
1127 1126, QS Pos_C-Phrase_in_Utterance(Fw)==4{*@4=*}
1128 1127, QS Pos_C-Phrase_in_Utterance(Fw)<=2{*@1=*,*@2=*}
1129 1128, QS Pos_C-Phrase_in_Utterance(Fw)<=3{*@1=*,*@2=*,*@3=*}
1130 1129, QS Pos_C-Phrase_in_Utterance(Fw)<=4{*@1=*,*@2=*,*@3=*,*@4=*}
1131 1130, QS Pos_C-Phrase_in_Utterance(Bw)==1{*=1|*}
1132 1131, QS Pos_C-Phrase_in_Utterance(Bw)==2{*=2|*}
1133 1132, QS Pos_C-Phrase_in_Utterance(Bw)==3{*=3|*}
1134 1133, QS Pos_C-Phrase_in_Utterance(Bw)==4{*=4|*}
1135 1134, QS Pos_C-Phrase_in_Utterance(Bw)<=2{*=1|*,*=2|*}
1136 1135, QS Pos_C-Phrase_in_Utterance(Bw)<=3{*=1|*,*=2|*,*=3|*}
1137 1136, QS Pos_C-Phrase_in_Utterance(Bw)<=4{*=1|*,*=2|*,*=3|*,*=4|*}

```

```

1138 1137, QS R-Phrase_Num-Syls==0{*/I:0=*}
1139 1138, QS R-Phrase_Num-Syls==1{*/I:1=*}
1140 1139, QS R-Phrase_Num-Syls==2{*/I:2=*}
1141 1140, QS R-Phrase_Num-Syls==3{*/I:3=*}
1142 1141, QS R-Phrase_Num-Syls==4{*/I:4=*}
1143 1142, QS R-Phrase_Num-Syls==5{*/I:5=*}
1144 1143, QS R-Phrase_Num-Syls==6{*/I:6=*}
1145 1144, QS R-Phrase_Num-Syls==7{*/I:7=*}
1146 1145, QS R-Phrase_Num-Syls==8{*/I:8=*}
1147 1146, QS R-Phrase_Num-Syls==9{*/I:9=*}
1148 1147, QS R-Phrase_Num-Syls==10{*/I:10=*}
1149 1148, QS R-Phrase_Num-Syls==11{*/I:11=*}
1150 1149, QS R-Phrase_Num-Syls==12{*/I:12=*}
1151 1150, QS R-Phrase_Num-Syls==13{*/I:13=*}
1152 1151, QS R-Phrase_Num-Syls==14{*/I:14=*}
1153 1152, QS R-Phrase_Num-Syls==15{*/I:15=*}
1154 1153, QS R-Phrase_Num-Syls==16{*/I:16=*}
1155 1154, QS R-Phrase_Num-Syls==17{*/I:17=*}
1156 1155, QS R-Phrase_Num-Syls==18{*/I:18=*}
1157 1156, QS R-Phrase_Num-Syls==19{*/I:19=*}
1158 1157, QS R-Phrase_Num-Syls==20{*/I:20=*}
1159 1158, QS R-Phrase_Num-Syls<=1{*/I:0=*,*/I:1=*}
1160 1159, QS R-Phrase_Num-Syls<=2{*/I:0=*,*/I:1=*,*/I:2=*}
1161 1160, QS R-Phrase_Num-Syls<=3{*/I:0=*,*/I:1=*,*/I:2=*,*/I:3=*}
1162 1161, QS R-Phrase_Num-Syls<=4{*/I:0=*,*/I:1=*,*/I:2=*,*/I:3=*,*/I
:4=*}
1163 1162, QS R-Phrase_Num-Syls<=5{*/I:0=*,*/I:1=*,*/I:2=*,*/I:3=*,*/I
:4=*,*/I:5=*}
1164 1163, QS R-Phrase_Num-Syls<=6{*/I:0=*,*/I:1=*,*/I:2=*,*/I:3=*,*/I
:4=*,*/I:5=*,*/I:6=*}
1165 1164, QS R-Phrase_Num-Syls<=7{*/I:0=*,*/I:1=*,*/I:2=*,*/I:3=*,*/I
:4=*,*/I:5=*,*/I:6=*,*/I:7=*}
1166 1165, QS R-Phrase_Num-Syls<=8{*/I:0=*,*/I:1=*,*/I:2=*,*/I:3=*,*/I
:4=*,*/I:5=*,*/I:6=*,*/I:7=*,*/I:8=*}
1167 1166, QS R-Phrase_Num-Syls<=9{*/I:?=*,*/I:10=*}
1168 1167, QS R-Phrase_Num-Syls<=10{*/I:?=*,*/I:10=*}
1169 1168, QS R-Phrase_Num-Syls<=11{*/I:?=*,*/I:10=*,*/I:11=*}
1170 1169, QS R-Phrase_Num-Syls<=12{*/I:?=*,*/I:10=*,*/I:11=*,*/I:12=*}
1171 1170, QS R-Phrase_Num-Syls<=13{*/I:?=*,*/I:10=*,*/I:11=*,*/I:12=*,*/
I:13=*}
1172 1171, QS R-Phrase_Num-Syls<=14{*/I:?=*,*/I:10=*,*/I:11=*,*/I:12=*,*/
I:13=*,*/I:14=*}
1173 1172, QS R-Phrase_Num-Syls<=15{*/I:?=*,*/I:10=*,*/I:11=*,*/I:12=*,*/
I:13=*,*/I:14=*,*/I:15=*}
1174 1173, QS R-Phrase_Num-Syls<=16{*/I:?=*,*/I:10=*,*/I:11=*,*/I:12=*,*/
I:13=*,*/I:14=*,*/I:15=*,*/I:16=*}

```



```

1175 1174, QS R-Phrase_Num-Syls <=17{* / I : ? = * , * / I : 10 = * , * / I : 11 = * , * / I : 12 = * , * /
      I : 13 = * , * / I : 14 = * , * / I : 15 = * , * / I : 16 = * , * / I : 17 = *}
1176 1175, QS R-Phrase_Num-Syls <=18{* / I : ? = * , * / I : 10 = * , * / I : 11 = * , * / I : 12 = * , * /
      I : 13 = * , * / I : 14 = * , * / I : 15 = * , * / I : 16 = * , * / I : 17 = * , * / I : 18 = *}
1177 1176, QS R-Phrase_Num-Syls <=19{* / I : ? = * , * / I : 1? = *}
1178 1177, QS R-Phrase_Num-Syls <=20{* / I : ? = * , * / I : 1? = * , * / I : 20 = *}
1179 1178, QS R-Phrase_Num-Words ==0{* =0 / J : *}
1180 1179, QS R-Phrase_Num-Words ==1{* =1 / J : *}
1181 1180, QS R-Phrase_Num-Words ==2{* =2 / J : *}
1182 1181, QS R-Phrase_Num-Words ==3{* =3 / J : *}
1183 1182, QS R-Phrase_Num-Words ==4{* =4 / J : *}
1184 1183, QS R-Phrase_Num-Words ==5{* =5 / J : *}
1185 1184, QS R-Phrase_Num-Words ==6{* =6 / J : *}
1186 1185, QS R-Phrase_Num-Words ==7{* =7 / J : *}
1187 1186, QS R-Phrase_Num-Words ==8{* =8 / J : *}
1188 1187, QS R-Phrase_Num-Words ==9{* =9 / J : *}
1189 1188, QS R-Phrase_Num-Words ==10{* =10 / J : *}
1190 1189, QS R-Phrase_Num-Words ==11{* =11 / J : *}
1191 1190, QS R-Phrase_Num-Words ==12{* =12 / J : *}
1192 1191, QS R-Phrase_Num-Words ==13{* =13 / J : *}
1193 1192, QS R-Phrase_Num-Words ==14{* =14 / J : *}
1194 1193, QS R-Phrase_Num-Words ==15{* =15 / J : *}
1195 1194, QS R-Phrase_Num-Words <=1{* =0 / J : * , * =1 / J : *}
1196 1195, QS R-Phrase_Num-Words <=2{* =0 / J : * , * =1 / J : * , * =2 / J : *}
1197 1196, QS R-Phrase_Num-Words <=3{* =0 / J : * , * =1 / J : * , * =2 / J : * , * =3 / J : *}
1198 1197, QS R-Phrase_Num-Words <=4{* =0 / J : * , * =1 / J : * , * =2 / J : * , * =3 / J : * , * =4 / J
      : *}
1199 1198, QS R-Phrase_Num-Words <=5{* =0 / J : * , * =1 / J : * , * =2 / J : * , * =3 / J : * , * =4 / J
      : * , * =5 / J : *}
1200 1199, QS R-Phrase_Num-Words <=6{* =0 / J : * , * =1 / J : * , * =2 / J : * , * =3 / J : * , * =4 / J
      : * , * =5 / J : * , * =6 / J : *}
1201 1200, QS R-Phrase_Num-Words <=7{* =0 / J : * , * =1 / J : * , * =2 / J : * , * =3 / J : * , * =4 / J
      : * , * =5 / J : * , * =6 / J : * , * =7 / J : *}
1202 1201, QS R-Phrase_Num-Words <=8{* =0 / J : * , * =1 / J : * , * =2 / J : * , * =3 / J : * , * =4 / J
      : * , * =5 / J : * , * =6 / J : * , * =7 / J : * , * =8 / J : *}
1203 1202, QS R-Phrase_Num-Words <=9{* =? / J : *}
1204 1203, QS R-Phrase_Num-Words <=10{* =? / J : * , * =10 / J : *}
1205 1204, QS R-Phrase_Num-Words <=11{* =? / J : * , * =10 / J : * , * =11 / J : *}
1206 1205, QS R-Phrase_Num-Words <=12{* =? / J : * , * =10 / J : * , * =11 / J : * , * =12 / J : *}
1207 1206, QS R-Phrase_Num-Words <=13{* =? / J : * , * =10 / J : * , * =11 / J : * , * =12 / J
      : * , * =13 / J : *}
1208 1207, QS R-Phrase_Num-Words <=14{* =? / J : * , * =10 / J : * , * =11 / J : * , * =12 / J
      : * , * =13 / J : * , * =14 / J : *}
1209 1208, QS R-Phrase_Num-Words <=15{* =? / J : * , * =10 / J : * , * =11 / J : * , * =12 / J
      : * , * =13 / J : * , * =14 / J : * , * =15 / J : *}
1210 1209, QS Num-Syls_in_Utterance ==1{* / J : 1 + *}
1211 1210, QS Num-Syls_in_Utterance ==2{* / J : 2 + *}

```


1212 1211, QS Num-Syls_in_Utterance==3{*/J:3+*}
1213 1212, QS Num-Syls_in_Utterance==4{*/J:4+*}
1214 1213, QS Num-Syls_in_Utterance==5{*/J:5+*}
1215 1214, QS Num-Syls_in_Utterance==6{*/J:6+*}
1216 1215, QS Num-Syls_in_Utterance==7{*/J:7+*}
1217 1216, QS Num-Syls_in_Utterance==8{*/J:8+*}
1218 1217, QS Num-Syls_in_Utterance==9{*/J:9+*}
1219 1218, QS Num-Syls_in_Utterance==10{*/J:10+*}
1220 1219, QS Num-Syls_in_Utterance==11{*/J:11+*}
1221 1220, QS Num-Syls_in_Utterance==12{*/J:12+*}
1222 1221, QS Num-Syls_in_Utterance==13{*/J:13+*}
1223 1222, QS Num-Syls_in_Utterance==14{*/J:14+*}
1224 1223, QS Num-Syls_in_Utterance==15{*/J:15+*}
1225 1224, QS Num-Syls_in_Utterance==16{*/J:16+*}
1226 1225, QS Num-Syls_in_Utterance==17{*/J:17+*}
1227 1226, QS Num-Syls_in_Utterance==18{*/J:18+*}
1228 1227, QS Num-Syls_in_Utterance==19{*/J:19+*}
1229 1228, QS Num-Syls_in_Utterance==20{*/J:20+*}
1230 1229, QS Num-Syls_in_Utterance==21{*/J:21+*}
1231 1230, QS Num-Syls_in_Utterance==22{*/J:22+*}
1232 1231, QS Num-Syls_in_Utterance==23{*/J:23+*}
1233 1232, QS Num-Syls_in_Utterance==24{*/J:24+*}
1234 1233, QS Num-Syls_in_Utterance==25{*/J:25+*}
1235 1234, QS Num-Syls_in_Utterance==26{*/J:26+*}
1236 1235, QS Num-Syls_in_Utterance==27{*/J:27+*}
1237 1236, QS Num-Syls_in_Utterance==28{*/J:28+*}
1238 1237, QS Num-Syls_in_Utterance<=2{*/J:1+*,*/J:2+*}
1239 1238, QS Num-Syls_in_Utterance<=3{*/J:1+*,*/J:2+*,*/J:3+*}
1240 1239, QS Num-Syls_in_Utterance<=4{*/J:1+*,*/J:2+*,*/J:3+*,*/J:4+*}
1241 1240, QS Num-Syls_in_Utterance<=5{*/J:1+*,*/J:2+*,*/J:3+*,*/J:4+*,*/
J:5+*}
1242 1241, QS Num-Syls_in_Utterance<=6{*/J:1+*,*/J:2+*,*/J:3+*,*/J:4+*,*/
J:5+*,*/J:6+*}
1243 1242, QS Num-Syls_in_Utterance<=7{*/J:1+*,*/J:2+*,*/J:3+*,*/J:4+*,*/
J:5+*,*/J:6+*,*/J:7+*}
1244 1243, QS Num-Syls_in_Utterance<=8{*/J:1+*,*/J:2+*,*/J:3+*,*/J:4+*,*/
J:5+*,*/J:6+*,*/J:7+*,*/J:8+*}
1245 1244, QS Num-Syls_in_Utterance<=9{*/J:?+*}
1246 1245, QS Num-Syls_in_Utterance<=10{*/J:?+*,*/J:10+*}
1247 1246, QS Num-Syls_in_Utterance<=11{*/J:?+*,*/J:10+*,*/J:11+*}
1248 1247, QS Num-Syls_in_Utterance<=12{*/J:?+*,*/J:10+*,*/J:11+*,*/J
:12+*}
1249 1248, QS Num-Syls_in_Utterance<=13{*/J:?+*,*/J:10+*,*/J:11+*,*/J
:12+*,*/J:13+*}
1250 1249, QS Num-Syls_in_Utterance<=14{*/J:?+*,*/J:10+*,*/J:11+*,*/J
:12+*,*/J:13+*,*/J:14+*}

```

1251 1250, QS Num-Syls_in_Utterance <=15{*/J:?+*,*/J:10+*,*/J:11+*,*/J
      :12+*,*/J:13+*,*/J:14+*,*/J:15+*}
1252 1251, QS Num-Syls_in_Utterance <=16{*/J:?+*,*/J:10+*,*/J:11+*,*/J
      :12+*,*/J:13+*,*/J:14+*,*/J:15+*,*/J:16+*}
1253 1252, QS Num-Syls_in_Utterance <=17{*/J:?+*,*/J:10+*,*/J:11+*,*/J
      :12+*,*/J:13+*,*/J:14+*,*/J:15+*,*/J:16+*,*/J:17+*}
1254 1253, QS Num-Syls_in_Utterance <=18{*/J:?+*,*/J:10+*,*/J:11+*,*/J
      :12+*,*/J:13+*,*/J:14+*,*/J:15+*,*/J:16+*,*/J:17+*,*/J:18+*}
1255 1254, QS Num-Syls_in_Utterance <=19{*/J:?+*,*/J:1?+*}
1256 1255, QS Num-Syls_in_Utterance <=20{*/J:?+*,*/J:1?+*,*/J:20+*}
1257 1256, QS Num-Syls_in_Utterance <=21{*/J:?+*,*/J:1?+*,*/J:20+*,*/J
      :21+*}
1258 1257, QS Num-Syls_in_Utterance <=22{*/J:?+*,*/J:1?+*,*/J:20+*,*/J
      :21+*,*/J:22+*}
1259 1258, QS Num-Syls_in_Utterance <=23{*/J:?+*,*/J:1?+*,*/J:20+*,*/J
      :21+*,*/J:22+*,*/J:23+*}
1260 1259, QS Num-Syls_in_Utterance <=24{*/J:?+*,*/J:1?+*,*/J:20+*,*/J
      :21+*,*/J:22+*,*/J:23+*,*/J:24+*}
1261 1260, QS Num-Syls_in_Utterance <=25{*/J:?+*,*/J:1?+*,*/J:20+*,*/J
      :21+*,*/J:22+*,*/J:23+*,*/J:24+*,*/J:25+*}
1262 1261, QS Num-Syls_in_Utterance <=26{*/J:?+*,*/J:1?+*,*/J:20+*,*/J
      :21+*,*/J:22+*,*/J:23+*,*/J:24+*,*/J:25+*,*/J:26+*}
1263 1262, QS Num-Syls_in_Utterance <=27{*/J:?+*,*/J:1?+*,*/J:20+*,*/J
      :21+*,*/J:22+*,*/J:23+*,*/J:24+*,*/J:25+*,*/J:26+*,*/J:27+*}
1264 1263, QS Num-Syls_in_Utterance <=28{*/J:?+*,*/J:1?+*,*/J:20+*,*/J
      :21+*,*/J:22+*,*/J:23+*,*/J:24+*,*/J:25+*,*/J:26+*,*/J:27+*,*/J
      :28+*}
1265 1264, QS Num-Words_in_Utterance ==1{**+1-*}
1266 1265, QS Num-Words_in_Utterance ==2{**+2-*}
1267 1266, QS Num-Words_in_Utterance ==3{**+3-*}
1268 1267, QS Num-Words_in_Utterance ==4{**+4-*}
1269 1268, QS Num-Words_in_Utterance ==5{**+5-*}
1270 1269, QS Num-Words_in_Utterance ==6{**+6-*}
1271 1270, QS Num-Words_in_Utterance ==7{**+7-*}
1272 1271, QS Num-Words_in_Utterance ==8{**+8-*}
1273 1272, QS Num-Words_in_Utterance ==9{**+9-*}
1274 1273, QS Num-Words_in_Utterance ==10{**+10-*}
1275 1274, QS Num-Words_in_Utterance ==11{**+11-*}
1276 1275, QS Num-Words_in_Utterance ==12{**+12-*}
1277 1276, QS Num-Words_in_Utterance ==13{**+13-*}
1278 1277, QS Num-Words_in_Utterance <=2{**+1-*,**+2-*}
1279 1278, QS Num-Words_in_Utterance <=3{**+1-*,**+2-*,**+3-*}
1280 1279, QS Num-Words_in_Utterance <=4{**+1-*,**+2-*,**+3-*,**+4-*}
1281 1280, QS Num-Words_in_Utterance <=5{**+1-*,**+2-*,**+3-*,**+4-*,**+5-*}
1282 1281, QS Num-Words_in_Utterance
      <=6{**+1-*,**+2-*,**+3-*,**+4-*,**+5-*,**+6-*}

```

```

1283 1282, QS Num-Words_in_Utterance
      <=7{**1-*,**2-*,**3-*,**4-*,**5-*,**6-*,**7-*}
1284 1283, QS Num-Words_in_Utterance
      <=8{**1-*,**2-*,**3-*,**4-*,**5-*,**6-*,**7-*,**8-*}
1285 1284, QS Num-Words_in_Utterance <=9{**?-*}
1286 1285, QS Num-Words_in_Utterance <=10{**?-*,**10-*}
1287 1286, QS Num-Words_in_Utterance <=11{**?-*,**10-*,**11-*}
1288 1287, QS Num-Words_in_Utterance <=12{**?-*,**10-*,**11-*,**12-*}
1289 1288, QS Num-Words_in_Utterance
      <=13{**?-*,**10-*,**11-*,**12-*,**13-*}
1290 1289, QS Num-Phrases_in_Utterance==1{* -1/K:*}
1291 1290, QS Num-Phrases_in_Utterance==2{* -2/K:*}
1292 1291, QS Num-Phrases_in_Utterance==3{* -3/K:*}
1293 1292, QS Num-Phrases_in_Utterance==4{* -4/K:*}
1294 1293, QS Num-Phrases_in_Utterance <=2{* -1/K:*,* -2/K:*}
1295 1294, QS Num-Phrases_in_Utterance <=3{* -1/K:*,* -2/K:*,* -3/K:*}
1296 1295, QS Num-Phrases_in_Utterance <=4{* -1/K:*,* -2/K:*,* -3/K:*,* -4/K
      :*}
1297 1296, vuv

```

Listing A.1: A comma-separated listing of feature indexes and their codenames including the HTK HED matching patterns used for extraction from the full-context labels.

Listing of Feature Groups

This section contains the definition of the nested feature groups in form of a raw Python script in Listing B.1. These groups were used in this work to provide a number of different perspectives on the relevance of the input features on the F_0 contour predicted by the intonation model. The listing has a bottom-up structure with the most concrete groups defined as simple lists of individual features at the beginning and the more abstract feature groups towards the end where the actual groups used in the relevance analysis are defined. The individual features are listed and explained in Appendix A of this work.

The first group `VUV` defined in lines 1–3 contains a single feature that represents the voicing information.

After that, in lines 5–24 groups that hold features that represent whether a segment is a vowel or a consonant are defined as `QUINTPHONE_SEGMENT_VC_*`. The `*` is a placeholder for one of the five quintphone contexts; the leftmost `LL`, immediate left `L`, central (current) `C`, immediate right `R` and rightmost `RR` segments.

In a similar manner, features related to the phonetic and phonological identity of a segment (`QUINTPHONE_SEGMENT_ARTICULATORY_TYPE_**`) are grouped by the quintphone context in lines 26–430.

Feature groups defined as `NUM_SEG_IN_SYL_*DIRECTION*_*RELATION*` in lines 432–470 contain features that encode the number of other segments in the current syllable. The `*DIRECTION*` part of the group name denotes whether the previous `BW` of the next `FW` segments are counted and `*RELATION*` denotes whether the number is equal `EQ` or less-than-or-equal `LTE` to the given number.

Feature groups in lines 472–508, group features of the immediate left neighbour of the current syllable. These include features related to stress, accent, accent type (ToBI tone) and length of the syllable expressed in number of segments respectively. The following groups in lines 510–548 contain the same features but related to the current syllable.

Additionally, in lines 550–677 groups related to the forward and backward position of the syllable in the current word and phrase are defined. Lines 679–884 contain

definitions of feature groups related to the number of stressed and accented syllables before and after the current syllable in the current phrase, as well as, the forward and backward distance to the closest stressed and accented syllables, expressed in number of syllables, respectively.

An additional group holding all features related to the phonetic and phonological identity of the syllable nucleus is defined in lines 886-904.

Lines 906–942 contain feature groups related to the immediate right neighbour of the current syllable defined in the same manner as groups containing features for the immediate left neighbour defined in lines 472–508.

Groups holding features related to the length of the immediate left and current word expressed in the number of syllables are defined in lines 944–962 and 963–981.

Groups containing features related to the forward and backward position of the current word in the current phrase are defined in lines 983–1045.

Features that express the number of content (accented) words before and after the current word in the current phrase and the distance to the nearest previous and next content word are structured into groups in lines 1047–1126.

Groups defined in lines 1128–1146 contain features related to the length of the immediate left word expressed in number of syllables.

Feature groups in lines 1148–1224 include the lengths of the immediate left phrase expressed in number of syllables and words.

Similarly, the groups in lines 1226–1306 contain features related to the length of the current phrase, also expressed in number of syllables and words. Additionally, for the current phrase, in lines 1308–1330, feature groups related to the forward and backward position of the current phrase in the current utterance are defined.

The feature groups with syllable- and word-lengths of the immediate right phrase are listed in lines 1332–1412.

The inventory of low level groups ends with groupings of features related to the length of the whole utterance, expressed in number of syllables, words and phrases.

The following groups provide an even higher abstraction level as they are defined as lists of lower level groups.

The `QUINTPHONE_SEGMENT_VC` group in lines 1524–1530 contains all low level groups of features encoding whether specific segments of the quintphone (i.e. the leftmost, immediate left, central, immediate right and rightmost segments) are a vowel or a consonant. Similarly, `QUINTPHONE_SEGMENT_ARTICULATORY_TYPE` defined in 1531–1537 lists all five feature groups containing features related to the phonetic and phonological identities and properties of the quintphone segments.

`SEG_POS_IN_SYL`, in turn (lines 1538–1543) groups all lower-level feature groups related to the position of a segment in current syllable, forward and backward, and expressed through both the equality and inequality checks.

`QUINTPHONE_SEGMENTAL_FEATURES` in lines 1544–1547 was defined as a list of the high level groups related to the vowel/consonant distinction and the phonetic and phonological properties and identities of a segment (defined just a few lines above).

The `QUINTPHONE` group, in turn, contains all of the higher level groups related to the quintphone segments.

All of the syllable-related low-level groups were structured in a similar fashion. The groups containing features related to the information about the previous, next and current syllable length, nucleus type, stress and accent presence and type are defined in lines 1560–1586.

Groups containing positional features for syllables were first grouped keeping the distinction of the parent unit type in which the position is calculated, i.e. word and phrase, and the direction towards which the position is determined, i.e. forward and backward (lines 1588–1603). Additional groups related to the enclosing parent unit only, are then created in lines 1604–1611 to include the various positional groups which are all finally grouped under `SYLLABLE_POSITION` in lines 1613–1616.

Groups in lines 1618–1641 include all lower-level groups related to the number of stressed and accented syllables before and after the current syllable. These groups are then included in a single general group `SYLLABLE_NEIGHBOURHOOD` in lines 1643–1650.

In the same manner, groups of features connected with the position of the current syllable in relation to the closest accented and stressed syllables in both directions are gradually structured in lines 1652–1675, to be finally included in a single high-level group `RELATIVE_SYLLABLE_POSITION` in lines 1677–1684.

The groups of features related to the length of the previous, current and next word measured in number of syllables are arranged into `WORD_LEN_IN_SYLS` in lines 1690–1697. Feature groups related to the position of the current words are first arranged into a number of more granular groups in lines 1698–1709, and then included in a single group called `WORD_POSITION` (lines 1710–1714), similar to the scheme used in case of the syllable.

In a very similar fashion the feature groups related to the number of content (accented) word before and after the current word in the current phrase are gradually structured into `WORD_SURROUNDINGS` in lines 1715–1726, whereas the feature groups related to the position of the current word in the current phrase in relation to closest next and previous content word are structured into `RELATIVE_WORD_POSITION` in lines 1727–1738.

All low-level phrase length groups are listed in `PHRASE_LEN_IN_SYLS` (lines 1744–1751), `PHRASE_LEN_IN_WORDS` (lines 1752–1759) and eventually in `PHRASE_LEN` in lines 1760–1763. Phrase position groups are structured in a similar fashion in the following lines (1764–1775).

`UTTERANCE_LEN` is constructed in lines 1781–1797.

After that the previously defined high-level groups are further structured into groups that provide an even higher and more intuitive level of abstraction. All groups related to the position of a segment are listed under `SEGMENTAL_POSITIONAL_ABSOLUTE` in lines 1801–1806. On the other hand, all groups related the phonetic and phonological qualities of a segment are structured under `SEGMENTAL_QUALITATIVE`.

Similarly, `SYLLABIC_POSITIONAL_ABSOLUTE` defined in lines 1823–1832 contains all groups related to the absolute position of a syllable in the enclosing unit, and `SYLLABIC_POSITIONAL_RELATIVE` (1834–1843) contain groups related to the relative position of a syllable.

Qualitative syllable-related feature groups define the `SYLLABIC_QUALITATIVE` group in lines 1845–1856.

In 1858–1865 all feature groups related to the composition (so length expressed in number of subordinate units) are listed under `SYLLABIC_COMPOSITIONAL`.

`SYLLABIC_PARENTAL_COMPOSITION` (lines 1867–1876), in turn, contain all syllabic feature groups related to the number of stressed and accented syllables in the current phrase.

`WORD_POSITIONAL_ABSOLUTE` and `WORD_POSITIONAL_RELATIVE` in lines 1880–1892 are defined with all objective and subjective word position feature groups.

`WORD_COMPOSITIONAL` (lines 1894–1901) and `WORD_PARENTAL_COMPOSITION` (1903–1908), similarly to their syllabic counterparts, contain groups that list features related to the length of word expressed in number of syllables, and the number of content words in the current phrase.

Feature groups related to the absolute position of phrase and its length expressed in subordinate units are grouped under `PHRASAL_POSITIONAL_ABSOLUTE` (lines 1912–1917) and `UTTERANCE_COMPOSITIONAL` (lines 1919–1932).

As the name suggests, `UTTERANCE_COMPOSITIONAL` (lines 1936–1943) contains all feature groups related to the composition (length expressed in all classes of subordinate units) of the utterance.

Next, all positional feature groups, regardless of the unit type are listed under `POSITIONAL_ABSOLUTE` and `POSITIONAL_RELATIVE` in lines 1945–1955.

Feature groups containing qualitative features of segments and syllables are grouped under `QUALITATIVE` in lines 1957–1960. Similarly, feature groups related to the composition (length expressed in number of subordinate units) and to the composition of the parental unit (number of stressed and accented syllables and content words in the current phrase) are listed together under `COMPOSITIONAL` and `PARENTAL_COMPOSITION` (lines 1962–1972) respectively.

After that all feature groups related to a single organisational unit type are listed together under `SEGMENTAL`, `SYLLABIC`, `WORD`, `PHRASAL` and `UTTERANCE` (lines 1974–2001).

The three feature groups that follow were defined to represent the three main abstraction levels used in the relevance analysis in the current work. First group defined in lines 2006–2027 – `FEATURE_GROUPS` – represents the highest level of abstraction as it consists of the most general high-level feature groups only. The next group, `DETAILED_GROUPS` (lines 2029–2074) provides a more granular list. The distinction between forward and backward positional and forward backward compositional features is kept here, for example. Also the qualitative features are not grouped under a single name but kept separate. The last abstraction level, defined in `ALL_GROUPS` (lines 2076–2172) is simply a list of all of the lower level feature groups. The quintphone segments are represented by separate items. Also the features that are defined separately for the previous, current and next unit are kept separate here. Additionally, separate list items are included for the different types of

relations (equal and less-than-or-equal) used for positional and other quantitative features.

At the end of the listing, some additional target groups were defined to provide a yet another look at the data. The separate groups defined for different organizational units (segment, syllable, word and phrase) are structured under a single group `LINGUISTIC_LEVEL_GROUPS` in line 2228–2234. Also groups listing all positional, qualitative, compositional, etc. features are defined separately in lines 2199–2226 and later included together under `FEATURE_TYPE_GROUPS` in lines 2236–2242. The last group provides a list combined of groups related to different organizational units and feature type (these are feature groups defined earlier in lines 1945–1972) at the same time.

```

1  VUV = [
2      "vuv",
3  ]
4
5  QUINTPHONE_SEGMENT_VC_LL = [
6      "QS LL-Vowel{i^*,y^*,e^*,a^*,o^*,u^*,schwa^*}",
7      "QS LL-Consonant{gs^*,p^*,b^*,t^*,d^*,k^*,g^*,ki^*,gi^*,f^*,v
      ^*,s^*,si^*,z^*,zi^*,sz^*,rz^*,x^*,c^*,dz^*,cz^*,drz^*,ci^*,
      dzi^*,m^*,n^*,ni^*,ng^*,l^*,r^*,w^*,ww^*,j^*,jj^*}",
8  ]
9  QUINTPHONE_SEGMENT_VC_L = [
10     "QS L-Vowel{*^schwa-,*^i-,*^y-,*^e-,*^a-,*^o-,*^u-}",
11     "QS L-Consonant
      {^gs-,^p-,^b-,^t-,^d
      -,^k-,^g-,^ki-,^gi-,^f-,^v-,^s-,^si-,^z-,^zi-,^sz
      -,^rz-,^x-,^c-,^dz-,^cz-,^drz-,^ci-,^dzi-,^m-,^n
      -,^ni-,^ng-,^l-,^r-,^w-,^ww-,^j-,^jj-}",
12 ]
13 QUINTPHONE_SEGMENT_VC_C = [
14     "QS C-Vowel{*schwa+,*-i+,*-y+,*-e+,*-a+,*-o+,*-u+}",
15     "QS C-Consonant
      {-gs+,-p+,-b+,-t
      +,-d+,-k+,-g+,-ki+,-gi+,-f+,-v+,-s+,-si
      +,-z+,-zi+,-sz+,-rz+,-x+,-c+,-dz+,-cz+,-
      drz+,-ci+,-dzi+,-m+,-n+,-ni+,-ng+,-l+,-r
      +,-w+,-ww+,-j+,-jj+}",
16 ]
17 QUINTPHONE_SEGMENT_VC_R = [
18     "QS R-Vowel{*i=,*y=,*e=,*a=,*o=,*u=,*schwa=*}",
19     "QS R-Consonant
      {+gs=,+p=,+b=,+t
      =,+d=,+k=,+g=,+ki=,+gi=,+f=,+v=,+s=,+si
      =,+z=,+zi=,+sz=,+rz=,+x=,+c=,+dz=,+cz=,+
      drz=,+ci=,+dzi=,+m=,+n=,+ni=,+ng=,+l=,+r
      =,+w=,+ww=,+j=,+jj=*}",
20 ]

```

```

21  QUINTPHONE_SEGMENT_VC_RR = [
22      "QS RR-Vowel{*=i@*,*=y@*,*=e@*,*=a@*,*=o@*,*=u@*,*=schwa@*}",
23      "QS RR-Consonant                                     {*=gs@*,*=p@*,*=b@*,*=t@
        *,*=d@*,*=k@*,*=g@*,*=ki@*,*=gi@*,*=f@*,*=v@*,*=s@*,*=si@
        *,*=z@*,*=zi@*,*=sz@*,*=rz@*,*=x@*,*=c@*,*=dz@*,*=cz@*,*=
        drz@*,*=ci@*,*=dzi@*,*=m@*,*=n@*,*=ni@*,*=ng@*,*=l@*,*=r@
        *,*=w@*,*=ww@*,*=j@*,*=jj@*}"
24  ]
25
26  QUINTPHONE_SEGMENT_ARTICULATORY_TYPE_LL = [
27      "QS LL-Stop{gs^*,p^*,b^*,t^*,d^*,k^*,g^*}",
28      "QS LL-Nasal{ww^*,jj^*,m^*,n^*,ni^*,ng^*}",
29      "QS LL-Fricative{f^*,v^*,s^*,si^*,z^*,zi^*,sz^*,rz^*,x^*}",
30      "QS LL-Front{e^*,i^*,y^*,f^*,v^*,p^*,b^*,m^*,w^*,ww^*}",
31      "QS LL-Central{schwa^*,a^*,t^*,d^*,s^*,si^*,z^*,zi^*,n^*,r^*,l
        ^*,t^*,d^*,sz^*,rz^*,cz^*,drz^*,c^*,dz^*,ci^*,dzi^*}",
32      "QS LL-Back{o^*,u^*,k^*,g^*,ki^*,gi^*,ng^*,x^*,gs^*}",
33      "QS LL-Front_Vowel{e^*,i^*,y^*}",
34      "QS LL-Central_Vowel{a^*,schwa^*}",
35      "QS LL-Back_Vowel{o^*,u^*}",
36      "QS LL-High_Vowel{i^*,y^*,u^*}",
37      "QS LL-Medium_Vowel{e^*,o^*}",
38      "QS LL-Low_Vowel{a^*}",
39      "QS LL-Rounded_Vowel{o^*,u^*}",
40      "QS LL-Unrounded_Vowel{a^*,e^*,i^*,y^*}",
41      "QS LL-IVowel{i^*}",
42      "QS LL-EVowel{e^*}",
43      "QS LL-AVowel{a^*}",
44      "QS LL-OVowel{o^*}",
45      "QS LL-UVowel{u^*}",
46      "QS LL-YVowel                                     {y^*}",
47      "QS LL-SCHWAVowel                                   {schwa^*}",
48      "QS LL-Unvoiced_Consonant{gs^*,p^*,t^*,k^*,ki^*,f^*,v^*,s^*,sz
        ^*,x^*,c^*,cz^*,ci^*}",
49      "QS LL-Voiced_Consonant{b^*,d^*,g^*,gi^*,v^*,z^*,zi^*,rz^*,dz
        ^*,drz^*,dzi^*,m^*,n^*,ni^*,ng^*,l^*,r^*,w^*,ww^*,j^*,jj
        ^*}",
50      "QS LL-Front_Consonant{f^*,v^*,f^*,p^*,b^*,m^*,w^*,ww^*}",
51      "QS LL-Central_Consonant{t^*,d^*,s^*,si^*,z^*,zi^*,n^*,r^*,l^*,
        t^*,d^*,sz^*,rz^*,cz^*,drz^*,c^*,dz^*,ci^*,dzi^*}",
52      "QS LL-Back_Consonant{gs^*,k^*,g^*,ki^*,gi^*,ng^*,x^*}",
53      "QS LL-Fortis_Consonant{gs^*,cz^*,f^*,k^*,p^*,s^*,sz^*,t^*,ci
        ^*,c^*,ki^*}",
54      "QS LL-Lenis_Consonant{drz^*,v^*,g^*,b^*,rz^*,z^*,d^*,dzi^*,dz
        ^*,gi^*,zi^*}",
55      "QS LL-Neigther_F_or_L{m^*,n^*,ni^*,ng^*,l^*,r^*,w^*,ww^*,j^*,
        jj^*}"

```

```

56 "QS LL-Voiced_Stop{b^*,d^*,g^*}",
57 "QS LL-Unvoiced_Stop{p^*,t^*,k^*,gs^*}",
58 "QS LL-Front_Stop{b^*,p^*}",
59 "QS LL-Central_Stop{d^*,t^*}",
60 "QS LL-Back_Stop{g^*,k^*,gs^*}",
61 "QS LL-Voiced_Fricative{v^*,z^*,zi^*,rz^*}",
62 "QS LL-Unvoiced_Fricative{f^*,s^*,si^*,sz^*,x^*}",
63 "QS LL-Front_Fricative{f^*,v^*}",
64 "QS LL-Affricate_Consonant{dz^*,drz^*,dzi^*,c^*,cz^*,ci^*}",
65 "QS LL-silences{pau^*}",
66 "QS LL-schwa {schwa^*}",
67 "QS LL-a{a^*}",
68 "QS LL-e{e^*}",
69 "QS LL-i{i^*}",
70 "QS LL-y{y^*}",
71 "QS LL-o{o^*}",
72 "QS LL-u{u^*}",
73 "QS LL-p{p^*}",
74 "QS LL-b{b^*}",
75 "QS LL-t{t^*}",
76 "QS LL-d{d^*}",
77 "QS LL-k{k^*}",
78 "QS LL-ki{ki^*}",
79 "QS LL-g{g^*}",
80 "QS LL-gi{gi^*}",
81 "QS LL-f{f^*}",
82 "QS LL-v{v^*}",
83 "QS LL-s{s^*}",
84 "QS LL-si{si^*}",
85 "QS LL-z{z^*}",
86 "QS LL-zi{zi^*}",
87 "QS LL-sz{sz^*}",
88 "QS LL-rz{rz^*}",
89 "QS LL-x{x^*}",
90 "QS LL-c{c^*}",
91 "QS LL-dz{dz^*}",
92 "QS LL-cz{cz^*}",
93 "QS LL-drz{drz^*}",
94 "QS LL-ci{ci^*}",
95 "QS LL-dzi{dzi^*}",
96 "QS LL-m{m^*}",
97 "QS LL-n{n^*}",
98 "QS LL-ni{ni^*}",
99 "QS LL-ng{ng^*}",
100 "QS LL-l{l^*}",
101 "QS LL-r{r^*}",
102 "QS LL-w{w^*}",

```

```

103     "QS LL-ww{ww^*}",
104     "QS LL-j{j^*}",
105     "QS LL-jj{jj^*}",
106     "QS LL-gs                                {gs^*}",
107 ]
108 QUINTPHONE_SEGMENT_ARTICULATORY_TYPE_L = [
109     "QS L-Stop{*^p-,*^b-,*^t-,*^d-,*^k-,*^g-,*^gs-}",
110     "QS L-Nasal{*^ww-,*^jj-,*^m-,*^n-,*^ni-,*^ng-}",
111     "QS L-Fricative{*^f-,*^v-,*^s-,*^si-,*^z-,*^zi-,*^sz-,*^rz-,*^x-
    -}",
112     "QS L-Front{*^e-,*^i-,*^y-,*^f-,*^v-,*^p-,*^b-,*^m-,*^w-,*^ww
    -}",
113     "QS L-Central{*^schwa-,*^a-,*^t-,*^d-,*^s-,*^si-,*^z-,*^zi-,*^n
    -,*^r-,*^l-,*^t-,*^d-,*^sz-,*^rz-,*^cz-,*^drz-,*^c-,*^dz-,*^
    ci-,*^dzi-}",
114     "QS L-Back{*^o-,*^u-,*^k-,*^g-,*^ki-,*^gi-,*^ng-,*^x-,*^gs-}",
115     "QS L-Front_Vowel{*^e-,*^i-,*^y-}",
116     "QS L-Central_Vowel{*^a-,*^schwa-}",
117     "QS L-Back_Vowel{*^o-,*^u-}",
118     "QS L-High_Vowel{*^i-,*^y-,*^u-}",
119     "QS L-Medium_Vowel{*^e-,*^o-}",
120     "QS L-Low_Vowel{*^a-}",
121     "QS L-Rounded_Vowel{*^o-,*^u-}",
122     "QS L-Unrounded_Vowel{*^a-,*^e-,*^i-,*^y-}",
123     "QS L-IVowel{*^i-}",
124     "QS L-EVowel{*^e-}",
125     "QS L-AVowel{*^a-}",
126     "QS L-OVowel{*^o-}",
127     "QS L-UVowel{*^u-}",
128     "QS L-YVowel                                {*^y-}",
129     "QS LL-SCHWAVowel                            {*^schwa-}",
130     "QS L-Unvoiced_Consonant{*^gs-,*^p-,*^t-,*^k-,*^ki-,*^f-,*^v
    -,*^s-,*^sz-,*^x-,*^c-,*^cz-,*^ci-}",
131     "QS L-Voiced_Consonant{*^b-,*^d-,*^g-,*^gi-,*^v-,*^z-,*^zi-,*^
    rz-,*^dz-,*^drz-,*^dzi-,*^m-,*^n-,*^ni-,*^ng-,*^l-,*^r-,*^w
    -,*^ww-,*^j-,*^jj-}",
132     "QS L-Front_Consonant{*^f-,*^v-,*^f-,*^p-,*^b-,*^m-,*^w-,*^ww
    -}",
133     "QS L-Central_Consonant{*^t-,*^d-,*^s-,*^si-,*^z-,*^zi-,*^n-,*^
    r-,*^l-,*^t-,*^d-,*^sz-,*^rz-,*^cz-,*^drz-,*^c-,*^dz-,*^ci
    -,*^dzi-}",
134     "QS L-Back_Consonant{*^gs-,*^k-,*^g-,*^ki-,*^gi-,*^ng-,*^x-}",
135     "QS L-Fortis_Consonant{*^gs-,*^cz-,*^f-,*^k-,*^p-,*^s-,*^sz-,*^
    t-,*^ci-,*^c-,*^ki-}",
136     "QS L-Lenis_Consonant{*^drz-,*^v-,*^g-,*^b-,*^rz-,*^z-,*^d-,*^
    dzi-,*^dz-,*^gi-,*^zi-}",

```

```

137 "QS L-Neigther_F_or_L{*^m-,*^n-,*^ni-,*^ng-,*^l-,*^r-,*^w-,*^ww
    -,*^j-,*^jj-}",
138 "QS L-Voiced_Stop{*^b-,*^d-,*^g-}",
139 "QS L-Unvoiced_Stop{*^gs-,*^p-,*^t-,*^k-}",
140 "QS L-Front_Stop{*^b-,*^p-}",
141 "QS L-Central_Stop{*^d-,*^t-}",
142 "QS L-Back_Stop{*^gs-,*^g-,*^k-}",
143 "QS L-Voiced_Fricative{*^v-,*^z-,*^zi-,*^rz-}",
144 "QS L-Unvoiced_Fricative{*^f-,*^s-,*^si-,*^sz-,*^x-}",
145 "QS L-Front_Fricative{*^f-,*^v-}",
146 "QS L-Affricate_Consonant{*^dz-,*^drz-,*^dzi-,*^c-,*^cz-,*^ci
    -}",
147 "QS L-silences{*^pau-}",
148 "QS L-schwa {*^schwa-}",
149 "QS L-a{*^a-}",
150 "QS L-e{*^e-}",
151 "QS L-i{*^i-}",
152 "QS L-y{*^y-}",
153 "QS L-o{*^o-}",
154 "QS L-u{*^u-}",
155 "QS L-p{*^p-}",
156 "QS L-b{*^b-}",
157 "QS L-t{*^t-}",
158 "QS L-d{*^d-}",
159 "QS L-k{*^k-}",
160 "QS L-ki{*^ki-}",
161 "QS L-g{*^g-}",
162 "QS L-gi{*^gi-}",
163 "QS L-f{*^f-}",
164 "QS L-v{*^v-}",
165 "QS L-s{*^s-}",
166 "QS L-si{*^si-}",
167 "QS L-z{*^z-}",
168 "QS L-zi{*^zi-}",
169 "QS L-sz{*^sz-}",
170 "QS L-rz{*^rz-}",
171 "QS L-x{*^x-}",
172 "QS L-c{*^c-}",
173 "QS L-dz{*^dz-}",
174 "QS L-cz{*^cz-}",
175 "QS L-drz{*^drz-}",
176 "QS L-ci{*^ci-}",
177 "QS L-dzi{*^dzi-}",
178 "QS L-m{*^m-}",
179 "QS L-n{*^n-}",
180 "QS L-ni{*^ni-}",
181 "QS L-ng{*^ng-}",

```

```

182     "QS L-l{*^l-}",
183     "QS L-r{*^r-}",
184     "QS L-w{*^w-}",
185     "QS L-ww{*^ww-}",
186     "QS L-j{*^j-}",
187     "QS L-jj{*^jj-}",
188     "QS L-gs                                {*^gs-}",
189 ]
190 QUINTPHONE_SEGMENT_ARTICULATORY_TYPE_C = [
191     "QS C-Stop{*~p+*,*~b+*,*~t+*,*~d+*,*~k+*,*~g+*,*~gs+*}",
192     "QS C-Nasal{*~ww+*,*~jj+*,*~m+*,*~n+*,*~ni+*,*~ng+*}",
193     "QS C-Fricative{*~f+*,*~v+*,*~s+*,*~si+*,*~z+*,*~zi+*,*~sz+*,*~
        rz+*,*~x+*}",
194     "QS C-Front{*~e+*,*~i+*,*~y+*,*~f+*,*~v+*,*~p+*,*~b+*,*~m+*,*~w
        +*,*~ww+*}",
195     "QS C-Central{*~schwa+*,*~a+*,*~t+*,*~d+*,*~s+*,*~si+*,*~z+*,*~
        zi+*,*~n+*,*~r+*,*~l+*,*~t+*,*~d+*,*~sz+*,*~rz+*,*~cz+*,*~
        drz+*,*~c+*,*~dz+*,*~ci+*,*~dzi+*}",
196     "QS C-Back{*~o+*,*~u+*,*~k+*,*~g+*,*~ki+*,*~gi+*,*~ng+*,*~x
        +*,*~gs+*}",
197     "QS C-Front_Vowel{*~e+*,*~i+*,*~y+*}",
198     "QS C-Central_Vowel{*~a+*,*~schwa+*}",
199     "QS C-Back_Vowel{*~o+*,*~u+*}",
200     "QS C-High_Vowel{*~i+*,*~y+*,*~u+*}",
201     "QS C-Medium_Vowel{*~e+*,*~o+*}",
202     "QS C-Low_Vowel{*~a+*}",
203     "QS C-Rounded_Vowel{*~o+*,*~u+*}",
204     "QS C-Unrounded_Vowel{*~a+*,*~e+*,*~i+*,*~y+*}",
205     "QS C-IVowel{*~i+*}",
206     "QS C-EVowel{*~e+*}",
207     "QS C-AVowel{*~a+*}",
208     "QS C-OVowel{*~o+*}",
209     "QS C-UVowel{*~u+*}",
210     "QS C-YVowel                                {*~y+*}",
211     "QS C-SCHWAVowel                            {*~schwa+*}",
212     "QS C-Unvoiced_Consonant{*~p+*,*~t+*,*~k+*,*~ki+*,*~f+*,*~v
        +*,*~s+*,*~sz+*,*~x+*,*~c+*,*~cz+*,*~ci+*,*~gs+*}",
213     "QS C-Voiced_Consonant{*~b+*,*~d+*,*~g+*,*~gi+*,*~v+*,*~z+*,*~
        zi+*,*~rz+*,*~dz+*,*~drz+*,*~dzi+*,*~m+*,*~n+*,*~ni+*,*~ng
        +*,*~l+*,*~r+*,*~w+*,*~ww+*,*~j+*,*~jj+*}",
214     "QS C-Front_Consonant{*~f+*,*~v+*,*~f+*,*~p+*,*~b+*,*~m+*,*~w
        +*,*~ww+*}",
215     "QS C-Central_Consonant{*~t+*,*~d+*,*~s+*,*~si+*,*~z+*,*~zi
        +*,*~n+*,*~r+*,*~l+*,*~t+*,*~d+*,*~sz+*,*~rz+*,*~cz+*,*~drz
        +*,*~c+*,*~dz+*,*~ci+*,*~dzi+*}",
216     "QS C-Back_Consonant{*~k+*,*~g+*,*~ki+*,*~gi+*,*~ng+*,*~x+*,*~
        gs+*}",

```

```

217 "QS C-Fortis_Consonant{* -cz+*,* -f+*,* -k+*,* -p+*,* -s+*,* -sz+*,* -
    t+*,* -ci+*,* -c+*,* -ki+*,* -gs+*}" ,
218 "QS C-Lenis_Consonant{* -drz+*,* -v+*,* -g+*,* -b+*,* -rz+*,* -z+*,* -
    d+*,* -dzi+*,* -dz+*,* -gi+*,* -zi+*}" ,
219 "QS C-Neither_F_or_L{* -m+*,* -n+*,* -ni+*,* -ng+*,* -l+*,* -r+*,* -w
    +*,* -ww+*,* -j+*,* -jj+*}" ,
220 "QS C-Voiced_Stop{* -b+*,* -d+*,* -g+*}" ,
221 "QS C-Unvoiced_Stop{* -p+*,* -t+*,* -k+*,* -gs+*}" ,
222 "QS C-Front_Stop{* -b+*,* -p+*}" ,
223 "QS C-Central_Stop{* -d+*,* -t+*}" ,
224 "QS C-Back_Stop{* -g+*,* -k+*,* -gs+*}" ,
225 "QS C-Voiced_Fricative{* -v+*,* -z+*,* -zi+*,* -rz+*}" ,
226 "QS C-Unvoiced_Fricative{* -f+*,* -s+*,* -si+*,* -sz+*,* -x+*}" ,
227 "QS C-Front_Fricative{* -f+*,* -v+*}" ,
228 "QS C-Affricate_Consonant{* -dz+*,* -drz+*,* -dzi+*,* -c+*,* -cz
    +*,* -ci+*}" ,
229 "QS C-silences{* -pau+*}" ,
230 "QS C-schwa                                     {* -schwa+*}" ,
231 "QS C-a{* -a+*}" ,
232 "QS C-e{* -e+*}" ,
233 "QS C-i{* -i+*}" ,
234 "QS C-y{* -y+*}" ,
235 "QS C-o{* -o+*}" ,
236 "QS C-u{* -u+*}" ,
237 "QS C-p{* -p+*}" ,
238 "QS C-b{* -b+*}" ,
239 "QS C-t{* -t+*}" ,
240 "QS C-d{* -d+*}" ,
241 "QS C-k{* -k+*}" ,
242 "QS C-ki{* -ki+*}" ,
243 "QS C-g{* -g+*}" ,
244 "QS C-gi{* -gi+*}" ,
245 "QS C-f{* -f+*}" ,
246 "QS C-v{* -v+*}" ,
247 "QS C-s{* -s+*}" ,
248 "QS C-si{* -si+*}" ,
249 "QS C-z{* -z+*}" ,
250 "QS C-zi{* -zi+*}" ,
251 "QS C-sz{* -sz+*}" ,
252 "QS C-rz{* -rz+*}" ,
253 "QS C-x{* -x+*}" ,
254 "QS C-c{* -c+*}" ,
255 "QS C-dz{* -dz+*}" ,
256 "QS C-cz{* -cz+*}" ,
257 "QS C-drz{* -drz+*}" ,
258 "QS C-ci{* -ci+*}" ,
259 "QS C-dzi{* -dzi+*}" ,

```



```

260 "QS C-m{*m+*}",
261 "QS C-n{*n+*}",
262 "QS C-ni{*ni+*}",
263 "QS C-ng{*ng+*}",
264 "QS C-l{*l+*}",
265 "QS C-r{*r+*}",
266 "QS C-w{*w+*}",
267 "QS C-ww{*ww+*}",
268 "QS C-j{*j+*}",
269 "QS C-jj{*jj+*}",
270 "QS C-gs                                     {*-gs+*}",
271 ]
272 QUINTPHONE_SEGMENT_ARTICULATORY_TYPE_R = [
273 "QS R-Stop{*p=*,*b=*,*t=*,*d=*,*k=*,*g=*,*gs=*}",
274 "QS R-Nasal{*w=*,*jj=*,*m=*,*n=*,*ni=*,*ng=*}",
275 "QS R-Fricative{*f=*,*v=*,*s=*,*si=*,*z=*,*zi=*,*sz=*,*
    rz=*,*x=*}",
276 "QS R-Front{*e=*,*i=*,*y=*,*f=*,*v=*,*p=*,*b=*,*m=*,*w
    =*,*ww=*}",
277 "QS R-Central{*schwa=*,*a=*,*t=*,*d=*,*s=*,*si=*,*z=*,*
    zi=*,*n=*,*r=*,*l=*,*t=*,*d=*,*sz=*,*rz=*,*cz=*,*
    drz=*,*c=*,*dz=*,*ci=*,*dzi=*}",
278 "QS R-Back{*o=*,*u=*,*k=*,*g=*,*ki=*,*gi=*,*ng=*,*x
    =*,*gs=*}",
279 "QS R-Front_Vowel{*e=*,*i=*,*y=*}",
280 "QS R-Central_Vowel{*a=*,*schwa=*}",
281 "QS R-Back_Vowel{*o=*,*u=*}",
282 "QS R-High_Vowel{*i=*,*y=*,*u=*}",
283 "QS R-Medium_Vowel{*e=*,*o=*}",
284 "QS R-Low_Vowel{*a=*}",
285 "QS R-Rounded_Vowel{*o=*,*u=*}",
286 "QS R-Unrounded_Vowel{*a=*,*e=*,*i=*,*y=*}",
287 "QS R-IVowel{*i=*}",
288 "QS R-OVowel{*o=*}",
289 "QS R-UVowel{*u=*}",
290 "QS R-YVowel                                     {*y=*}",
291 "QS R-SCHWAVowel                                 {*schwa=*}",
292 "QS R-Unvoiced_Consonant{*p=*,*t=*,*k=*,*ki=*,*f=*,*v
    =*,*s=*,*sz=*,*x=*,*c=*,*cz=*,*ci=*,*gs=*}",
293 "QS R-Voiced_Consonant{*b=*,*d=*,*g=*,*gi=*,*v=*,*z=*,*
    zi=*,*rz=*,*dz=*,*drz=*,*dzi=*,*m=*,*n=*,*ni=*,*ng
    =*,*l=*,*r=*,*w=*,*ww=*,*j=*,*jj=*}",
294 "QS R-Front_Consonant{*f=*,*v=*,*f=*,*p=*,*b=*,*m=*,*w
    =*,*ww=*}",
295 "QS R-Central_Consonant{*t=*,*d=*,*s=*,*si=*,*z=*,*zi
    =*,*n=*,*r=*,*l=*,*t=*,*d=*,*sz=*,*rz=*,*cz=*,*drz
    =*,*c=*,*dz=*,*ci=*,*dzi=*}",

```

```

296 "QS R-Back_Consonant{**k=*,**g=*,**ki=*,**gi=*,**ng=*,**x=*,**
    gs=*}",
297 "QS R-Fortis_Consonant{**cz=*,**f=*,**k=*,**p=*,**s=*,**sz=*,**
    t=*,**ci=*,**c=*,**ki=*,**gs=*}",
298 "QS R-Lenis_Consonant{**drz=*,**v=*,**g=*,**b=*,**rz=*,**z=*,**
    d=*,**dzi=*,**dz=*,**gi=*,**zi=*}",
299 "QS R-Neigther_F_or_L{**m=*,**n=*,**ni=*,**ng=*,**l=*,**r=*,**w
    =*,**ww=*,**j=*,**jj=*}",
300 "QS R-Voiced_Stop{**b=*,**d=*,**g=*}",
301 "QS R-Unvoiced_Stop{**p=*,**t=*,**k=*,**gs=*}",
302 "QS R-Front_Stop{**b=*,**p=*}",
303 "QS R-Central_Stop{**d=*,**t=*}",
304 "QS R-Back_Stop{**g=*,**k=*,**gs=*}",
305 "QS R-Voiced_Fricative{**v=*,**z=*,**zi=*,**rz=*}",
306 "QS R-Unvoiced_Fricative{**f=*,**s=*,**si=*,**sz=*,**x=*}",
307 "QS R-Front_Fricative{**f=*,**v=*}",
308 "QS R-Affricate_Consonant{**dz=*,**drz=*,**dzi=*,**c=*,**cz
    =*,**ci=*}",
309 "QS R-silences{**pau=*}",
310 "QS R-schwa {**schwa=*}",
311 "QS R-a{**a=*}",
312 "QS R-e{**e=*}",
313 "QS R-i{**i=*}",
314 "QS R-y{**y=*}",
315 "QS R-o{**o=*}",
316 "QS R-u{**u=*}",
317 "QS R-p{**p=*}",
318 "QS R-b{**b=*}",
319 "QS R-t{**t=*}",
320 "QS R-d{**d=*}",
321 "QS R-k{**k=*}",
322 "QS R-ki{**ki=*}",
323 "QS R-g{**g=*}",
324 "QS R-gi{**gi=*}",
325 "QS R-f{**f=*}",
326 "QS R-v{**v=*}",
327 "QS R-s{**s=*}",
328 "QS R-si{**si=*}",
329 "QS R-z{**z=*}",
330 "QS R-zi{**zi=*}",
331 "QS R-sz{**sz=*}",
332 "QS R-rz{**rz=*}",
333 "QS R-x{**x=*}",
334 "QS R-c{**c=*}",
335 "QS R-dz{**dz=*}",
336 "QS R-cz{**cz=*}",
337 "QS R-drz{**drz=*}",

```

```

338 "QS R-ci{**ci=*}",
339 "QS R-dzi{**dzi=*}",
340 "QS R-m{**m=*}",
341 "QS R-n{**n=*}",
342 "QS R-ni{**ni=*}",
343 "QS R-ng{**ng=*}",
344 "QS R-l{**l=*}",
345 "QS R-r{**r=*}",
346 "QS R-w{**w=*}",
347 "QS R-ww{**ww=*}",
348 "QS R-j{**j=*}",
349 "QS R-jj{**jj=*}",
350 "QS R-gs{**gs=*}",
351 ]
352 QUINTPHONE_SEGMENT_ARTICULATORY_TYPE_RR = [
353 "QS RR-Stop{**gs@*,**p@*,**b@*,**t@*,**d@*,**k@*,**g@*}",
354 "QS RR-Nasal{**ww@*,**jj@*,**m@*,**n@*,**ni@*,**ng@*}",
355 "QS RR-Fricative{**f@*,**v@*,**s@*,**si@*,**z@*,**zi@*,**sz@
    *,**rz@*,**x@*}",
356 "QS RR-Front{**e@*,**i@*,**y@*,**f@*,**v@*,**p@*,**b@*,**m@*,**
    w@*,**ww@*}",
357 "QS RR-Central{**schwa@*,**a@*,**t@*,**d@*,**s@*,**si@*,**z@
    *,**zi@*,**n@*,**r@*,**l@*,**t@*,**d@*,**sz@*,**rz@*,**cz@
    *,**drz@*,**c@*,**dz@*,**ci@*,**dzi@*}",
358 "QS RR-Back{**o@*,**u@*,**k@*,**g@*,**ki@*,**gi@*,**ng@*,**x@
    *,**gs@*}",
359 "QS RR-Front_Vowel{**e@*,**i@*,**y@*}",
360 "QS RR-Central_Vowel{**a@*,**schwa@*}",
361 "QS RR-Back_Vowel{**o@*,**u@*}",
362 "QS RR-High_Vowel{**i@*,**y@*,**u@*}",
363 "QS RR-Medium_Vowel{**e@*,**o@*}",
364 "QS RR-Low_Vowel{**a@*}",
365 "QS RR-Rounded_Vowel{**o@*,**u@*}",
366 "QS RR-Unrounded_Vowel{**a@*,**e@*,**i@*,**y@*}",
367 "QS RR-IVowel{**i@*}",
368 "QS RR-OVowel{**o@*}",
369 "QS RR-UVowel{**u@*}",
370 "QS RR-YVowel {**y@*}",
371 "QS RR-SCHWAVowel {**schwa@*}",
372 "QS RR-Unvoiced_Consonant{**p@*,**t@*,**k@*,**ki@*,**f@*,**v@
    *,**s@*,**sz@*,**x@*,**c@*,**cz@*,**ci@*,**gs@*}",
373 "QS RR-Voiced_Consonant{**b@*,**d@*,**g@*,**gi@*,**v@*,**z@*,**
    zi@*,**rz@*,**dz@*,**drz@*,**dzi@*,**m@*,**n@*,**ni@*,**ng@
    *,**l@*,**r@*,**w@*,**ww@*,**j@*,**jj@*}",
374 "QS RR-Front_Consonant{**f@*,**v@*,**f@*,**p@*,**b@*,**m@*,**w@
    *,**ww@*}",

```

```

375 "QS RR-Central_Consonant{*=t@*,*=d@*,*=s@*,*=si@*,*=z@*,*=zi@
    *,*=n@*,*=r@*,*=l@*,*=t@*,*=d@*,*=sz@*,*=rz@*,*=cz@*,*=drz@
    *,*=c@*,*=dz@*,*=ci@*,*=dzi@*}",
376 "QS RR-Back_Consonant{*=k@*,*=g@*,*=ki@*,*=gi@*,*=ng@*,*=x@*,*=
    gs@*}",
377 "QS RR-Fortis_Consonant{*=cz@*,*=f@*,*=k@*,*=p@*,*=s@*,*=sz@
    *,*=t@*,*=ci@*,*=c@*,*=ki@*,*=gs@*}",
378 "QS RR-Lenis_Consonant{*=drz@*,*=v@*,*=g@*,*=b@*,*=rz@*,*=z@
    *,*=d@*,*=dzi@*,*=dz@*,*=gi@*,*=zi@*}",
379 "QS RR-Neither_F_or_L{*=m@*,*=n@*,*=ni@*,*=ng@*,*=l@*,*=r@*,*=
    w@*,*=ww@*,*=j@*,*=jj@*}",
380 "QS RR-Voiced_Stop{*=b@*,*=d@*,*=g@*}",
381 "QS RR-Unvoiced_Stop{*=p@*,*=t@*,*=k@*,*=gs@*}",
382 "QS RR-Front_Stop{*=b@*,*=p@*}",
383 "QS RR-Central_Stop{*=d@*,*=t@*}",
384 "QS RR-Back_Stop{*=g@*,*=k@*,*=gs@*}",
385 "QS RR-Voiced_Fricative{*=v@*,*=z@*,*=zi@*,*=rz@*}",
386 "QS RR-Unvoiced_Fricative{*=f@*,*=s@*,*=si@*,*=sz@*,*=x@*}",
387 "QS RR-Front_Fricative{*=f@*,*=v@*}",
388 "QS RR-Affricate_Consonant{*=dz@*,*=drz@*,*=dzi@*,*=c@*,*=cz@
    *,*=ci@*}",
389 "QS RR-silences{*=pau@*}",
390 "QS RR-a{*=a@*}",
391 "QS RR-e{*=e@*}",
392 "QS RR-i{*=i@*}",
393 "QS RR-y{*=y@*}",
394 "QS RR-o{*=o@*}",
395 "QS RR-u{*=u@*}",
396 "QS RR-p{*=p@*}",
397 "QS RR-b{*=b@*}",
398 "QS RR-t{*=t@*}",
399 "QS RR-d{*=d@*}",
400 "QS RR-k{*=k@*}",
401 "QS RR-ki{*=ki@*}",
402 "QS RR-g{*=g@*}",
403 "QS RR-gi{*=gi@*}",
404 "QS RR-f{*=f@*}",
405 "QS RR-v{*=v@*}",
406 "QS RR-s{*=s@*}",
407 "QS RR-si{*=si@*}",
408 "QS RR-z{*=z@*}",
409 "QS RR-zi{*=zi@*}",
410 "QS RR-sz{*=sz@*}",
411 "QS RR-rz{*=rz@*}",
412 "QS RR-x{*=x@*}",
413 "QS RR-c{*=c@*}",
414 "QS RR-dz{*=dz@*}",

```

```

415     "QS RR-cz{*=cz@*}",
416     "QS RR-drz{*=drz@*}",
417     "QS RR-ci{*=ci@*}",
418     "QS RR-dzi{*=dzi@*}",
419     "QS RR-m{*=m@*}",
420     "QS RR-n{*=n@*}",
421     "QS RR-ni{*=ni@*}",
422     "QS RR-ng{*=ng@*}",
423     "QS RR-l{*=l@*}",
424     "QS RR-r{*=r@*}",
425     "QS RR-w{*=w@*}",
426     "QS RR-ww{*=ww@*}",
427     "QS RR-j{*=j@*}",
428     "QS RR-jj{*=jj@*}",
429     "QS RR-gs                                     {*=gs@*}",
430 ]
431
432 NUM_SEG_IN_SYL_FW_EQ = [
433     "QS Seg_Fw==x{*@x_*}",
434     "QS Seg_Fw==1{*@1_*}",
435     "QS Seg_Fw==2{*@2_*}",
436     "QS Seg_Fw==3{*@3_*}",
437     "QS Seg_Fw==4{*@4_*}",
438     "QS Seg_Fw==5{*@5_*}",
439     "QS Seg_Fw==6{*@6_*}",
440     "QS Seg_Fw==7{*@7_*}",
441 ]
442 NUM_SEG_IN_SYL_FW_LTE = [
443     "QS Seg_Fw<=1{*@x_*,*@1_*}",
444     "QS Seg_Fw<=2{*@x_*,*@1_*,*@2_*}",
445     "QS Seg_Fw<=3{*@x_*,*@1_*,*@2_*,*@3_*}",
446     "QS Seg_Fw<=4{*@x_*,*@1_*,*@2_*,*@3_*,*@4_*}",
447     "QS Seg_Fw<=5{*@x_*,*@1_*,*@2_*,*@3_*,*@4_*,*@5_*}",
448     "QS Seg_Fw<=6{*@x_*,*@1_*,*@2_*,*@3_*,*@4_*,*@5_*,*@6_*}",
449     "QS Seg_Fw<=7{*@x_*,*@1_*,*@2_*,*@3_*,*@4_*,*@5_*,*@6_*,*@7_
        *}",
450 ]
451 NUM_SEG_IN_SYL_BW_EQ = [
452     "QS Seg_Bw==x{*_x/A:*}",
453     "QS Seg_Bw==1{*_1/A:*}",
454     "QS Seg_Bw==2{*_2/A:*}",
455     "QS Seg_Bw==3{*_3/A:*}",
456     "QS Seg_Bw==4{*_4/A:*}",
457     "QS Seg_Bw==5{*_5/A:*}",
458     "QS Seg_Bw==6{*_6/A:*}",
459     "QS Seg_Bw==7{*_7/A:*}",
460 ]

```

```

461 NUM_SEG_IN_SYL_BW_LTE = [
462     "QS Seg_Bw<=0{*_x/A:*,*_0/A:*}",
463     "QS Seg_Bw<=1{*_x/A:*,*_0/A:*,*_1/A:*}",
464     "QS Seg_Bw<=2{*_x/A:*,*_0/A:*,*_1/A:*,*_2/A:*}",
465     "QS Seg_Bw<=3{*_x/A:*,*_0/A:*,*_1/A:*,*_2/A:*,*_3/A:*}",
466     "QS Seg_Bw<=4{*_x/A:*,*_0/A:*,*_1/A:*,*_2/A:*,*_3/A:*,*_4/A
        :*}" ,
467     "QS Seg_Bw<=5{*_x/A:*,*_0/A:*,*_1/A:*,*_2/A:*,*_3/A:*,*_4/A:*,*
        _5/A:*}" ,
468     "QS Seg_Bw<=6{*_x/A:*,*_0/A:*,*_1/A:*,*_2/A:*,*_3/A:*,*_4/A:*,*
        _5/A:*,*_6/A:*}" ,
469     "QS Seg_Bw<=7{*_x/A:*,*_0/A:*,*_1/A:*,*_2/A:*,*_3/A:*,*_4/A:*,*
        _5/A:*,*_6/A:*,*_7/A:*}" ,
470 ]
471
472 SYL_STRESSED_L = [
473     "QS L-Syl_Stress==1{*_1_*}" ,
474     "QS L-Syl_Stress==0{*_0_*}" ,
475 ]
476 SYL_ACCENTED_L = [
477     "QS L-Syl_Accent==1{*_1_*}" ,
478     "QS L-Syl_Accent==0{*_0_*}" ,
479 ]
480 SYL_ACCENT_TYPE_L = [
481     "QS L-Syl_TOBI_Accent-H{*_K:H?/L:*}" ,
482     "QS L-Syl_TOBI_Accent-L{*_K:L?/L:*}" ,
483     "QS L-Syl_TOBI_Accent-L+H{*_K:L?+H*}" ,
484     "QS L-Syl_TOBI_Accent-L+H{*_K:L+H*}" ,
485     "QS L-Syl_TOBI_Accent-O{*_K:O*}" ,
486     "QS L-Syl_TOBI_Accent-NONE{*_K:NONE*}" ,
487     "QS L-Syl_TOBI_Accent-x{*_K:x*}" ,
488 ]
489
490 SYL_LEN_IN_NO_SEG_EQ_L = [
491     "QS L-Syl_Num-Segs==0{*_0/B:*}" ,
492     "QS L-Syl_Num-Segs==1{*_1/B:*}" ,
493     "QS L-Syl_Num-Segs==2{*_2/B:*}" ,
494     "QS L-Syl_Num-Segs==3{*_3/B:*}" ,
495     "QS L-Syl_Num-Segs==4{*_4/B:*}" ,
496     "QS L-Syl_Num-Segs==5{*_5/B:*}" ,
497     "QS L-Syl_Num-Segs==6{*_6/B:*}" ,
498     "QS L-Syl_Num-Segs==7{*_7/B:*}" ,
499 ]
500 SYL_LEN_IN_NO_SEG_LTE_L = [
501     "QS L-Syl_Num-Segs<=1{*_0/B:*,*_1/B:*}" ,
502     "QS L-Syl_Num-Segs<=2{*_0/B:*,*_1/B:*,*_2/B:*}" ,
503     "QS L-Syl_Num-Segs<=3{*_0/B:*,*_1/B:*,*_2/B:*,*_3/B:*}" ,

```

```

504     "QS L-Syl_Num-Segs <=4{*_0/B:*,*_1/B:*,*_2/B:*,*_3/B:*,*_4/B
      :*}",
505     "QS L-Syl_Num-Segs <=5{*_0/B:*,*_1/B:*,*_2/B:*,*_3/B:*,*_4/B:*,*
      _5/B:*}",
506     "QS L-Syl_Num-Segs <=6{*_0/B:*,*_1/B:*,*_2/B:*,*_3/B:*,*_4/B:*,*
      _5/B:*,*_6/B:*}",
507     "QS L-Syl_Num-Segs <=7{*_0/B:*,*_1/B:*,*_2/B:*,*_3/B:*,*_4/B:*,*
      _5/B:*,*_6/B:*,*_7/B:*}",
508 ]
509
510 SYL_STRESSED_C = [
511     "QS C-Syl_Stress==1{*/B:1-*}",
512     "QS C-Syl_Stress==0{*/B:0-*}",
513     "QS C-Syl_Stress==x{*/B:x-*}",
514 ]
515 SYL_ACCENTED_C = [
516     "QS C-Syl_Accent==1{*1-*}",
517     "QS C-Syl_Accent==0{*0-*}",
518     "QS C-Syl_Accent==x{*x-*}",
519 ]
520 SYL_ACCENT_TYPE_C = [
521     "QS C-Syl_TOBI_Accent-H{*/*L:H*/M:*}",
522     "QS C-Syl_TOBI_Accent-L{*/*L:L*/M:*}",
523     "QS C-Syl_TOBI_Accent-L*+H{*/*L:L?+H*}",
524     "QS C-Syl_TOBI_Accent-L+H{*/*L:L+H*}",
525     "QS C-Syl_TOBI_Accent-O{*/*L:0*}",
526     "QS C-Syl_TOBI_Accent-NONE{*/*L:NONE*}",
527     "QS C-Syl_TOBI_Accent-x{*/*L:x*}",
528 ]
529
530 SYL_LEN_IN_NO_SEG_EQ_C = [
531     "QS C-Syl_Num-Segs==x{*x@*}",
532     "QS C-Syl_Num-Segs==1{*1@*}",
533     "QS C-Syl_Num-Segs==2{*2@*}",
534     "QS C-Syl_Num-Segs==3{*3@*}",
535     "QS C-Syl_Num-Segs==4{*4@*}",
536     "QS C-Syl_Num-Segs==5{*5@*}",
537     "QS C-Syl_Num-Segs==6{*6@*}",
538     "QS C-Syl_Num-Segs==7{*7@*}",
539 ]
540 SYL_LEN_IN_NO_SEG_LTE_C = [
541     "QS C-Syl_Num-Segs <=1{*x@*,*-1@*}",
542     "QS C-Syl_Num-Segs <=2{*x@*,*-1@*,*-2@*}",
543     "QS C-Syl_Num-Segs <=3{*x@*,*-1@*,*-2@*,*-3@*}",
544     "QS C-Syl_Num-Segs <=4{*x@*,*-1@*,*-2@*,*-3@*,*-4@*}",
545     "QS C-Syl_Num-Segs <=5{*x@*,*-1@*,*-2@*,*-3@*,*-4@*,*-5@*}",

```

```

546     "QS C-Syl_Num-Segs<=6{*~x@*,~-1@*,~-2@*,~-3@*,~-4@*,~-5@*,~-6@
547     *}",
548 ]
549
550 SYL_POSITION_IN_WORD_FW_EQ = [
551     "QS Pos_C-Syl_in_C-Word(Fw)==x{*@x-*}",
552     "QS Pos_C-Syl_in_C-Word(Fw)==1{*@1-*}",
553     "QS Pos_C-Syl_in_C-Word(Fw)==2{*@2-*}",
554     "QS Pos_C-Syl_in_C-Word(Fw)==3{*@3-*}",
555     "QS Pos_C-Syl_in_C-Word(Fw)==4{*@4-*}",
556     "QS Pos_C-Syl_in_C-Word(Fw)==5{*@5-*}",
557     "QS Pos_C-Syl_in_C-Word(Fw)==6{*@6-*}",
558     "QS Pos_C-Syl_in_C-Word(Fw)==7{*@7-*}",
559 ]
560 SYL_POSITION_IN_WORD_FW_LTE = [
561     "QS Pos_C-Syl_in_C-Word(Fw)<=1{*@x-*,*@1-*}",
562     "QS Pos_C-Syl_in_C-Word(Fw)<=2{*@x-*,*@1-*,*@2-*}",
563     "QS Pos_C-Syl_in_C-Word(Fw)<=3{*@x-*,*@1-*,*@2-*,*@3-*}",
564     "QS Pos_C-Syl_in_C-Word(Fw)<=4{*@x-*,*@1-*,*@2-*,*@3-*,*@4-*}",
565     "QS Pos_C-Syl_in_C-Word(Fw)<=5{*@x-*,*@1-*,*@2-*,*@3-*,*@4-*,*
566     @5-*}",
567     "QS Pos_C-Syl_in_C-Word(Fw)<=6{*@x-*,*@1-*,*@2-*,*@3-*,*@4-*,*
568     @5-*,*@6-*}",
569     "QS Pos_C-Syl_in_C-Word(Fw)<=7{*@x-*,*@1-*,*@2-*,*@3-*,*@4-*,*
570     @5-*,*@6-*,*@7-*}",
571 ]
572 SYL_POSITION_IN_WORD_BW_EQ = [
573     "QS Pos_C-Syl_in_C-Word(Bw)==x{*~x&*}",
574     "QS Pos_C-Syl_in_C-Word(Bw)==1{*~1&*}",
575     "QS Pos_C-Syl_in_C-Word(Bw)==2{*~2&*}",
576     "QS Pos_C-Syl_in_C-Word(Bw)==3{*~3&*}",
577     "QS Pos_C-Syl_in_C-Word(Bw)==4{*~4&*}",
578     "QS Pos_C-Syl_in_C-Word(Bw)==5{*~5&*}",
579     "QS Pos_C-Syl_in_C-Word(Bw)==6{*~6&*}",
580     "QS Pos_C-Syl_in_C-Word(Bw)==7{*~7&*}",
581 ]
582 SYL_POSITION_IN_WORD_BW_LTE = [
583     "QS Pos_C-Syl_in_C-Word(Bw)<=1{*~x&*,~-1&*}",
584     "QS Pos_C-Syl_in_C-Word(Bw)<=2{*~x&*,~-1&*,~-2&*}",
585     "QS Pos_C-Syl_in_C-Word(Bw)<=3{*~x&*,~-1&*,~-2&*,~-3&*}",
586     "QS Pos_C-Syl_in_C-Word(Bw)<=4{*~x&*,~-1&*,~-2&*,~-3&*,~-4&*}",
587     "QS Pos_C-Syl_in_C-Word(Bw)<=5{*~x
588     &*,~-1&*,~-2&*,~-3&*,~-4&*,~-5&*}",
589     "QS Pos_C-Syl_in_C-Word(Bw)<=6{*~x
590     &*,~-1&*,~-2&*,~-3&*,~-4&*,~-5&*,~-6&*}",

```



```

586     "QS Pos_C-Syl_in_C-Word(Bw) <=7{*x
        &*,*-1&*,*-2&*,*-3&*,*-4&*,*-5&*,*-6&*,*-7&}",
587 ]
588 SYL_POSITION_IN_PHRASE_FW_EQ = [
589     "QS Pos_C-Syl_in_C-Phrase(Fw) ==x{*&x-*}",
590     "QS Pos_C-Syl_in_C-Phrase(Fw) ==1{*&1-*}",
591     "QS Pos_C-Syl_in_C-Phrase(Fw) ==2{*&2-*}",
592     "QS Pos_C-Syl_in_C-Phrase(Fw) ==3{*&3-*}",
593     "QS Pos_C-Syl_in_C-Phrase(Fw) ==4{*&4-*}",
594     "QS Pos_C-Syl_in_C-Phrase(Fw) ==5{*&5-*}",
595     "QS Pos_C-Syl_in_C-Phrase(Fw) ==6{*&6-*}",
596     "QS Pos_C-Syl_in_C-Phrase(Fw) ==7{*&7-*}",
597     "QS Pos_C-Syl_in_C-Phrase(Fw) ==8{*&8-*}",
598     "QS Pos_C-Syl_in_C-Phrase(Fw) ==9{*&9-*}",
599     "QS Pos_C-Syl_in_C-Phrase(Fw) ==10{*&10-*}",
600     "QS Pos_C-Syl_in_C-Phrase(Fw) ==11{*&11-*}",
601     "QS Pos_C-Syl_in_C-Phrase(Fw) ==12{*&12-*}",
602     "QS Pos_C-Syl_in_C-Phrase(Fw) ==13{*&13-*}",
603     "QS Pos_C-Syl_in_C-Phrase(Fw) ==14{*&14-*}",
604     "QS Pos_C-Syl_in_C-Phrase(Fw) ==15{*&15-*}",
605     "QS Pos_C-Syl_in_C-Phrase(Fw) ==16{*&16-*}",
606     "QS Pos_C-Syl_in_C-Phrase(Fw) ==17{*&17-*}",
607     "QS Pos_C-Syl_in_C-Phrase(Fw) ==18{*&18-*}",
608     "QS Pos_C-Syl_in_C-Phrase(Fw) ==19{*&19-*}",
609     "QS Pos_C-Syl_in_C-Phrase(Fw) ==20{*&20-*}",
610 ]
611 SYL_POSITION_IN_PHRASE_FW_LTE = [
612     "QS Pos_C-Syl_in_C-Phrase(Fw) <=1{*&x-*,*&0-*,*&1-*}",
613     "QS Pos_C-Syl_in_C-Phrase(Fw) <=2{*&x-*,*&0-*,*&1-*,*&2-*}",
614     "QS Pos_C-Syl_in_C-Phrase(Fw) <=3{*&x
        -*,*&0-*,*&1-*,*&2-*,*&3-*}",
615     "QS Pos_C-Syl_in_C-Phrase(Fw) <=4{*&x
        -*,*&0-*,*&1-*,*&2-*,*&3-*,*&4-*}",
616     "QS Pos_C-Syl_in_C-Phrase(Fw) <=5{*&x
        -*,*&0-*,*&1-*,*&2-*,*&3-*,*&4-*,*&5-*}",
617     "QS Pos_C-Syl_in_C-Phrase(Fw) <=6{*&x
        -*,*&0-*,*&1-*,*&2-*,*&3-*,*&4-*,*&5-*,*&6-*}",
618     "QS Pos_C-Syl_in_C-Phrase(Fw) <=7{*&x
        -*,*&0-*,*&1-*,*&2-*,*&3-*,*&4-*,*&5-*,*&6-*,*&7-*}",
619     "QS Pos_C-Syl_in_C-Phrase(Fw) <=8{*&x
        -*,*&0-*,*&1-*,*&2-*,*&3-*,*&4-*,*&5-*,*&6-*,*&7-*,*&8-*}",
620     "QS Pos_C-Syl_in_C-Phrase(Fw) <=9{*&?-*}",
621     "QS Pos_C-Syl_in_C-Phrase(Fw) <=10{*&?-*,*&10-*}",
622     "QS Pos_C-Syl_in_C-Phrase(Fw) <=11{*&?-*,*&10-*,*&11-*}",
623     "QS Pos_C-Syl_in_C-Phrase(Fw) <=12{*&?-*,*&10-*,*&11-*,*&12-*}",
624     "QS Pos_C-Syl_in_C-Phrase(Fw)
        <=13{*&?-*,*&10-*,*&11-*,*&12-*,*&13-*}",

```

```

625 "QS Pos_C-Syl_in_C-Phrase(Fw)
    <=14{*&?-*,*&10-*,*&11-*,*&12-*,*&13-*,*&14-*}",
626 "QS Pos_C-Syl_in_C-Phrase(Fw)
    <=15{*&?-*,*&10-*,*&11-*,*&12-*,*&13-*,*&14-*,*&15-*}",
627 "QS Pos_C-Syl_in_C-Phrase(Fw)
    <=16{*&?-*,*&10-*,*&11-*,*&12-*,*&13-*,*&14-*,*&15-*,*&16-*}",

628 "QS Pos_C-Syl_in_C-Phrase(Fw)
    <=17{*&?-*,*&10-*,*&11-*,*&12-*,*&13-*,*&14-*,*&15-*,*&16-*,*&17-*}",

629 "QS Pos_C-Syl_in_C-Phrase(Fw)
    <=18{*&?-*,*&10-*,*&11-*,*&12-*,*&13-*,*&14-*,*&15-*,*&16-*,*&17-*,*&18-*}",

630 "QS Pos_C-Syl_in_C-Phrase(Fw) <=19{*&?-*,*&1?-*}",
631 "QS Pos_C-Syl_in_C-Phrase(Fw) <=20{*&?-*,*&1?-*,*&20-*}",
632 ]
633 SYL_POSITION_IN_PHRASE_BW_EQ = [
634     "QS Pos_C-Syl_in_C-Phrase(Bw)==x{* -x#*}",
635     "QS Pos_C-Syl_in_C-Phrase(Bw)==1{* -1#*}",
636     "QS Pos_C-Syl_in_C-Phrase(Bw)==2{* -2#*}",
637     "QS Pos_C-Syl_in_C-Phrase(Bw)==3{* -3#*}",
638     "QS Pos_C-Syl_in_C-Phrase(Bw)==4{* -4#*}",
639     "QS Pos_C-Syl_in_C-Phrase(Bw)==5{* -5#*}",
640     "QS Pos_C-Syl_in_C-Phrase(Bw)==6{* -6#*}",
641     "QS Pos_C-Syl_in_C-Phrase(Bw)==7{* -7#*}",
642     "QS Pos_C-Syl_in_C-Phrase(Bw)==8{* -8#*}",
643     "QS Pos_C-Syl_in_C-Phrase(Bw)==9{* -9#*}",
644     "QS Pos_C-Syl_in_C-Phrase(Bw)==10{* -10#*}",
645     "QS Pos_C-Syl_in_C-Phrase(Bw)==11{* -11#*}",
646     "QS Pos_C-Syl_in_C-Phrase(Bw)==12{* -12#*}",
647     "QS Pos_C-Syl_in_C-Phrase(Bw)==13{* -13#*}",
648     "QS Pos_C-Syl_in_C-Phrase(Bw)==14{* -14#*}",
649     "QS Pos_C-Syl_in_C-Phrase(Bw)==15{* -15#*}",
650     "QS Pos_C-Syl_in_C-Phrase(Bw)==16{* -16#*}",
651     "QS Pos_C-Syl_in_C-Phrase(Bw)==17{* -17#*}",
652     "QS Pos_C-Syl_in_C-Phrase(Bw)==18{* -18#*}",
653     "QS Pos_C-Syl_in_C-Phrase(Bw)==19{* -19#*}",
654     "QS Pos_C-Syl_in_C-Phrase(Bw)==20{* -20#*}",
655 ]
656 SYL_POSITION_IN_PHRASE_BW_LTE = [
657     "QS Pos_C-Syl_in_C-Phrase(Bw) <=1{* -x#*,*-1#*}",
658     "QS Pos_C-Syl_in_C-Phrase(Bw) <=2{* -x#*,*-1#*,*-2#*}",
659     "QS Pos_C-Syl_in_C-Phrase(Bw) <=3{* -x#*,*-1#*,*-2#*,*-3#*}",
660     "QS Pos_C-Syl_in_C-Phrase(Bw) <=4{* -x
        #*,*-1#*,*-2#*,*-3#*,*-4#*}",
661     "QS Pos_C-Syl_in_C-Phrase(Bw) <=5{* -x
        #*,*-1#*,*-2#*,*-3#*,*-4#*,*-5#*}",

```

```

662 "QS Pos_C-Syl_in_C-Phrase(Bw) <=6{*x
    #*,*-1#*,*-2#*,*-3#*,*-4#*,*-5#*,*-6#*}",
663 "QS Pos_C-Syl_in_C-Phrase(Bw) <=7{*x
    #*,*-1#*,*-2#*,*-3#*,*-4#*,*-5#*,*-6#*,*-7#*}",
664 "QS Pos_C-Syl_in_C-Phrase(Bw) <=8{*x
    #*,*-1#*,*-2#*,*-3#*,*-4#*,*-5#*,*-6#*,*-7#*,*-8#*}",
665 "QS Pos_C-Syl_in_C-Phrase(Bw) <=9{*-*#*}",
666 "QS Pos_C-Syl_in_C-Phrase(Bw) <=10{*-*#*,*-10#*}",
667 "QS Pos_C-Syl_in_C-Phrase(Bw) <=11{*-*#*,*-10#*,*-11#*}",
668 "QS Pos_C-Syl_in_C-Phrase(Bw) <=12{*-*#*,*-10#*,*-11#*,*-12#*}",
669 "QS Pos_C-Syl_in_C-Phrase(Bw)
    <=13{*-*#*,*-10#*,*-11#*,*-12#*,*-13#*}",
670 "QS Pos_C-Syl_in_C-Phrase(Bw)
    <=14{*-*#*,*-10#*,*-11#*,*-12#*,*-13#*,*-14#*}",
671 "QS Pos_C-Syl_in_C-Phrase(Bw)
    <=15{*-*#*,*-10#*,*-11#*,*-12#*,*-13#*,*-14#*,*-15#*}",
672 "QS Pos_C-Syl_in_C-Phrase(Bw)
    <=16{*-*#*,*-10#*,*-11#*,*-12#*,*-13#*,*-14#*,*-15#*,*-16#*}",

673 "QS Pos_C-Syl_in_C-Phrase(Bw)
    <=17{*-*#*,*-10#*,*-11#*,*-12#*,*-13#*,*-14#*,*-15#*,*-16#*,*-17#*}",

674 "QS Pos_C-Syl_in_C-Phrase(Bw)
    <=18{*-*#*,*-10#*,*-11#*,*-12#*,*-13#*,*-14#*,*-15#*,*-16#*,*-17#*,*-18#*}",

675 "QS Pos_C-Syl_in_C-Phrase(Bw) <=19{*-*#*,*-1?#*}",
676 "QS Pos_C-Syl_in_C-Phrase(Bw) <=20{*-*#*,*-1?#*,*-20#*}",
677 ]
678
679 NUM_STRESSED_SYLS_BEFORE_CURRENT_SYL_IN_CURRENT_PHRASE_EQ = [
680 "QS Num-StressedSyl_before_C-Syl_in_C-Phrase==x{##x-*}",
681 "QS Num-StressedSyl_before_C-Syl_in_C-Phrase==1{##1-*}",
682 "QS Num-StressedSyl_before_C-Syl_in_C-Phrase==2{##2-*}",
683 "QS Num-StressedSyl_before_C-Syl_in_C-Phrase==3{##3-*}",
684 "QS Num-StressedSyl_before_C-Syl_in_C-Phrase==4{##4-*}",
685 "QS Num-StressedSyl_before_C-Syl_in_C-Phrase==5{##5-*}",
686 "QS Num-StressedSyl_before_C-Syl_in_C-Phrase==6{##6-*}",
687 "QS Num-StressedSyl_before_C-Syl_in_C-Phrase==7{##7-*}",
688 "QS Num-StressedSyl_before_C-Syl_in_C-Phrase==8{##8-*}",
689 "QS Num-StressedSyl_before_C-Syl_in_C-Phrase==9{##9-*}",
690 "QS Num-StressedSyl_before_C-Syl_in_C-Phrase==10{##10-*}",
691 "QS Num-StressedSyl_before_C-Syl_in_C-Phrase==11{##11-*}",
692 "QS Num-StressedSyl_before_C-Syl_in_C-Phrase==12{##12-*}",
693 ]
694 NUM_STRESSED_SYLS_BEFORE_CURRENT_SYL_IN_CURRENT_PHRASE_LTE = [
695 "QS Num-StressedSyl_before_C-Syl_in_C-Phrase<=1{##x-*,##1-*}",

```

```

696 "QS Num-StressedSyl_before_C-Syl_in_C-Phrase<=2{##x
    -*,##1-*,##2-*}",
697 "QS Num-StressedSyl_before_C-Syl_in_C-Phrase<=3{##x
    -*,##1-*,##2-*,##3-*}",
698 "QS Num-StressedSyl_before_C-Syl_in_C-Phrase<=4{##x
    -*,##1-*,##2-*,##3-*,##4-*}",
699 "QS Num-StressedSyl_before_C-Syl_in_C-Phrase<=5{##x
    -*,##1-*,##2-*,##3-*,##4-*,##5-*}",
700 "QS Num-StressedSyl_before_C-Syl_in_C-Phrase<=6{##x
    -*,##1-*,##2-*,##3-*,##4-*,##5-*,##6-*}",
701 "QS Num-StressedSyl_before_C-Syl_in_C-Phrase<=7{##x
    -*,##1-*,##2-*,##3-*,##4-*,##5-*,##6-*,##7-*}",
702 "QS Num-StressedSyl_before_C-Syl_in_C-Phrase<=8{##x
    -*,##1-*,##2-*,##3-*,##4-*,##5-*,##6-*,##7-*,##8-*}",
703 "QS Num-StressedSyl_before_C-Syl_in_C-Phrase<=9{##?-*}",
704 "QS Num-StressedSyl_before_C-Syl_in_C-Phrase
    <=10{##?-*,##10-*}",
705 "QS Num-StressedSyl_before_C-Syl_in_C-Phrase
    <=11{##?-*,##10-*,##11-*}",
706 "QS Num-StressedSyl_before_C-Syl_in_C-Phrase
    <=12{##?-*,##10-*,##11-*,##12-*}",
707 ]
708 NUM_STRESSED_SYLS_AFTER_CURRENT_SYL_IN_CURRENT_PHRASE_EQ = [
709 "QS Num-StressedSyl_after_C-Syl_in_C-Phrase==x{* -x$*}",
710 "QS Num-StressedSyl_after_C-Syl_in_C-Phrase==1{* -1$*}",
711 "QS Num-StressedSyl_after_C-Syl_in_C-Phrase==2{* -2$*}",
712 "QS Num-StressedSyl_after_C-Syl_in_C-Phrase==3{* -3$*}",
713 "QS Num-StressedSyl_after_C-Syl_in_C-Phrase==4{* -4$*}",
714 "QS Num-StressedSyl_after_C-Syl_in_C-Phrase==5{* -5$*}",
715 "QS Num-StressedSyl_after_C-Syl_in_C-Phrase==6{* -6$*}",
716 "QS Num-StressedSyl_after_C-Syl_in_C-Phrase==7{* -7$*}",
717 "QS Num-StressedSyl_after_C-Syl_in_C-Phrase==8{* -8$*}",
718 "QS Num-StressedSyl_after_C-Syl_in_C-Phrase==9{* -9$*}",
719 "QS Num-StressedSyl_after_C-Syl_in_C-Phrase==10{* -10$*}",
720 "QS Num-StressedSyl_after_C-Syl_in_C-Phrase==11{* -11$*}",
721 "QS Num-StressedSyl_after_C-Syl_in_C-Phrase==12{* -12$*}",
722 ]
723 NUM_STRESSED_SYLS_AFTER_CURRENT_SYL_IN_CURRENT_PHRASE_LTE = [
724 "QS Num-StressedSyl_after_C-Syl_in_C-Phrase<=1{* -x$*,*-1$*}",
725 "QS Num-StressedSyl_after_C-Syl_in_C-Phrase<=2{* -x$*,*-1$*,*-2$
    *}",
726 "QS Num-StressedSyl_after_C-Syl_in_C-Phrase<=3{* -x$*,*-1$*,*-2$
    *,*-3$*}",
727 "QS Num-StressedSyl_after_C-Syl_in_C-Phrase<=4{* -x$*,*-1$*,*-2$
    *,*-3$*,*-4$*}",
728 "QS Num-StressedSyl_after_C-Syl_in_C-Phrase<=5{* -x$*,*-1$*,*-2$
    *,*-3$*,*-4$*,*-5$*}",

```

```

729 "QS Num-StressedSyl_after_C-Syl_in_C-Phrase<=6{*~x$*,~-1$*,~-2$
    *,~-3$*,~-4$*,~-5$*,~-6$*}",
730 "QS Num-StressedSyl_after_C-Syl_in_C-Phrase<=7{*~x$*,~-1$*,~-2$
    *,~-3$*,~-4$*,~-5$*,~-6$*,~-7$*}",
731 "QS Num-StressedSyl_after_C-Syl_in_C-Phrase<=8{*~x$*,~-1$*,~-2$
    *,~-3$*,~-4$*,~-5$*,~-6$*,~-7$*,~-8$*}",
732 "QS Num-StressedSyl_after_C-Syl_in_C-Phrase<=9{*~?*$*}",
733 "QS Num-StressedSyl_after_C-Syl_in_C-Phrase<=10{*~?*$*,~-10$*}",
734 "QS Num-StressedSyl_after_C-Syl_in_C-Phrase<=11{*~?*$*,~-10$
    *,~-11$*}",
735 "QS Num-StressedSyl_after_C-Syl_in_C-Phrase<=12{*~?*$*,~-10$
    *,~-11$*,~-12$*}",
736 ]
737 NUM_ACCENTED_SYLS_BEFORE_CURRENT_SYL_IN_CURRENT_PHRASE_EQ = [
738 "QS Num-AccentedSyl_before_C-Syl_in_C-Phrase==x{*~$x-*}",
739 "QS Num-AccentedSyl_before_C-Syl_in_C-Phrase==1{*~$1-*}",
740 "QS Num-AccentedSyl_before_C-Syl_in_C-Phrase==2{*~$2-*}",
741 "QS Num-AccentedSyl_before_C-Syl_in_C-Phrase==3{*~$3-*}",
742 "QS Num-AccentedSyl_before_C-Syl_in_C-Phrase==4{*~$4-*}",
743 "QS Num-AccentedSyl_before_C-Syl_in_C-Phrase==5{*~$5-*}",
744 "QS Num-AccentedSyl_before_C-Syl_in_C-Phrase==6{*~$6-*}",
745 ]
746 NUM_ACCENTED_SYLS_BEFORE_CURRENT_SYL_IN_CURRENT_PHRASE_LTE = [
747 "QS Num-AccentedSyl_before_C-Syl_in_C-Phrase<=1{*~$x-*,*~$1-*}",
748 "QS Num-AccentedSyl_before_C-Syl_in_C-Phrase<=2{*~$x-*,*~$1-*,*~$2
    -*}",
749 "QS Num-AccentedSyl_before_C-Syl_in_C-Phrase<=3{*~$x-*,*~$1-*,*~$2
    -*,*~$3-*}",
750 "QS Num-AccentedSyl_before_C-Syl_in_C-Phrase<=4{*~$x-*,*~$1-*,*~$2
    -*,*~$3-*,*~$4-*}",
751 "QS Num-AccentedSyl_before_C-Syl_in_C-Phrase<=5{*~$x-*,*~$1-*,*~$2
    -*,*~$3-*,*~$4-*,*~$5-*}",
752 "QS Num-AccentedSyl_before_C-Syl_in_C-Phrase<=6{*~$x-*,*~$1-*,*~$2
    -*,*~$3-*,*~$4-*,*~$5-*,*~$6-*}",
753 ]
754 NUM_ACCENTED_SYLS_AFTER_CURRENT_SYL_IN_CURRENT_PHRASE_EQ = [
755 "QS Num-AccentedSyl_after_C-Syl_in_C-Phrase==x{*~x!*}",
756 "QS Num-AccentedSyl_after_C-Syl_in_C-Phrase==1{*~1!*}",
757 "QS Num-AccentedSyl_after_C-Syl_in_C-Phrase==2{*~2!*}",
758 "QS Num-AccentedSyl_after_C-Syl_in_C-Phrase==3{*~3!*}",
759 "QS Num-AccentedSyl_after_C-Syl_in_C-Phrase==4{*~4!*}",
760 "QS Num-AccentedSyl_after_C-Syl_in_C-Phrase==5{*~5!*}",
761 "QS Num-AccentedSyl_after_C-Syl_in_C-Phrase==6{*~6!*}",
762 "QS Num-AccentedSyl_after_C-Syl_in_C-Phrase==7{*~7!*}",
763 ]
764 NUM_ACCENTED_SYLS_AFTER_CURRENT_SYL_IN_CURRENT_PHRASE_LTE = [
765 "QS Num-AccentedSyl_after_C-Syl_in_C-Phrase<=1{*~x!*,*~1!*}",

```

```

766 "QS Num-AccentedSyl_after_C-Syl_in_C-Phrase<=2{*~x
    !*,*-1!*,*-2!*},
767 "QS Num-AccentedSyl_after_C-Syl_in_C-Phrase<=3{*~x
    !*,*-1!*,*-2!*,*-3!*},
768 "QS Num-AccentedSyl_after_C-Syl_in_C-Phrase<=4{*~x
    !*,*-1!*,*-2!*,*-3!*,*-4!*},
769 "QS Num-AccentedSyl_after_C-Syl_in_C-Phrase<=5{*~x
    !*,*-1!*,*-2!*,*-3!*,*-4!*,*-5!*},
770 "QS Num-AccentedSyl_after_C-Syl_in_C-Phrase<=6{*~x
    !*,*-1!*,*-2!*,*-3!*,*-4!*,*-5!*,*-6!*},
771 "QS Num-AccentedSyl_after_C-Syl_in_C-Phrase<=7{*~x
    !*,*-1!*,*-2!*,*-3!*,*-4!*,*-5!*,*-6!*,*-7!*},
772 ]
773 NUM_SYLS_FROM_PREV_STRESSED_SYL_EQ = [
774 "QS Num-Syl_from_prev-StressedSyl==x{*!x-*},
775 "QS Num-Syl_from_prev-StressedSyl==0{*!0-*},
776 "QS Num-Syl_from_prev-StressedSyl==1{*!1-*},
777 "QS Num-Syl_from_prev-StressedSyl==2{*!2-*},
778 "QS Num-Syl_from_prev-StressedSyl==3{*!3-*},
779 "QS Num-Syl_from_prev-StressedSyl==4{*!4-*},
780 "QS Num-Syl_from_prev-StressedSyl==5{*!5-*},
781 ]
782 NUM_SYLS_FROM_PREV_STRESSED_SYL_LTE = [
783 "QS Num-Syl_from_prev-StressedSyl<=0{*!x-*,*!0-*},
784 "QS Num-Syl_from_prev-StressedSyl<=1{*!x-*,*!0-*,*!1-*},
785 "QS Num-Syl_from_prev-StressedSyl<=2{*!x-*,*!0-*,*!1-*,*!2-*},
786 "QS Num-Syl_from_prev-StressedSyl<=3{*!x
    -*,*!0-*,*!1-*,*!2-*,*!3-*},
787 "QS Num-Syl_from_prev-StressedSyl<=4{*!x
    -*,*!0-*,*!1-*,*!2-*,*!3-*,*!4-*},
788 "QS Num-Syl_from_prev-StressedSyl<=5{*!x
    -*,*!0-*,*!1-*,*!2-*,*!3-*,*!4-*,*!5-*},
789 ]
790 NUM_SYLS_FROM_NEXT_STRESSED_SYL_EQ = [
791 "QS Num-Syl_from_next-StressedSyl==x{*~x;*},
792 "QS Num-Syl_from_next-StressedSyl==0{*~0;*},
793 "QS Num-Syl_from_next-StressedSyl==1{*~1;*},
794 "QS Num-Syl_from_next-StressedSyl==2{*~2;*},
795 "QS Num-Syl_from_next-StressedSyl==3{*~3;*},
796 "QS Num-Syl_from_next-StressedSyl==4{*~4;*},
797 "QS Num-Syl_from_next-StressedSyl==5{*~5;*},
798 ]
799 NUM_SYLS_FROM_NEXT_STRESSED_SYL_LTE = [
800 "QS Num-Syl_from_next-StressedSyl<=0{*~x;*,*~0;*},
801 "QS Num-Syl_from_next-StressedSyl<=1{*~x;*,*~0;*,*~1;*},
802 "QS Num-Syl_from_next-StressedSyl<=2{*~x;*,*~0;*,*~1;*,*~2;*},

```

```

803     "QS Num-Syl_from_next-StressedSyl<=3{* -x
      ;*,*-0;*,*-1;*,*-2;*,*-3;*}" ,
804     "QS Num-Syl_from_next-StressedSyl<=4{* -x
      ;*,*-0;*,*-1;*,*-2;*,*-3;*,*-4;*}" ,
805     "QS Num-Syl_from_next-StressedSyl<=5{* -x
      ;*,*-0;*,*-1;*,*-2;*,*-3;*,*-4;*,*-5;*}" ,
806 ]
807 NUM_SYLS_FROM_PREV_ACCENTED_SYL_EQ = [
808     "QS Num-Syl_from_prev-AccentedSyl==x{* ;x-*}" ,
809     "QS Num-Syl_from_prev-AccentedSyl==0{* ;0-*}" ,
810     "QS Num-Syl_from_prev-AccentedSyl==1{* ;1-*}" ,
811     "QS Num-Syl_from_prev-AccentedSyl==2{* ;2-*}" ,
812     "QS Num-Syl_from_prev-AccentedSyl==3{* ;3-*}" ,
813     "QS Num-Syl_from_prev-AccentedSyl==4{* ;4-*}" ,
814     "QS Num-Syl_from_prev-AccentedSyl==5{* ;5-*}" ,
815     "QS Num-Syl_from_prev-AccentedSyl==6{* ;6-*}" ,
816     "QS Num-Syl_from_prev-AccentedSyl==7{* ;7-*}" ,
817     "QS Num-Syl_from_prev-AccentedSyl==8{* ;8-*}" ,
818     "QS Num-Syl_from_prev-AccentedSyl==9{* ;9-*}" ,
819     "QS Num-Syl_from_prev-AccentedSyl==10{* ;10-*}" ,
820     "QS Num-Syl_from_prev-AccentedSyl==11{* ;11-*}" ,
821     "QS Num-Syl_from_prev-AccentedSyl==12{* ;12-*}" ,
822     "QS Num-Syl_from_prev-AccentedSyl==13{* ;13-*}" ,
823     "QS Num-Syl_from_prev-AccentedSyl==14{* ;14-*}" ,
824     "QS Num-Syl_from_prev-AccentedSyl==15{* ;15-*}" ,
825     "QS Num-Syl_from_prev-AccentedSyl==16{* ;16-*}" ,
826 ]
827 NUM_SYLS_FROM_PREV_ACCENTED_SYL_LTE = [
828     "QS Num-Syl_from_prev-AccentedSyl<=0{* ;x-*,*;0-*}" ,
829     "QS Num-Syl_from_prev-AccentedSyl<=1{* ;x-*,*;0-*,*;1-*}" ,
830     "QS Num-Syl_from_prev-AccentedSyl<=2{* ;x-*,*;0-*,*;1-*,*;2-*}" ,
831     "QS Num-Syl_from_prev-AccentedSyl<=3{* ;x
      -*,*;0-*,*;1-*,*;2-*,*;3-*}" ,
832     "QS Num-Syl_from_prev-AccentedSyl<=4{* ;x
      -*,*;0-*,*;1-*,*;2-*,*;3-*,*;4-*}" ,
833     "QS Num-Syl_from_prev-AccentedSyl<=5{* ;x
      -*,*;0-*,*;1-*,*;2-*,*;3-*,*;4-*,*;5-*}" ,
834     "QS Num-Syl_from_prev-AccentedSyl<=6{* ;x
      -*,*;0-*,*;1-*,*;2-*,*;3-*,*;4-*,*;5-*,*;6-*}" ,
835     "QS Num-Syl_from_prev-AccentedSyl<=7{* ;x
      -*,*;0-*,*;1-*,*;2-*,*;3-*,*;4-*,*;5-*,*;6-*,*;7-*}" ,
836     "QS Num-Syl_from_prev-AccentedSyl<=8{* ;x
      -*,*;0-*,*;1-*,*;2-*,*;3-*,*;4-*,*;5-*,*;6-*,*;7-*,*;8-*}" ,
837     "QS Num-Syl_from_prev-AccentedSyl<=9{* ;?-*}" ,
838     "QS Num-Syl_from_prev-AccentedSyl<=10{* ;?-*,*;10-*}" ,
839     "QS Num-Syl_from_prev-AccentedSyl<=11{* ;?-*,*;10-*,*;11-*}" ,

```

```

840     "QS Num-Syl_from_prev-AccentedSyl
      <=12{*;?-*,*;10-*,*;11-*,*;12-*}",
841     "QS Num-Syl_from_prev-AccentedSyl
      <=13{*;?-*,*;10-*,*;11-*,*;12-*,*;13-*}",
842     "QS Num-Syl_from_prev-AccentedSyl
      <=14{*;?-*,*;10-*,*;11-*,*;12-*,*;13-*,*;14-*}",
843     "QS Num-Syl_from_prev-AccentedSyl
      <=15{*;?-*,*;10-*,*;11-*,*;12-*,*;13-*,*;14-*,*;15-*}",
844     "QS Num-Syl_from_prev-AccentedSyl
      <=16{*;?-*,*;10-*,*;11-*,*;12-*,*;13-*,*;14-*,*;15-*,*;16-*}",

845 ]
846 NUM_SYLS_FROM_NEXT_ACCENTED_SYL_EQ = [
847     "QS Num-Syl_from_next-AccentedSyl==x{* -x|*}",
848     "QS Num-Syl_from_next-AccentedSyl==0{* -0|*}",
849     "QS Num-Syl_from_next-AccentedSyl==1{* -1|*}",
850     "QS Num-Syl_from_next-AccentedSyl==2{* -2|*}",
851     "QS Num-Syl_from_next-AccentedSyl==3{* -3|*}",
852     "QS Num-Syl_from_next-AccentedSyl==4{* -4|*}",
853     "QS Num-Syl_from_next-AccentedSyl==5{* -5|*}",
854     "QS Num-Syl_from_next-AccentedSyl==6{* -6|*}",
855     "QS Num-Syl_from_next-AccentedSyl==7{* -7|*}",
856     "QS Num-Syl_from_next-AccentedSyl==8{* -8|*}",
857     "QS Num-Syl_from_next-AccentedSyl==9{* -9|*}",
858     "QS Num-Syl_from_next-AccentedSyl==10{* -10|*}",
859     "QS Num-Syl_from_next-AccentedSyl==11{* -11|*}",
860     "QS Num-Syl_from_next-AccentedSyl==12{* -12|*}",
861     "QS Num-Syl_from_next-AccentedSyl==13{* -13|*}",
862     "QS Num-Syl_from_next-AccentedSyl==14{* -14|*}",
863     "QS Num-Syl_from_next-AccentedSyl==15{* -15|*}",
864     "QS Num-Syl_from_next-AccentedSyl==16{* -16|*}",
865 ]
866 NUM_SYLS_FROM_NEXT_ACCENTED_SYL_LTE = [
867     "QS Num-Syl_from_next-AccentedSyl<=0{* -x|*,*-0|*}",
868     "QS Num-Syl_from_next-AccentedSyl<=1{* -x|*,*-0|*,*-1|*}",
869     "QS Num-Syl_from_next-AccentedSyl<=2{* -x|*,*-0|*,*-1|*,*-2|*}",
870     "QS Num-Syl_from_next-AccentedSyl<=3{* -x
      |*,*-0|*,*-1|*,*-2|*,*-3|*}",
871     "QS Num-Syl_from_next-AccentedSyl<=4{* -x
      |*,*-0|*,*-1|*,*-2|*,*-3|*,*-4|*}",
872     "QS Num-Syl_from_next-AccentedSyl<=5{* -x
      |*,*-0|*,*-1|*,*-2|*,*-3|*,*-4|*,*-5|*}",
873     "QS Num-Syl_from_next-AccentedSyl<=6{* -x
      |*,*-0|*,*-1|*,*-2|*,*-3|*,*-4|*,*-5|*,*-6|*}",
874     "QS Num-Syl_from_next-AccentedSyl<=7{* -x
      |*,*-0|*,*-1|*,*-2|*,*-3|*,*-4|*,*-5|*,*-6|*,*-7|*}",

```



```

875 "QS Num-Syl_from_next-AccentedSyl<=8{*-x
      |*,*-0|*,*-1|*,*-2|*,*-3|*,*-4|*,*-5|*,*-6|*,*-7|*,*-8|*}",
876 "QS Num-Syl_from_next-AccentedSyl<=9{*-?|*}",
877 "QS Num-Syl_from_next-AccentedSyl<=10{*-?|*,*-10|*}",
878 "QS Num-Syl_from_next-AccentedSyl<=11{*-?|*,*-10|*,*-11|*}",
879 "QS Num-Syl_from_next-AccentedSyl
      <=12{*-?|*,*-10|*,*-11|*,*-12|*}",
880 "QS Num-Syl_from_next-AccentedSyl
      <=13{*-?|*,*-10|*,*-11|*,*-12|*,*-13|*}",
881 "QS Num-Syl_from_next-AccentedSyl
      <=14{*-?|*,*-10|*,*-11|*,*-12|*,*-13|*,*-14|*}",
882 "QS Num-Syl_from_next-AccentedSyl
      <=15{*-?|*,*-10|*,*-11|*,*-12|*,*-13|*,*-14|*,*-15|*}",
883 "QS Num-Syl_from_next-AccentedSyl
      <=16{*-?|*,*-10|*,*-11|*,*-12|*,*-13|*,*-14|*,*-15|*,*-16|*}",

884 ]
885
886 SYL_VOWEL_TYPE = [
887     "QS C-Syl_Vowel==x{*|x/C:*}",
888     "QS C-Syl_Vowel{*|i/C:*,*|y/C:*,*|e/C:*,*|a/C:*,*|o/C:*,*|u/C
      :*,*|schwa/C:*}",
889     "QS C-Syl_Front_Vowel{*|e/C:*,*|i/C:*,*|y/C:*}",
890     "QS C-Syl_Central_Vowel{*|a/C:*,*|schwa/C:*}",
891     "QS C-Syl_Back_Vowel{*|o/C:*,*|u/C:*}",
892     "QS C-Syl_High_Vowel{*|i/C:*,*|y/C:*,*|u/C:*}",
893     "QS C-Syl_Medium_Vowel{*|e/C:*,*|o/C:*,*|schwa/C:*}",
894     "QS C-Syl_Low_Vowel{*|a/C:*}",
895     "QS C-Syl_Rounded_Vowel{*|o/C:*,*|u/C:*}",
896     "QS C-Syl_Unrounded_Vowel{*|a/C:*,*|e/C:*,*|i/C:*,*|y/C:*}",
897     "QS C-Syl_IVowel{*|i/C:*}",
898     "QS C-Syl_EVowel{*|e/C:*}",
899     "QS C-Syl_AVowel{*|a/C:*}",
900     "QS C-Syl_OVowel{*|o/C:*}",
901     "QS C-Syl_UVowel{*|u/C:*}",
902     "QS C-Syl_YVowel{*|y/C:*}",
903     "QS C-Syl_SCHWAVowel                                {*|schwa/C:*}",
904 ]
905
906 SYL_STRESSED_R = [
907     "QS R-Syl_Stress==1{*/C:1+*}",
908     "QS R-Syl_Stress==0{*/C:0+*}",
909 ]
910 SYL_ACCENTED_R = [
911     "QS R-Syl_Accent==1{**+1+*}",
912     "QS R-Syl_Accent==0{**+0+*}",
913 ]

```

```

914 SYL_ACCENT_TYPE_R = [
915     "QS R-Syl_TOBI_Accent-H*{/M:H?}" ,
916     "QS R-Syl_TOBI_Accent-L*{/M:L?}" ,
917     "QS R-Syl_TOBI_Accent-L*+H{/M:L?+H}" ,
918     "QS R-Syl_TOBI_Accent-L+H*{/M:L+H?}" ,
919     "QS R-Syl_TOBI_Accent-O*{/M:O}" ,
920     "QS R-Syl_TOBI_Accent-NONE*{/M:NONE}" ,
921     "QS R-Syl_TOBI_Accent-x*{/M:x}" ,
922 ]
923
924 SYL_LEN_IN_NO_SEG_EQ_R = [
925     "QS R-Syl_Num-Segs==0{**0/D:*}" ,
926     "QS R-Syl_Num-Segs==1{**1/D:*}" ,
927     "QS R-Syl_Num-Segs==2{**2/D:*}" ,
928     "QS R-Syl_Num-Segs==3{**3/D:*}" ,
929     "QS R-Syl_Num-Segs==4{**4/D:*}" ,
930     "QS R-Syl_Num-Segs==5{**5/D:*}" ,
931     "QS R-Syl_Num-Segs==6{**6/D:*}" ,
932     "QS R-Syl_Num-Segs==7{**7/D:*}" ,
933 ]
934 SYL_LEN_IN_NO_SEG_LTE_R = [
935     "QS R-Syl_Num-Segs<=1{**0/D:*,**1/D:*}" ,
936     "QS R-Syl_Num-Segs<=2{**0/D:*,**1/D:*,**2/D:*}" ,
937     "QS R-Syl_Num-Segs<=3{**0/D:*,**1/D:*,**2/D:*,**3/D:*}" ,
938     "QS R-Syl_Num-Segs<=4{**0/D:*,**1/D:*,**2/D:*,**3/D:*,**4/D
939     :*}" ,
940     "QS R-Syl_Num-Segs<=5{**0/D:*,**1/D:*,**2/D:*,**3/D:*,**4/D
941     :*,**5/D:*}" ,
942     "QS R-Syl_Num-Segs<=6{**0/D:*,**1/D:*,**2/D:*,**3/D:*,**4/D
943     :*,**5/D:*,**6/D:*}" ,
944     "QS R-Syl_Num-Segs<=7{**0/D:*,**1/D:*,**2/D:*,**3/D:*,**4/D
945     :*,**5/D:*,**6/D:*,**7/D:*}" ,
946 ]
947
948 WORD_LEN_IN_NO_SYLS_EQ_L = [
949     "QS L-Word_Num-Syls==0{*_0/E:*}" ,
950     "QS L-Word_Num-Syls==1{*_1/E:*}" ,
951     "QS L-Word_Num-Syls==2{*_2/E:*}" ,
952     "QS L-Word_Num-Syls==3{*_3/E:*}" ,
953     "QS L-Word_Num-Syls==4{*_4/E:*}" ,
954     "QS L-Word_Num-Syls==5{*_5/E:*}" ,
955     "QS L-Word_Num-Syls==6{*_6/E:*}" ,
956     "QS L-Word_Num-Syls==7{*_7/E:*}" ,
957 ]
958 WORD_LEN_IN_NO_SYLS_LTE_L = [
959     "QS L-Word_Num-Syls<=1{*_0/E:*,*_1/E:*}" ,
960     "QS L-Word_Num-Syls<=2{*_0/E:*,*_1/E:*,*_2/E:*}" ,

```

```

957 "QS L-Word_Num-Syls<=3{*_0/E:*,*_1/E:*,*_2/E:*,*_3/E:*}",
958 "QS L-Word_Num-Syls<=4{*_0/E:*,*_1/E:*,*_2/E:*,*_3/E:*,*_4/E
    :*}",
959 "QS L-Word_Num-Syls<=5{*_0/E:*,*_1/E:*,*_2/E:*,*_3/E:*,*_4/E
    :*,*_5/E:*}",
960 "QS L-Word_Num-Syls<=6{*_0/E:*,*_1/E:*,*_2/E:*,*_3/E:*,*_4/E
    :*,*_5/E:*,*_6/E:*}",
961 "QS L-Word_Num-Syls<=7{*_0/E:*,*_1/E:*,*_2/E:*,*_3/E:*,*_4/E
    :*,*_5/E:*,*_6/E:*,*_7/E:*}",
962 ]
963 WORD_LEN_IN_NO_SYLS_EQ_C = [
964     "QS C-Word_Num-Syls==x{**x@*}",
965     "QS C-Word_Num-Syls==1{**1@*}",
966     "QS C-Word_Num-Syls==2{**2@*}",
967     "QS C-Word_Num-Syls==3{**3@*}",
968     "QS C-Word_Num-Syls==4{**4@*}",
969     "QS C-Word_Num-Syls==5{**5@*}",
970     "QS C-Word_Num-Syls==6{**6@*}",
971     "QS C-Word_Num-Syls==7{**7@*}",
972 ]
973 WORD_LEN_IN_NO_SYLS_LTE_C = [
974     "QS C-Word_Num-Syls<=1{**x@*,**1@*}",
975     "QS C-Word_Num-Syls<=2{**x@*,**1@*,**2@*}",
976     "QS C-Word_Num-Syls<=3{**x@*,**1@*,**2@*,**3@*}",
977     "QS C-Word_Num-Syls<=4{**x@*,**1@*,**2@*,**3@*,**4@*}",
978     "QS C-Word_Num-Syls<=5{**x@*,**1@*,**2@*,**3@*,**4@*,**5@*}",
979     "QS C-Word_Num-Syls<=6{**x@*,**1@*,**2@*,**3@*,**4@*,**5@*,**6@
        *}",
980     "QS C-Word_Num-Syls<=7{**x@*,**1@*,**2@*,**3@*,**4@*,**5@*,**6@
        *,**7@*}",
981 ]
982
983 WORD_POSITION_IN_PHRASE_FW_EQ = [
984     "QS Pos_C-Word_in_C-Phrase(Fw)==x{*@x+*}",
985     "QS Pos_C-Word_in_C-Phrase(Fw)==1{*@1+*}",
986     "QS Pos_C-Word_in_C-Phrase(Fw)==2{*@2+*}",
987     "QS Pos_C-Word_in_C-Phrase(Fw)==3{*@3+*}",
988     "QS Pos_C-Word_in_C-Phrase(Fw)==4{*@4+*}",
989     "QS Pos_C-Word_in_C-Phrase(Fw)==5{*@5+*}",
990     "QS Pos_C-Word_in_C-Phrase(Fw)==6{*@6+*}",
991     "QS Pos_C-Word_in_C-Phrase(Fw)==7{*@7+*}",
992     "QS Pos_C-Word_in_C-Phrase(Fw)==8{*@8+*}",
993     "QS Pos_C-Word_in_C-Phrase(Fw)==9{*@9+*}",
994     "QS Pos_C-Word_in_C-Phrase(Fw)==10{*@10+*}",
995     "QS Pos_C-Word_in_C-Phrase(Fw)==11{*@11+*}",
996     "QS Pos_C-Word_in_C-Phrase(Fw)==12{*@12+*}",
997     "QS Pos_C-Word_in_C-Phrase(Fw)==13{*@13+*}",

```

```

998 ]
999 WORD_POSITION_IN_PHRASE_FW_LTE = [
1000     "QS Pos_C-Word_in_C-Phrase(Fw) <=1{*@x+*,*@1+*}",
1001     "QS Pos_C-Word_in_C-Phrase(Fw) <=2{*@x+*,*@1+*,*@2+*}",
1002     "QS Pos_C-Word_in_C-Phrase(Fw) <=3{*@x+*,*@1+*,*@2+*,*@3+*}",
1003     "QS Pos_C-Word_in_C-Phrase(Fw) <=4{*@x+*,*@1+*,*@2+*,*@3+*,*@4
        +*}",
1004     "QS Pos_C-Word_in_C-Phrase(Fw) <=5{*@x+*,*@1+*,*@2+*,*@3+*,*@4
        +*,*@5+*}",
1005     "QS Pos_C-Word_in_C-Phrase(Fw) <=6{*@x+*,*@1+*,*@2+*,*@3+*,*@4
        +*,*@5+*,*@6+*}",
1006     "QS Pos_C-Word_in_C-Phrase(Fw) <=7{*@x+*,*@1+*,*@2+*,*@3+*,*@4
        +*,*@5+*,*@6+*,*@7+*}",
1007     "QS Pos_C-Word_in_C-Phrase(Fw) <=8{*@x+*,*@1+*,*@2+*,*@3+*,*@4
        +*,*@5+*,*@6+*,*@7+*,*@8+*}",
1008     "QS Pos_C-Word_in_C-Phrase(Fw) <=9{*@?+*}",
1009     "QS Pos_C-Word_in_C-Phrase(Fw) <=10{*@?+*,*@10+*}",
1010     "QS Pos_C-Word_in_C-Phrase(Fw) <=11{*@?+*,*@10+*,*@11+*}",
1011     "QS Pos_C-Word_in_C-Phrase(Fw) <=12{*@?+*,*@10+*,*@11+*,*@12
        +*}",
1012     "QS Pos_C-Word_in_C-Phrase(Fw) <=13{*@?+*,*@10+*,*@11+*,*@12+*,*
        @13+*}",
1013 ]
1014 WORD_POSITION_IN_PHRASE_BW_EQ = [
1015     "QS Pos_C-Word_in_C-Phrase(Bw)==x{*+x&*}",
1016     "QS Pos_C-Word_in_C-Phrase(Bw)==0{*+0&*}",
1017     "QS Pos_C-Word_in_C-Phrase(Bw)==1{*+1&*}",
1018     "QS Pos_C-Word_in_C-Phrase(Bw)==2{*+2&*}",
1019     "QS Pos_C-Word_in_C-Phrase(Bw)==3{*+3&*}",
1020     "QS Pos_C-Word_in_C-Phrase(Bw)==4{*+4&*}",
1021     "QS Pos_C-Word_in_C-Phrase(Bw)==5{*+5&*}",
1022     "QS Pos_C-Word_in_C-Phrase(Bw)==6{*+6&*}",
1023     "QS Pos_C-Word_in_C-Phrase(Bw)==7{*+7&*}",
1024     "QS Pos_C-Word_in_C-Phrase(Bw)==8{*+8&*}",
1025     "QS Pos_C-Word_in_C-Phrase(Bw)==9{*+9&*}",
1026     "QS Pos_C-Word_in_C-Phrase(Bw)==10{*+10&*}",
1027     "QS Pos_C-Word_in_C-Phrase(Bw)==11{*+11&*}",
1028     "QS Pos_C-Word_in_C-Phrase(Bw)==12{*+12&*}",
1029     "QS Pos_C-Word_in_C-Phrase(Bw)==13{*+13&*}",
1030 ]
1031 WORD_POSITION_IN_PHRASE_BW_LTE = [
1032     "QS Pos_C-Word_in_C-Phrase(Bw) <=1{*+x&*,*+0&*,*+1&*}",
1033     "QS Pos_C-Word_in_C-Phrase(Bw) <=2{*+x&*,*+0&*,*+1&*,*+2&*}",
1034     "QS Pos_C-Word_in_C-Phrase(Bw) <=3{*+x
        &*,*+0&*,*+1&*,*+2&*,*+3&*}",
1035     "QS Pos_C-Word_in_C-Phrase(Bw) <=4{*+x
        &*,*+0&*,*+1&*,*+2&*,*+3&*,*+4&*}",

```

```

1036 "QS Pos_C-Word_in_C-Phrase(Bw) <=5{**x
      &*,**0&*,**1&*,**2&*,**3&*,**4&*,**5&}",
1037 "QS Pos_C-Word_in_C-Phrase(Bw) <=6{**x
      &*,**0&*,**1&*,**2&*,**3&*,**4&*,**5&*,**6&}",
1038 "QS Pos_C-Word_in_C-Phrase(Bw) <=7{**x
      &*,**0&*,**1&*,**2&*,**3&*,**4&*,**5&*,**6&*,**7&}",
1039 "QS Pos_C-Word_in_C-Phrase(Bw) <=8{**x
      &*,**0&*,**1&*,**2&*,**3&*,**4&*,**5&*,**6&*,**7&*,**8&}",
1040 "QS Pos_C-Word_in_C-Phrase(Bw) <=9{**?&}",
1041 "QS Pos_C-Word_in_C-Phrase(Bw) <=10{**?&*,**10&}",
1042 "QS Pos_C-Word_in_C-Phrase(Bw) <=11{**?&*,**10&*,**11&}",
1043 "QS Pos_C-Word_in_C-Phrase(Bw)
      <=12{**?&*,**10&*,**11&*,**12&}",
1044 "QS Pos_C-Word_in_C-Phrase(Bw)
      <=13{**?&*,**10&*,**11&*,**12&*,**13&}",
1045 ]
1046
1047 NUM_CONT_WORDS_BEFORE_CURRENT_WORD_IN_PHRASE_EQ = [
1048 "QS Num-ContWord_before_C-Word_in_C-Phrase==x{*&x+*}",
1049 "QS Num-ContWord_before_C-Word_in_C-Phrase==1{*&1+*}",
1050 "QS Num-ContWord_before_C-Word_in_C-Phrase==2{*&2+*}",
1051 "QS Num-ContWord_before_C-Word_in_C-Phrase==3{*&3+*}",
1052 "QS Num-ContWord_before_C-Word_in_C-Phrase==4{*&4+*}",
1053 "QS Num-ContWord_before_C-Word_in_C-Phrase==5{*&5+*}",
1054 "QS Num-ContWord_before_C-Word_in_C-Phrase==6{*&6+*}",
1055 "QS Num-ContWord_before_C-Word_in_C-Phrase==7{*&7+*}",
1056 "QS Num-ContWord_before_C-Word_in_C-Phrase==8{*&8+*}",
1057 "QS Num-ContWord_before_C-Word_in_C-Phrase==9{*&9+*}",
1058 ]
1059 NUM_CONT_WORDS_BEFORE_CURRENT_WORD_IN_PHRASE_LTE = [
1060 "QS Num-ContWord_before_C-Word_in_C-Phrase<=1{*&x+*,*&1+*}",
1061 "QS Num-ContWord_before_C-Word_in_C-Phrase<=2{*&x
      +*,*&1+*,*&2+*}",
1062 "QS Num-ContWord_before_C-Word_in_C-Phrase<=3{*&x
      +*,*&1+*,*&2+*,*&3+*}",
1063 "QS Num-ContWord_before_C-Word_in_C-Phrase<=4{*&x
      +*,*&1+*,*&2+*,*&3+*,*&4+*}",
1064 "QS Num-ContWord_before_C-Word_in_C-Phrase<=5{*&x
      +*,*&1+*,*&2+*,*&3+*,*&4+*,*&5+*}",
1065 "QS Num-ContWord_before_C-Word_in_C-Phrase<=6{*&x
      +*,*&1+*,*&2+*,*&3+*,*&4+*,*&5+*,*&6+*}",
1066 "QS Num-ContWord_before_C-Word_in_C-Phrase<=7{*&x
      +*,*&1+*,*&2+*,*&3+*,*&4+*,*&5+*,*&6+*,*&7+*}",
1067 "QS Num-ContWord_before_C-Word_in_C-Phrase<=8{*&x
      +*,*&1+*,*&2+*,*&3+*,*&4+*,*&5+*,*&6+*,*&7+*,*&8+*}",
1068 "QS Num-ContWord_before_C-Word_in_C-Phrase<=9{*&?+*}",
1069 ]

```

```

1070 NUM_CONT_WORDS_AFTER_CURRENT_WORD_IN_PHRASE_EQ = [
1071     "QS Num-ContWord_after_C-Word_in_C-Phrase==x{**x**}",
1072     "QS Num-ContWord_after_C-Word_in_C-Phrase==0{**0**}",
1073     "QS Num-ContWord_after_C-Word_in_C-Phrase==1{**1**}",
1074     "QS Num-ContWord_after_C-Word_in_C-Phrase==2{**2**}",
1075     "QS Num-ContWord_after_C-Word_in_C-Phrase==3{**3**}",
1076     "QS Num-ContWord_after_C-Word_in_C-Phrase==4{**4**}",
1077     "QS Num-ContWord_after_C-Word_in_C-Phrase==5{**5**}",
1078     "QS Num-ContWord_after_C-Word_in_C-Phrase==6{**6**}",
1079     "QS Num-ContWord_after_C-Word_in_C-Phrase==7{**7**}",
1080     "QS Num-ContWord_after_C-Word_in_C-Phrase==8{**8**}",
1081 ]
1082 NUM_CONT_WORDS_AFTER_CURRENT_WORD_IN_PHRASE_LTE = [
1083     "QS Num-ContWord_after_C-Word_in_C-Phrase<=0{**x**,**0**}",
1084     "QS Num-ContWord_after_C-Word_in_C-Phrase<=1{**x
1085         **,**0**,**1**}",
1086     "QS Num-ContWord_after_C-Word_in_C-Phrase<=2{**x
1087         **,**0**,**1**,**2**}",
1088     "QS Num-ContWord_after_C-Word_in_C-Phrase<=3{**x
1089         **,**0**,**1**,**2**,**3**}",
1090     "QS Num-ContWord_after_C-Word_in_C-Phrase<=4{**x
1091         **,**0**,**1**,**2**,**3**,**4**}",
1092     "QS Num-ContWord_after_C-Word_in_C-Phrase<=5{**x
1093         **,**0**,**1**,**2**,**3**,**4**,**5**}",
1094     "QS Num-ContWord_after_C-Word_in_C-Phrase<=6{**x
1095         **,**0**,**1**,**2**,**3**,**4**,**5**,**6**}",
1096     "QS Num-ContWord_after_C-Word_in_C-Phrase<=7{**x
1097         **,**0**,**1**,**2**,**3**,**4**,**5**,**6**,**7**}",
1098     "QS Num-ContWord_after_C-Word_in_C-Phrase<=8{**x
1099         **,**0**,**1**,**2**,**3**,**4**,**5**,**6**,**7**,**8**}",
1100 ]
1101 NUM_WORDS_FROM_PREV_CONT_WORD_EQ = [
1102     "QS Num-Words_from_prev-ContWord==x{**x**}",
1103     "QS Num-Words_from_prev-ContWord==0{**0**}",
1104     "QS Num-Words_from_prev-ContWord==1{**1**}",
1105     "QS Num-Words_from_prev-ContWord==2{**2**}",
1106     "QS Num-Words_from_prev-ContWord==3{**3**}",
1107     "QS Num-Words_from_prev-ContWord==4{**4**}",
1108     "QS Num-Words_from_prev-ContWord==5{**5**}",
1109 ]
1110 NUM_WORDS_FROM_PREV_CONT_WORD_LTE = [
1111     "QS Num-Words_from_prev-ContWord<=0{**x**,**0**}",
1112     "QS Num-Words_from_prev-ContWord<=1{**x**,**0**,**1**}",
1113     "QS Num-Words_from_prev-ContWord<=2{**x**,**0**,**1**,**2**}",
1114     "QS Num-Words_from_prev-ContWord<=3{**x
1115         **,**0**,**1**,**2**,**3**}",

```

```

1107 "QS Num-Words_from_prev-ContWord<=4{##x
      ++,##0++,##1++,##2++,##3++,##4++}",
1108 "QS Num-Words_from_prev-ContWord<=5{##x
      ++,##0++,##1++,##2++,##3++,##4++,##5++}",
1109 ]
1110 NUM_WORDS_FROM_NEXT_CONT_WORD_EQ = [
1111 "QS Num-Words_from_next-ContWord==x{++x/F:*}",
1112 "QS Num-Words_from_next-ContWord==0{++0/F:*}",
1113 "QS Num-Words_from_next-ContWord==1{++1/F:*}",
1114 "QS Num-Words_from_next-ContWord==2{++2/F:*}",
1115 "QS Num-Words_from_next-ContWord==3{++3/F:*}",
1116 "QS Num-Words_from_next-ContWord==4{++4/F:*}",
1117 "QS Num-Words_from_next-ContWord==5{++5/F:*}",
1118 ]
1119 NUM_WORDS_FROM_NEXT_CONT_WORD_LTE = [
1120 "QS Num-Words_from_next-ContWord<=0{++x/F:*,++0/F:*}",
1121 "QS Num-Words_from_next-ContWord<=1{++x/F:*,++0/F:*,++1/F:*}",
1122 "QS Num-Words_from_next-ContWord<=2{++x/F:*,++0/F:*,++1/F
      :*,++2/F:*}",
1123 "QS Num-Words_from_next-ContWord<=3{++x/F:*,++0/F:*,++1/F
      :*,++2/F:*,++3/F:*}",
1124 "QS Num-Words_from_next-ContWord<=4{++x/F:*,++0/F:*,++1/F
      :*,++2/F:*,++3/F:*,++4/F:*}",
1125 "QS Num-Words_from_next-ContWord<=5{++x/F:*,++0/F:*,++1/F
      :*,++2/F:*,++3/F:*,++4/F:*,++5/F:*}",
1126 ]
1127
1128 WORD_LEN_IN_NO_SYLS_EQ_R = [
1129 "QS R-Word_Num-Syls==0{*_0/G:*}",
1130 "QS R-Word_Num-Syls==1{*_1/G:*}",
1131 "QS R-Word_Num-Syls==2{*_2/G:*}",
1132 "QS R-Word_Num-Syls==3{*_3/G:*}",
1133 "QS R-Word_Num-Syls==4{*_4/G:*}",
1134 "QS R-Word_Num-Syls==5{*_5/G:*}",
1135 "QS R-Word_Num-Syls==6{*_6/G:*}",
1136 "QS R-Word_Num-Syls==7{*_7/G:*}",
1137 ]
1138 WORD_LEN_IN_NO_SYLS_LTE_R = [
1139 "QS R-Word_Num-Syls<=1{*_0/G:*,*_1/G:*}",
1140 "QS R-Word_Num-Syls<=2{*_0/G:*,*_1/G:*,*_2/G:*}",
1141 "QS R-Word_Num-Syls<=3{*_0/G:*,*_1/G:*,*_2/G:*,*_3/G:*}",
1142 "QS R-Word_Num-Syls<=4{*_0/G:*,*_1/G:*,*_2/G:*,*_3/G:*,*_4/G
      :*}",
1143 "QS R-Word_Num-Syls<=5{*_0/G:*,*_1/G:*,*_2/G:*,*_3/G:*,*_4/G
      :*,*_5/G:*}",
1144 "QS R-Word_Num-Syls<=6{*_0/G:*,*_1/G:*,*_2/G:*,*_3/G:*,*_4/G
      :*,*_5/G:*,*_6/G:*}",

```

```

1145 "QS R-Word_Num-Syls<=7{*_0/G:*,*_1/G:*,*_2/G:*,*_3/G:*,*_4/G
1146 :*,*_5/G:*,*_6/G:*,*_7/G:*)"",
1147 ]
1148 PHRASE_LEN_IN_NO_SYLS_EQ_L = [
1149 "QS L-Phrase_Num-Syls==0{*/G:0_*}",
1150 "QS L-Phrase_Num-Syls==1{*/G:1_*}",
1151 "QS L-Phrase_Num-Syls==2{*/G:2_*}",
1152 "QS L-Phrase_Num-Syls==3{*/G:3_*}",
1153 "QS L-Phrase_Num-Syls==4{*/G:4_*}",
1154 "QS L-Phrase_Num-Syls==5{*/G:5_*}",
1155 "QS L-Phrase_Num-Syls==6{*/G:6_*}",
1156 "QS L-Phrase_Num-Syls==7{*/G:7_*}",
1157 "QS L-Phrase_Num-Syls==8{*/G:8_*}",
1158 "QS L-Phrase_Num-Syls==9{*/G:9_*}",
1159 "QS L-Phrase_Num-Syls==10{*/G:10_*}",
1160 "QS L-Phrase_Num-Syls==11{*/G:11_*}",
1161 "QS L-Phrase_Num-Syls==12{*/G:12_*}",
1162 "QS L-Phrase_Num-Syls==13{*/G:13_*}",
1163 "QS L-Phrase_Num-Syls==14{*/G:14_*}",
1164 "QS L-Phrase_Num-Syls==15{*/G:15_*}",
1165 "QS L-Phrase_Num-Syls==16{*/G:16_*}",
1166 "QS L-Phrase_Num-Syls==17{*/G:17_*}",
1167 "QS L-Phrase_Num-Syls==18{*/G:18_*}",
1168 "QS L-Phrase_Num-Syls==19{*/G:19_*}",
1169 "QS L-Phrase_Num-Syls==20{*/G:20_*}",
1170 ]
1171 PHRASE_LEN_IN_NO_SYLS_LTE_L = [
1172 "QS L-Phrase_Num-Syls<=1{*/G:0_*,*/G:1_*}",
1173 "QS L-Phrase_Num-Syls<=2{*/G:0_*,*/G:1_*,*/G:2_*}",
1174 "QS L-Phrase_Num-Syls<=3{*/G:0_*,*/G:1_*,*/G:2_*,*/G:3_*}",
1175 "QS L-Phrase_Num-Syls<=4{*/G:0_*,*/G:1_*,*/G:2_*,*/G:3_*,*/G:4_
1176 *}",
1177 "QS L-Phrase_Num-Syls<=5{*/G:0_*,*/G:1_*,*/G:2_*,*/G:3_*,*/G:4_
1178 *,*/G:5_*}",
1179 "QS L-Phrase_Num-Syls<=6{*/G:0_*,*/G:1_*,*/G:2_*,*/G:3_*,*/G:4_
1180 *,*/G:5_*,*/G:6_*}",
1181 "QS L-Phrase_Num-Syls<=7{*/G:0_*,*/G:1_*,*/G:2_*,*/G:3_*,*/G:4_
1182 *,*/G:5_*,*/G:6_*,*/G:7_*}",
1183 "QS L-Phrase_Num-Syls<=8{*/G:0_*,*/G:1_*,*/G:2_*,*/G:3_*,*/G:4_
1184 *,*/G:5_*,*/G:6_*,*/G:7_*,*/G:8_*}",
1185 "QS L-Phrase_Num-Syls<=9{*/G:?_*}",
1186 "QS L-Phrase_Num-Syls<=10{*/G:?_*,*/G:10_*}",
1187 "QS L-Phrase_Num-Syls<=11{*/G:?_*,*/G:10_*,*/G:11_*}",
1188 "QS L-Phrase_Num-Syls<=12{*/G:?_*,*/G:10_*,*/G:11_*,*/G:12_*}",
1189 "QS L-Phrase_Num-Syls<=13{*/G:?_*,*/G:10_*,*/G:11_*,*/G:12_*,*/
1190 G:13_*}",

```



```

1185 "QS L-Phrase_Num-Syls<=14{*/G:?_*,*/G:10_*,*/G:11_*,*/G:12_*,*/
      G:13_*,*/G:14_*}",
1186 "QS L-Phrase_Num-Syls<=15{*/G:?_*,*/G:10_*,*/G:11_*,*/G:12_*,*/
      G:13_*,*/G:14_*,*/G:15_*}",
1187 "QS L-Phrase_Num-Syls<=16{*/G:?_*,*/G:10_*,*/G:11_*,*/G:12_*,*/
      G:13_*,*/G:14_*,*/G:15_*,*/G:16_*}",
1188 "QS L-Phrase_Num-Syls<=17{*/G:?_*,*/G:10_*,*/G:11_*,*/G:12_*,*/
      G:13_*,*/G:14_*,*/G:15_*,*/G:16_*,*/G:17_*}",
1189 "QS L-Phrase_Num-Syls<=18{*/G:?_*,*/G:10_*,*/G:11_*,*/G:12_*,*/
      G:13_*,*/G:14_*,*/G:15_*,*/G:16_*,*/G:17_*,*/G:18_*}",
1190 "QS L-Phrase_Num-Syls<=19{*/G:?_*,*/G:1?_*}",
1191 "QS L-Phrase_Num-Syls<=20{*/G:?_*,*/G:1?_*,*/G:20_*}",
1192 ]
1193
1194 PHRASE_LEN_IN_NO_WORDS_EQ_L = [
1195     "QS L-Phrase_Num-Words==0{*_0/H:*}",
1196     "QS L-Phrase_Num-Words==1{*_1/H:*}",
1197     "QS L-Phrase_Num-Words==2{*_2/H:*}",
1198     "QS L-Phrase_Num-Words==3{*_3/H:*}",
1199     "QS L-Phrase_Num-Words==4{*_4/H:*}",
1200     "QS L-Phrase_Num-Words==5{*_5/H:*}",
1201     "QS L-Phrase_Num-Words==6{*_6/H:*}",
1202     "QS L-Phrase_Num-Words==7{*_7/H:*}",
1203     "QS L-Phrase_Num-Words==8{*_8/H:*}",
1204     "QS L-Phrase_Num-Words==9{*_9/H:*}",
1205     "QS L-Phrase_Num-Words==10{*_10/H:*}",
1206     "QS L-Phrase_Num-Words==11{*_11/H:*}",
1207     "QS L-Phrase_Num-Words==12{*_12/H:*}",
1208     "QS L-Phrase_Num-Words==13{*_13/H:*}",
1209 ]
1210 PHRASE_LEN_IN_NO_WORDS_LTE_L = [
1211     "QS L-Phrase_Num-Words<=1{*_0/H:*,*_1/H:*}",
1212     "QS L-Phrase_Num-Words<=2{*_0/H:*,*_1/H:*,*_2/H:*}",
1213     "QS L-Phrase_Num-Words<=3{*_0/H:*,*_1/H:*,*_2/H:*,*_3/H:*}",
1214     "QS L-Phrase_Num-Words<=4{*_0/H:*,*_1/H:*,*_2/H:*,*_3/H:*,*_4/H
      :*}",
1215     "QS L-Phrase_Num-Words<=5{*_0/H:*,*_1/H:*,*_2/H:*,*_3/H:*,*_4/H
      :*,*_5/H:*}",
1216     "QS L-Phrase_Num-Words<=6{*_0/H:*,*_1/H:*,*_2/H:*,*_3/H:*,*_4/H
      :*,*_5/H:*,*_6/H:*}",
1217     "QS L-Phrase_Num-Words<=7{*_0/H:*,*_1/H:*,*_2/H:*,*_3/H:*,*_4/H
      :*,*_5/H:*,*_6/H:*,*_7/H:*}",
1218     "QS L-Phrase_Num-Words<=8{*_0/H:*,*_1/H:*,*_2/H:*,*_3/H:*,*_4/H
      :*,*_5/H:*,*_6/H:*,*_7/H:*,*_8/H:*}",
1219     "QS L-Phrase_Num-Words<=9{*_?/H:*}",
1220     "QS L-Phrase_Num-Words<=10{*_?/H:*,*_10/H:*}",
1221     "QS L-Phrase_Num-Words<=11{*_?/H:*,*_10/H:*,*_11/H:*}",

```

```

1222 "QS L-Phrase_Num-Words<=12{*_?/H:*,*_10/H:*,*_11/H:*,*_12/H
      :*}",
1223 "QS L-Phrase_Num-Words<=13{*_?/H:*,*_10/H:*,*_11/H:*,*_12/H:*,*
      _13/H:*}",
1224 ]
1225
1226 PHRASE_LEN_IN_NO_SYLS_EQ_C = [
1227     "QS C-Phrase_Num-Syls==x{*/H:x=*}",
1228     "QS C-Phrase_Num-Syls==0{*/H:0=*}",
1229     "QS C-Phrase_Num-Syls==1{*/H:1=*}",
1230     "QS C-Phrase_Num-Syls==2{*/H:2=*}",
1231     "QS C-Phrase_Num-Syls==3{*/H:3=*}",
1232     "QS C-Phrase_Num-Syls==4{*/H:4=*}",
1233     "QS C-Phrase_Num-Syls==5{*/H:5=*}",
1234     "QS C-Phrase_Num-Syls==6{*/H:6=*}",
1235     "QS C-Phrase_Num-Syls==7{*/H:7=*}",
1236     "QS C-Phrase_Num-Syls==8{*/H:8=*}",
1237     "QS C-Phrase_Num-Syls==9{*/H:9=*}",
1238     "QS C-Phrase_Num-Syls==10{*/H:10=*}",
1239     "QS C-Phrase_Num-Syls==11{*/H:11=*}",
1240     "QS C-Phrase_Num-Syls==12{*/H:12=*}",
1241     "QS C-Phrase_Num-Syls==13{*/H:13=*}",
1242     "QS C-Phrase_Num-Syls==14{*/H:14=*}",
1243     "QS C-Phrase_Num-Syls==15{*/H:15=*}",
1244     "QS C-Phrase_Num-Syls==16{*/H:16=*}",
1245     "QS C-Phrase_Num-Syls==17{*/H:17=*}",
1246     "QS C-Phrase_Num-Syls==18{*/H:18=*}",
1247     "QS C-Phrase_Num-Syls==19{*/H:19=*}",
1248     "QS C-Phrase_Num-Syls==20{*/H:20=*}",
1249 ]
1250 PHRASE_LEN_IN_NO_SYLS_LTE_C = [
1251     "QS C-Phrase_Num-Syls<=0{*/H:x=*,*/H:0=*}",
1252     "QS C-Phrase_Num-Syls<=1{*/H:x=*,*/H:0=*,*/H:1=*}",
1253     "QS C-Phrase_Num-Syls<=2{*/H:x=*,*/H:0=*,*/H:1=*,*/H:2=*}",
1254     "QS C-Phrase_Num-Syls<=3{*/H:x=*,*/H:0=*,*/H:1=*,*/H:2=*,*/H
      :3=*}",
1255     "QS C-Phrase_Num-Syls<=4{*/H:x=*,*/H:0=*,*/H:1=*,*/H:2=*,*/H
      :3=*,*/H:4=*}",
1256     "QS C-Phrase_Num-Syls<=5{*/H:x=*,*/H:0=*,*/H:1=*,*/H:2=*,*/H
      :3=*,*/H:4=*,*/H:5=*}",
1257     "QS C-Phrase_Num-Syls<=6{*/H:x=*,*/H:0=*,*/H:1=*,*/H:2=*,*/H
      :3=*,*/H:4=*,*/H:5=*,*/H:6=*}",
1258     "QS C-Phrase_Num-Syls<=7{*/H:x=*,*/H:0=*,*/H:1=*,*/H:2=*,*/H
      :3=*,*/H:4=*,*/H:5=*,*/H:6=*,*/H:7=*}",
1259     "QS C-Phrase_Num-Syls<=8{*/H:x=*,*/H:0=*,*/H:1=*,*/H:2=*,*/H
      :3=*,*/H:4=*,*/H:5=*,*/H:6=*,*/H:7=*,*/H:8=*}",
1260     "QS C-Phrase_Num-Syls<=9{*/H:?=*}",

```

```

1261 "QS C-Phrase_Num-Syls<=10{*/H:?=* ,*/H:10=*}" ,
1262 "QS C-Phrase_Num-Syls<=11{*/H:?=* ,*/H:10=* ,*/H:11=*}" ,
1263 "QS C-Phrase_Num-Syls<=12{*/H:?=* ,*/H:10=* ,*/H:11=* ,*/H:12=*}" ,
1264 "QS C-Phrase_Num-Syls<=13{*/H:?=* ,*/H:10=* ,*/H:11=* ,*/H:12=* ,*/
      H:13=*}" ,
1265 "QS C-Phrase_Num-Syls<=14{*/H:?=* ,*/H:10=* ,*/H:11=* ,*/H:12=* ,*/
      H:13=* ,*/H:14=*}" ,
1266 "QS C-Phrase_Num-Syls<=15{*/H:?=* ,*/H:10=* ,*/H:11=* ,*/H:12=* ,*/
      H:13=* ,*/H:14=* ,*/H:15=*}" ,
1267 "QS C-Phrase_Num-Syls<=16{*/H:?=* ,*/H:10=* ,*/H:11=* ,*/H:12=* ,*/
      H:13=* ,*/H:14=* ,*/H:15=* ,*/H:16=*}" ,
1268 "QS C-Phrase_Num-Syls<=17{*/H:?=* ,*/H:10=* ,*/H:11=* ,*/H:12=* ,*/
      H:13=* ,*/H:14=* ,*/H:15=* ,*/H:16=* ,*/H:17=*}" ,
1269 "QS C-Phrase_Num-Syls<=18{*/H:?=* ,*/H:10=* ,*/H:11=* ,*/H:12=* ,*/
      H:13=* ,*/H:14=* ,*/H:15=* ,*/H:16=* ,*/H:17=* ,*/H:18=*}" ,
1270 "QS C-Phrase_Num-Syls<=19{*/H:?=* ,*/H:1?=*}" ,
1271 "QS C-Phrase_Num-Syls<=20{*/H:?=* ,*/H:1?=* ,*/H:20=*}" ,
1272 ]
1273
1274 PHRASE_LEN_IN_NO_WORDS_EQ_C = [
1275     "QS C-Phrase_Num-Words==x{**x@*}" ,
1276     "QS C-Phrase_Num-Words==0{**0@*}" ,
1277     "QS C-Phrase_Num-Words==1{**1@*}" ,
1278     "QS C-Phrase_Num-Words==2{**2@*}" ,
1279     "QS C-Phrase_Num-Words==3{**3@*}" ,
1280     "QS C-Phrase_Num-Words==4{**4@*}" ,
1281     "QS C-Phrase_Num-Words==5{**5@*}" ,
1282     "QS C-Phrase_Num-Words==6{**6@*}" ,
1283     "QS C-Phrase_Num-Words==7{**7@*}" ,
1284     "QS C-Phrase_Num-Words==8{**8@*}" ,
1285     "QS C-Phrase_Num-Words==9{**9@*}" ,
1286     "QS C-Phrase_Num-Words==10{**10@*}" ,
1287     "QS C-Phrase_Num-Words==11{**11@*}" ,
1288     "QS C-Phrase_Num-Words==12{**12@*}" ,
1289     "QS C-Phrase_Num-Words==13{**13@*}" ,
1290 ]
1291 PHRASE_LEN_IN_NO_WORDS_LTE_C =[
1292     "QS C-Phrase_Num-Words<=0{**x@* ,**0@*}" ,
1293     "QS C-Phrase_Num-Words<=1{**x@* ,**0@* ,**1@*}" ,
1294     "QS C-Phrase_Num-Words<=2{**x@* ,**0@* ,**1@* ,**2@*}" ,
1295     "QS C-Phrase_Num-Words<=3{**x@* ,**0@* ,**1@* ,**2@* ,**3@*}" ,
1296     "QS C-Phrase_Num-Words<=4{**x@* ,**0@* ,**1@* ,**2@* ,**3@* ,**4@
      *}" ,
1297     "QS C-Phrase_Num-Words<=5{**x@* ,**0@* ,**1@* ,**2@* ,**3@* ,**4@
      * ,**5@*}" ,
1298     "QS C-Phrase_Num-Words<=6{**x@* ,**0@* ,**1@* ,**2@* ,**3@* ,**4@
      * ,**5@* ,**6@*}" ,

```

```

1299 "QS C-Phrase_Num-Words<=7{*=x@*,*=0@*,*=1@*,*=2@*,*=3@*,*=4@
      *,*=5@*,*=6@*,*=7@*}",
1300 "QS C-Phrase_Num-Words<=8{*=x@*,*=0@*,*=1@*,*=2@*,*=3@*,*=4@
      *,*=5@*,*=6@*,*=7@*,*=8@*}",
1301 "QS C-Phrase_Num-Words<=9{*=x@*,*=0@*,*=1@*,*=2@*,*=3@*,*=4@
      *,*=5@*,*=6@*,*=7@*,*=8@*,*=9@*}",
1302 "QS C-Phrase_Num-Words<=10{*=x@*,*=0@*,*=1@*,*=2@*,*=3@*,*=4@
      *,*=5@*,*=6@*,*=7@*,*=8@*,*=9@*,*=10@*}",
1303 "QS C-Phrase_Num-Words<=11{*=x@*,*=0@*,*=1@*,*=2@*,*=3@*,*=4@
      *,*=5@*,*=6@*,*=7@*,*=8@*,*=9@*,*=10@*,*=11@*}",
1304 "QS C-Phrase_Num-Words<=12{*=x@*,*=0@*,*=1@*,*=2@*,*=3@*,*=4@
      *,*=5@*,*=6@*,*=7@*,*=8@*,*=9@*,*=10@*,*=11@*,*=12@*}",
1305 "QS C-Phrase_Num-Words<=13{*=x@*,*=0@*,*=1@*,*=2@*,*=3@*,*=4@
      *,*=5@*,*=6@*,*=7@*,*=8@*,*=9@*,*=10@*,*=11@*,*=12@*,*=13@
      *}" ,
1306 ]
1307
1308 PHRASE_POSITION_IN_UTTERANCE_FW_EQ_C = [
1309     "QS Pos_C-Phrase_in_Utterance(Fw)==1{*@1=*}",
1310     "QS Pos_C-Phrase_in_Utterance(Fw)==2{*@2=*}",
1311     "QS Pos_C-Phrase_in_Utterance(Fw)==3{*@3=*}",
1312     "QS Pos_C-Phrase_in_Utterance(Fw)==4{*@4=*}",
1313 ]
1314 PHRASE_POSITION_IN_UTTERANCE_FW_LTE_C = [
1315     "QS Pos_C-Phrase_in_Utterance(Fw)<=2{*@1=*,*@2=*}",
1316     "QS Pos_C-Phrase_in_Utterance(Fw)<=3{*@1=*,*@2=*,*@3=*}",
1317     "QS Pos_C-Phrase_in_Utterance(Fw)<=4{*@1=*,*@2=*,*@3=*,*@4=*}",
1318 ]
1319
1320 PHRASE_POSITION_IN_UTTERANCE_BW_EQ_C = [
1321     "QS Pos_C-Phrase_in_Utterance(Bw)==1{*=1|*}",
1322     "QS Pos_C-Phrase_in_Utterance(Bw)==2{*=2|*}",
1323     "QS Pos_C-Phrase_in_Utterance(Bw)==3{*=3|*}",
1324     "QS Pos_C-Phrase_in_Utterance(Bw)==4{*=4|*}",
1325 ]
1326 PHRASE_POSITION_IN_UTTERANCE_BW_LTE_C = [
1327     "QS Pos_C-Phrase_in_Utterance(Bw)<=2{*=1|*,*=2|*}",
1328     "QS Pos_C-Phrase_in_Utterance(Bw)<=3{*=1|*,*=2|*,*=3|*}",
1329     "QS Pos_C-Phrase_in_Utterance(Bw)<=4{*=1|*,*=2|*,*=3|*,*=4|*}",
1330 ]
1331
1332 PHRASE_LEN_IN_NO_SYLS_EQ_R = [
1333     "QS R-Phrase_Num-Syls==0{*/I:0=*}",
1334     "QS R-Phrase_Num-Syls==1{*/I:1=*}",
1335     "QS R-Phrase_Num-Syls==2{*/I:2=*}",
1336     "QS R-Phrase_Num-Syls==3{*/I:3=*}",
1337     "QS R-Phrase_Num-Syls==4{*/I:4=*}",

```

```

1338 "QS R-Phrase_Num-Syls==5{*/I:5=*}" ,
1339 "QS R-Phrase_Num-Syls==6{*/I:6=*}" ,
1340 "QS R-Phrase_Num-Syls==7{*/I:7=*}" ,
1341 "QS R-Phrase_Num-Syls==8{*/I:8=*}" ,
1342 "QS R-Phrase_Num-Syls==9{*/I:9=*}" ,
1343 "QS R-Phrase_Num-Syls==10{*/I:10=*}" ,
1344 "QS R-Phrase_Num-Syls==11{*/I:11=*}" ,
1345 "QS R-Phrase_Num-Syls==12{*/I:12=*}" ,
1346 "QS R-Phrase_Num-Syls==13{*/I:13=*}" ,
1347 "QS R-Phrase_Num-Syls==14{*/I:14=*}" ,
1348 "QS R-Phrase_Num-Syls==15{*/I:15=*}" ,
1349 "QS R-Phrase_Num-Syls==16{*/I:16=*}" ,
1350 "QS R-Phrase_Num-Syls==17{*/I:17=*}" ,
1351 "QS R-Phrase_Num-Syls==18{*/I:18=*}" ,
1352 "QS R-Phrase_Num-Syls==19{*/I:19=*}" ,
1353 "QS R-Phrase_Num-Syls==20{*/I:20=*}" ,
1354 ]
1355 PHRASE_LEN_IN_NO_SYLS_LTE_R = [
1356 "QS R-Phrase_Num-Syls<=1{*/I:0=*,*/I:1=*}" ,
1357 "QS R-Phrase_Num-Syls<=2{*/I:0=*,*/I:1=*,*/I:2=*}" ,
1358 "QS R-Phrase_Num-Syls<=3{*/I:0=*,*/I:1=*,*/I:2=*,*/I:3=*}" ,
1359 "QS R-Phrase_Num-Syls<=4{*/I:0=*,*/I:1=*,*/I:2=*,*/I:3=*,*/I
:4=*}" ,
1360 "QS R-Phrase_Num-Syls<=5{*/I:0=*,*/I:1=*,*/I:2=*,*/I:3=*,*/I
:4=*,*/I:5=*}" ,
1361 "QS R-Phrase_Num-Syls<=6{*/I:0=*,*/I:1=*,*/I:2=*,*/I:3=*,*/I
:4=*,*/I:5=*,*/I:6=*}" ,
1362 "QS R-Phrase_Num-Syls<=7{*/I:0=*,*/I:1=*,*/I:2=*,*/I:3=*,*/I
:4=*,*/I:5=*,*/I:6=*,*/I:7=*}" ,
1363 "QS R-Phrase_Num-Syls<=8{*/I:0=*,*/I:1=*,*/I:2=*,*/I:3=*,*/I
:4=*,*/I:5=*,*/I:6=*,*/I:7=*,*/I:8=*}" ,
1364 "QS R-Phrase_Num-Syls<=9{*/I:?=*,*/I:10=*}" ,
1365 "QS R-Phrase_Num-Syls<=10{*/I:?=*,*/I:10=*,*/I:11=*}" ,
1366 "QS R-Phrase_Num-Syls<=11{*/I:?=*,*/I:10=*,*/I:11=*,*/I:12=*}" ,
1367 "QS R-Phrase_Num-Syls<=12{*/I:?=*,*/I:10=*,*/I:11=*,*/I:12=*,*/
I:13=*}" ,
1368 "QS R-Phrase_Num-Syls<=13{*/I:?=*,*/I:10=*,*/I:11=*,*/I:12=*,*/
I:13=*,*/I:14=*}" ,
1369 "QS R-Phrase_Num-Syls<=14{*/I:?=*,*/I:10=*,*/I:11=*,*/I:12=*,*/
I:13=*,*/I:14=*,*/I:15=*}" ,
1370 "QS R-Phrase_Num-Syls<=15{*/I:?=*,*/I:10=*,*/I:11=*,*/I:12=*,*/
I:13=*,*/I:14=*,*/I:15=*,*/I:16=*}" ,
1371 "QS R-Phrase_Num-Syls<=16{*/I:?=*,*/I:10=*,*/I:11=*,*/I:12=*,*/
I:13=*,*/I:14=*,*/I:15=*,*/I:16=*,*/I:17=*}" ,
1372 "QS R-Phrase_Num-Syls<=17{*/I:?=*,*/I:10=*,*/I:11=*,*/I:12=*,*/
I:13=*,*/I:14=*,*/I:15=*,*/I:16=*,*/I:17=*,*/I:18=*}" ,
1373 "QS R-Phrase_Num-Syls<=18{*/I:?=*,*/I:10=*,*/I:11=*,*/I:12=*,*/
I:13=*,*/I:14=*,*/I:15=*,*/I:16=*,*/I:17=*,*/I:18=*}" ,

```

```

1374 "QS R-Phrase_Num-Syls<=19{*/I:?=?,*/I:1?=?}" ,
1375 "QS R-Phrase_Num-Syls<=20{*/I:?=?,*/I:1?=?,*/I:20=?}" ,
1376 ]
1377
1378 PHRASE_LEN_IN_NO_WORDS_EQ_R = [
1379 "QS R-Phrase_Num-Words==0{*=0/J:*}" ,
1380 "QS R-Phrase_Num-Words==1{*=1/J:*}" ,
1381 "QS R-Phrase_Num-Words==2{*=2/J:*}" ,
1382 "QS R-Phrase_Num-Words==3{*=3/J:*}" ,
1383 "QS R-Phrase_Num-Words==4{*=4/J:*}" ,
1384 "QS R-Phrase_Num-Words==5{*=5/J:*}" ,
1385 "QS R-Phrase_Num-Words==6{*=6/J:*}" ,
1386 "QS R-Phrase_Num-Words==7{*=7/J:*}" ,
1387 "QS R-Phrase_Num-Words==8{*=8/J:*}" ,
1388 "QS R-Phrase_Num-Words==9{*=9/J:*}" ,
1389 "QS R-Phrase_Num-Words==10{*=10/J:*}" ,
1390 "QS R-Phrase_Num-Words==11{*=11/J:*}" ,
1391 "QS R-Phrase_Num-Words==12{*=12/J:*}" ,
1392 "QS R-Phrase_Num-Words==13{*=13/J:*}" ,
1393 "QS R-Phrase_Num-Words==14{*=14/J:*}" ,
1394 "QS R-Phrase_Num-Words==15{*=15/J:*}" ,
1395 ]
1396 PHRASE_LEN_IN_NO_WORDS_LTE_R = [
1397 "QS R-Phrase_Num-Words<=1{*=0/J:*,*=1/J:*}" ,
1398 "QS R-Phrase_Num-Words<=2{*=0/J:*,*=1/J:*,*=2/J:*}" ,
1399 "QS R-Phrase_Num-Words<=3{*=0/J:*,*=1/J:*,*=2/J:*,*=3/J:*}" ,
1400 "QS R-Phrase_Num-Words<=4{*=0/J:*,*=1/J:*,*=2/J:*,*=3/J:*,*=4/J
    :*}" ,
1401 "QS R-Phrase_Num-Words<=5{*=0/J:*,*=1/J:*,*=2/J:*,*=3/J:*,*=4/J
    :*,*=5/J:*}" ,
1402 "QS R-Phrase_Num-Words<=6{*=0/J:*,*=1/J:*,*=2/J:*,*=3/J:*,*=4/J
    :*,*=5/J:*,*=6/J:*}" ,
1403 "QS R-Phrase_Num-Words<=7{*=0/J:*,*=1/J:*,*=2/J:*,*=3/J:*,*=4/J
    :*,*=5/J:*,*=6/J:*,*=7/J:*}" ,
1404 "QS R-Phrase_Num-Words<=8{*=0/J:*,*=1/J:*,*=2/J:*,*=3/J:*,*=4/J
    :*,*=5/J:*,*=6/J:*,*=7/J:*,*=8/J:*}" ,
1405 "QS R-Phrase_Num-Words<=9{*=?/J:*}" ,
1406 "QS R-Phrase_Num-Words<=10{*=?/J:*,*=10/J:*}" ,
1407 "QS R-Phrase_Num-Words<=11{*=?/J:*,*=10/J:*,*=11/J:*}" ,
1408 "QS R-Phrase_Num-Words<=12{*=?/J:*,*=10/J:*,*=11/J:*,*=12/J
    :*}" ,
1409 "QS R-Phrase_Num-Words<=13{*=?/J:*,*=10/J:*,*=11/J:*,*=12/J
    :*,*=13/J:*}" ,
1410 "QS R-Phrase_Num-Words<=14{*=?/J:*,*=10/J:*,*=11/J:*,*=12/J
    :*,*=13/J:*,*=14/J:*}" ,
1411 "QS R-Phrase_Num-Words<=15{*=?/J:*,*=10/J:*,*=11/J:*,*=12/J
    :*,*=13/J:*,*=14/J:*,*=15/J:*}" ,

```

```

1412 ]
1413
1414 UTTERANCE_LEN_IN_NO_SYLS_EQ = [
1415     "QS Num-Syls_in_Utterance==1{*/J:1+*}",
1416     "QS Num-Syls_in_Utterance==2{*/J:2+*}",
1417     "QS Num-Syls_in_Utterance==3{*/J:3+*}",
1418     "QS Num-Syls_in_Utterance==4{*/J:4+*}",
1419     "QS Num-Syls_in_Utterance==5{*/J:5+*}",
1420     "QS Num-Syls_in_Utterance==6{*/J:6+*}",
1421     "QS Num-Syls_in_Utterance==7{*/J:7+*}",
1422     "QS Num-Syls_in_Utterance==8{*/J:8+*}",
1423     "QS Num-Syls_in_Utterance==9{*/J:9+*}",
1424     "QS Num-Syls_in_Utterance==10{*/J:10+*}",
1425     "QS Num-Syls_in_Utterance==11{*/J:11+*}",
1426     "QS Num-Syls_in_Utterance==12{*/J:12+*}",
1427     "QS Num-Syls_in_Utterance==13{*/J:13+*}",
1428     "QS Num-Syls_in_Utterance==14{*/J:14+*}",
1429     "QS Num-Syls_in_Utterance==15{*/J:15+*}",
1430     "QS Num-Syls_in_Utterance==16{*/J:16+*}",
1431     "QS Num-Syls_in_Utterance==17{*/J:17+*}",
1432     "QS Num-Syls_in_Utterance==18{*/J:18+*}",
1433     "QS Num-Syls_in_Utterance==19{*/J:19+*}",
1434     "QS Num-Syls_in_Utterance==20{*/J:20+*}",
1435     "QS Num-Syls_in_Utterance==21{*/J:21+*}",
1436     "QS Num-Syls_in_Utterance==22{*/J:22+*}",
1437     "QS Num-Syls_in_Utterance==23{*/J:23+*}",
1438     "QS Num-Syls_in_Utterance==24{*/J:24+*}",
1439     "QS Num-Syls_in_Utterance==25{*/J:25+*}",
1440     "QS Num-Syls_in_Utterance==26{*/J:26+*}",
1441     "QS Num-Syls_in_Utterance==27{*/J:27+*}",
1442     "QS Num-Syls_in_Utterance==28{*/J:28+*}",
1443 ]
1444 UTTERANCE_LEN_IN_NO_SYLS_LTE = [
1445     "QS Num-Syls_in_Utterance<=2{*/J:1+*,*/J:2+*}",
1446     "QS Num-Syls_in_Utterance<=3{*/J:1+*,*/J:2+*,*/J:3+*}",
1447     "QS Num-Syls_in_Utterance<=4{*/J:1+*,*/J:2+*,*/J:3+*,*/J:4+*}",
1448     "QS Num-Syls_in_Utterance<=5{*/J:1+*,*/J:2+*,*/J:3+*,*/J:4+*,*/
        J:5+*}",
1449     "QS Num-Syls_in_Utterance<=6{*/J:1+*,*/J:2+*,*/J:3+*,*/J:4+*,*/
        J:5+*,*/J:6+*}",
1450     "QS Num-Syls_in_Utterance<=7{*/J:1+*,*/J:2+*,*/J:3+*,*/J:4+*,*/
        J:5+*,*/J:6+*,*/J:7+*}",
1451     "QS Num-Syls_in_Utterance<=8{*/J:1+*,*/J:2+*,*/J:3+*,*/J:4+*,*/
        J:5+*,*/J:6+*,*/J:7+*,*/J:8+*}",
1452     "QS Num-Syls_in_Utterance<=9{*/J:??+*}",
1453     "QS Num-Syls_in_Utterance<=10{*/J:??+*,*/J:10+*}",
1454     "QS Num-Syls_in_Utterance<=11{*/J:??+*,*/J:10+*,*/J:11+*}",

```

```

1455 "QS Num-Syls_in_Utterance<=12{*/J:?+*,*/J:10+*,*/J:11+*,*/J
      :12+*}",
1456 "QS Num-Syls_in_Utterance<=13{*/J:?+*,*/J:10+*,*/J:11+*,*/J
      :12+*,*/J:13+*}",
1457 "QS Num-Syls_in_Utterance<=14{*/J:?+*,*/J:10+*,*/J:11+*,*/J
      :12+*,*/J:13+*,*/J:14+*}",
1458 "QS Num-Syls_in_Utterance<=15{*/J:?+*,*/J:10+*,*/J:11+*,*/J
      :12+*,*/J:13+*,*/J:14+*,*/J:15+*}",
1459 "QS Num-Syls_in_Utterance<=16{*/J:?+*,*/J:10+*,*/J:11+*,*/J
      :12+*,*/J:13+*,*/J:14+*,*/J:15+*,*/J:16+*}",
1460 "QS Num-Syls_in_Utterance<=17{*/J:?+*,*/J:10+*,*/J:11+*,*/J
      :12+*,*/J:13+*,*/J:14+*,*/J:15+*,*/J:16+*,*/J:17+*}",
1461 "QS Num-Syls_in_Utterance<=18{*/J:?+*,*/J:10+*,*/J:11+*,*/J
      :12+*,*/J:13+*,*/J:14+*,*/J:15+*,*/J:16+*,*/J:17+*,*/J
      :18+*}",
1462 "QS Num-Syls_in_Utterance<=19{*/J:?+*,*/J:1?+*}",
1463 "QS Num-Syls_in_Utterance<=20{*/J:?+*,*/J:1?+*,*/J:20+*}",
1464 "QS Num-Syls_in_Utterance<=21{*/J:?+*,*/J:1?+*,*/J:20+*,*/J
      :21+*}",
1465 "QS Num-Syls_in_Utterance<=22{*/J:?+*,*/J:1?+*,*/J:20+*,*/J
      :21+*,*/J:22+*}",
1466 "QS Num-Syls_in_Utterance<=23{*/J:?+*,*/J:1?+*,*/J:20+*,*/J
      :21+*,*/J:22+*,*/J:23+*}",
1467 "QS Num-Syls_in_Utterance<=24{*/J:?+*,*/J:1?+*,*/J:20+*,*/J
      :21+*,*/J:22+*,*/J:23+*,*/J:24+*}",
1468 "QS Num-Syls_in_Utterance<=25{*/J:?+*,*/J:1?+*,*/J:20+*,*/J
      :21+*,*/J:22+*,*/J:23+*,*/J:24+*,*/J:25+*}",
1469 "QS Num-Syls_in_Utterance<=26{*/J:?+*,*/J:1?+*,*/J:20+*,*/J
      :21+*,*/J:22+*,*/J:23+*,*/J:24+*,*/J:25+*,*/J:26+*}",
1470 "QS Num-Syls_in_Utterance<=27{*/J:?+*,*/J:1?+*,*/J:20+*,*/J
      :21+*,*/J:22+*,*/J:23+*,*/J:24+*,*/J:25+*,*/J:26+*,*/J
      :27+*}",
1471 "QS Num-Syls_in_Utterance<=28{*/J:?+*,*/J:1?+*,*/J:20+*,*/J
      :21+*,*/J:22+*,*/J:23+*,*/J:24+*,*/J:25+*,*/J:26+*,*/J
      :27+*,*/J:28+*}",
1472 ]
1473
1474 UTTERANCE_LEN_IN_NO_WORDS_EQ = [
1475     "QS Num-Words_in_Utterance==1{**+1-*}",
1476     "QS Num-Words_in_Utterance==2{**+2-*}",
1477     "QS Num-Words_in_Utterance==3{**+3-*}",
1478     "QS Num-Words_in_Utterance==4{**+4-*}",
1479     "QS Num-Words_in_Utterance==5{**+5-*}",
1480     "QS Num-Words_in_Utterance==6{**+6-*}",
1481     "QS Num-Words_in_Utterance==7{**+7-*}",
1482     "QS Num-Words_in_Utterance==8{**+8-*}",
1483     "QS Num-Words_in_Utterance==9{**+9-*}",

```



```

1484     "QS Num-Words_in_Utterance==10{**10-*}",
1485     "QS Num-Words_in_Utterance==11{**11-*}",
1486     "QS Num-Words_in_Utterance==12{**12-*}",
1487     "QS Num-Words_in_Utterance==13{**13-*}",
1488 ]
1489 UTTERANCE_LEN_IN_NO_WORDS_LTE = [
1490     "QS Num-Words_in_Utterance<=2{**1-*,**2-*}",
1491     "QS Num-Words_in_Utterance<=3{**1-*,**2-*,**3-*}",
1492     "QS Num-Words_in_Utterance<=4{**1-*,**2-*,**3-*,**4-*}",
1493     "QS Num-Words_in_Utterance<=5{**1-*,**2-*,**3-*,**4-*,**5-*}",
1494     "QS Num-Words_in_Utterance
1495         <=6{**1-*,**2-*,**3-*,**4-*,**5-*,**6-*}",
1496     "QS Num-Words_in_Utterance
1497         <=7{**1-*,**2-*,**3-*,**4-*,**5-*,**6-*,**7-*}",
1498     "QS Num-Words_in_Utterance
1499         <=8{**1-*,**2-*,**3-*,**4-*,**5-*,**6-*,**7-*,**8-*}",
1500     "QS Num-Words_in_Utterance<=9{**?-*}",
1501     "QS Num-Words_in_Utterance<=10{**?-*,**10-*}",
1502     "QS Num-Words_in_Utterance<=11{**?-*,**10-*,**11-*}",
1503     "QS Num-Words_in_Utterance<=12{**?-*,**10-*,**11-*,**12-*}",
1504     "QS Num-Words_in_Utterance
1505         <=13{**?-*,**10-*,**11-*,**12-*,**13-*}",
1506 ]
1507
1508 UTTERANCE_LEN_IN_NO_PHRASES_EQ = [
1509     "QS Num-Phrases_in_Utterance==1{* -1/K:*}",
1510     "QS Num-Phrases_in_Utterance==2{* -2/K:*}",
1511     "QS Num-Phrases_in_Utterance==3{* -3/K:*}",
1512     "QS Num-Phrases_in_Utterance==4{* -4/K:*}",
1513 ]
1514
1515 UTTERANCE_LEN_IN_NO_PHRASES_LTE = [
1516     "QS Num-Phrases_in_Utterance<=2{* -1/K:*,**-2/K:*}",
1517     "QS Num-Phrases_in_Utterance<=3{* -1/K:*,**-2/K:*,**-3/K:*}",
1518     "QS Num-Phrases_in_Utterance<=4{* -1/K:*,**-2/K:*,**-3/K:*,**-4/K
1519         :*}",
1520 ]
1521
1522 #####
1523 #HIGH-LEVEL FEATURE GROUPS
1524 #####
1525
1526 #####
1527 # SEGMENTAL
1528 #####
1529
1530 QUINTPHONE_SEGMENT_VC = [
1531     QUINTPHONE_SEGMENT_VC_LL ,

```

```

1526     QUINTPHONE_SEGMENT_VC_L ,
1527     QUINTPHONE_SEGMENT_VC_C ,
1528     QUINTPHONE_SEGMENT_VC_R ,
1529     QUINTPHONE_SEGMENT_VC_RR ,
1530 ]
1531 QUINTPHONE_SEGMENT_ARTICULATORY_TYPE = [
1532     QUINTPHONE_SEGMENT_ARTICULATORY_TYPE_LL ,
1533     QUINTPHONE_SEGMENT_ARTICULATORY_TYPE_L ,
1534     QUINTPHONE_SEGMENT_ARTICULATORY_TYPE_C ,
1535     QUINTPHONE_SEGMENT_ARTICULATORY_TYPE_R ,
1536     QUINTPHONE_SEGMENT_ARTICULATORY_TYPE_RR ,
1537 ]
1538 SEG_POS_IN_SYL = [
1539     NUM_SEG_IN_SYL_FW_EQ ,
1540     NUM_SEG_IN_SYL_FW_LTE ,
1541     NUM_SEG_IN_SYL_BW_EQ ,
1542     NUM_SEG_IN_SYL_BW_LTE ,
1543 ]
1544 QUINTPHONE_SEGMENTAL_FEATURES = [
1545     QUINTPHONE_SEGMENT_VC ,
1546     QUINTPHONE_SEGMENT_ARTICULATORY_TYPE ,
1547 ]
1548
1549 QUINTPHONE = [
1550     QUINTPHONE_SEGMENT_VC ,
1551     QUINTPHONE_SEGMENT_ARTICULATORY_TYPE ,
1552     SEG_POS_IN_SYL ,
1553     QUINTPHONE_SEGMENTAL_FEATURES ,
1554 ]
1555
1556 #####
1557 # SYLLABIC
1558 #####
1559
1560 SYLLABLE_STRESS = [
1561     SYL_STRESSED_L ,
1562     SYL_STRESSED_C ,
1563     SYL_STRESSED_R ,
1564 ]
1565 SYLLABLE_ACCENTED = [
1566     SYL_ACCENTED_L ,
1567     SYL_ACCENTED_C ,
1568     SYL_ACCENTED_R ,
1569 ]
1570 SYLLABLE_ACCENT_TYPE = [
1571     SYL_ACCENT_TYPE_L ,
1572     SYL_ACCENT_TYPE_C ,

```

```

1573     SYL_ACCENT_TYPE_R ,
1574 ]
1575 SYLLABLE_LEN_IN_SEGMENTS = [
1576     SYL_LEN_IN_NO_SEG_EQ_L ,
1577     SYL_LEN_IN_NO_SEG_LTE_L ,
1578     SYL_LEN_IN_NO_SEG_EQ_C ,
1579     SYL_LEN_IN_NO_SEG_LTE_C ,
1580     SYL_LEN_IN_NO_SEG_EQ_R ,
1581     SYL_LEN_IN_NO_SEG_LTE_R ,
1582
1583 ]
1584 SYLLABLE_NUCLEUS = [
1585     SYL_VOWEL_TYPE ,
1586 ]
1587
1588 SYLLABLE_POSITION_IN_WORD_FW = [
1589     SYL_POSITION_IN_WORD_FW_EQ ,
1590     SYL_POSITION_IN_WORD_FW_LTE ,
1591 ]
1592 SYLLABLE_POSITION_IN_WORD_BW = [
1593     SYL_POSITION_IN_WORD_BW_EQ ,
1594     SYL_POSITION_IN_WORD_BW_LTE ,
1595 ]
1596 SYLLABLE_POSITION_IN_PHRASE_FW = [
1597     SYL_POSITION_IN_PHRASE_FW_EQ ,
1598     SYL_POSITION_IN_PHRASE_FW_LTE ,
1599 ]
1600 SYLLABLE_POSITION_IN_PHRASE_BW = [
1601     SYL_POSITION_IN_PHRASE_BW_EQ ,
1602     SYL_POSITION_IN_PHRASE_BW_LTE ,
1603 ]
1604 SYLLABLE_POSITION_IN_WORD = [
1605     SYLLABLE_POSITION_IN_WORD_FW ,
1606     SYLLABLE_POSITION_IN_WORD_BW ,
1607 ]
1608 SYLLABLE_POSITION_IN_PHRASE = [
1609     SYLLABLE_POSITION_IN_PHRASE_FW ,
1610     SYLLABLE_POSITION_IN_PHRASE_BW ,
1611 ]
1612
1613 SYLLABLE_POSITION = [
1614     SYLLABLE_POSITION_IN_WORD ,
1615     SYLLABLE_POSITION_IN_PHRASE ,
1616 ]
1617
1618 NUM_STRESSED_SYLS_IN_CURRENT_PHRASE_BEFORE_CURRENT_SYL = [
1619     NUM_STRESSED_SYLS_BEFORE_CURRENT_SYL_IN_CURRENT_PHRASE_EQ ,

```

```

1620     NUM_STRESSED_SYLS_BEFORE_CURRENT_SYL_IN_CURRENT_PHRASE_LTE ,
1621 ]
1622 NUM_STRESSED_SYLS_IN_CURRENT_PHRASE_AFTER_CURRENT_SYL = [
1623     NUM_STRESSED_SYLS_AFTER_CURRENT_SYL_IN_CURRENT_PHRASE_EQ ,
1624     NUM_STRESSED_SYLS_AFTER_CURRENT_SYL_IN_CURRENT_PHRASE_LTE ,
1625 ]
1626 NUM_STRESSED_SYLS_IN_CURRENT_PHRASE_IN_RELATION_TO_CURRENT_SYL_POSITION
    = [
1627     NUM_STRESSED_SYLS_IN_CURRENT_PHRASE_BEFORE_CURRENT_SYL ,
1628     NUM_STRESSED_SYLS_IN_CURRENT_PHRASE_AFTER_CURRENT_SYL ,
1629 ]
1630 NUM_ACCENTED_SYLS_IN_CURRENT_PHRASE_BEFORE_CURRENT_SYL = [
1631     NUM_ACCENTED_SYLS_BEFORE_CURRENT_SYL_IN_CURRENT_PHRASE_EQ ,
1632     NUM_ACCENTED_SYLS_BEFORE_CURRENT_SYL_IN_CURRENT_PHRASE_LTE ,
1633 ]
1634 NUM_ACCENTED_SYLS_IN_CURRENT_PHRASE_AFTER_CURRENT_SYL = [
1635     NUM_ACCENTED_SYLS_AFTER_CURRENT_SYL_IN_CURRENT_PHRASE_EQ ,
1636     NUM_ACCENTED_SYLS_AFTER_CURRENT_SYL_IN_CURRENT_PHRASE_LTE ,
1637 ]
1638 NUM_ACCENTED_SYLS_IN_CURRENT_PHRASE_IN_RELATION_TO_CURRENT_SYL_POSITION
    = [
1639     NUM_ACCENTED_SYLS_IN_CURRENT_PHRASE_BEFORE_CURRENT_SYL ,
1640     NUM_ACCENTED_SYLS_IN_CURRENT_PHRASE_AFTER_CURRENT_SYL ,
1641 ]
1642
1643 SYLLABLE_NEIGHBOURHOOD = [
1644     NUM_STRESSED_SYLS_IN_CURRENT_PHRASE_BEFORE_CURRENT_SYL ,
1645     NUM_STRESSED_SYLS_IN_CURRENT_PHRASE_AFTER_CURRENT_SYL ,
1646     NUM_STRESSED_SYLS_IN_CURRENT_PHRASE_IN_RELATION_TO_CURRENT_SYL_POSITION
        ,
1647     NUM_ACCENTED_SYLS_IN_CURRENT_PHRASE_BEFORE_CURRENT_SYL ,
1648     NUM_ACCENTED_SYLS_IN_CURRENT_PHRASE_AFTER_CURRENT_SYL ,
1649     NUM_ACCENTED_SYLS_IN_CURRENT_PHRASE_IN_RELATION_TO_CURRENT_SYL_POSITION
        ,
1650 ]
1651
1652 POSITION_OF_CURRENT_SYL_IN_RELATION_TO_NEXT_STRESSED_SYL = [
1653     NUM_SYLS_FROM_NEXT_STRESSED_SYL_EQ ,
1654     NUM_SYLS_FROM_NEXT_STRESSED_SYL_LTE ,
1655 ]
1656 POSITION_OF_CURRENT_SYL_IN_RELATION_TO_PREV_STRESSED_SYL = [
1657     NUM_SYLS_FROM_PREV_STRESSED_SYL_EQ ,
1658     NUM_SYLS_FROM_PREV_STRESSED_SYL_LTE
1659 ]
1660 POSITION_OF_CURRENT_SYL_IN_RELATION_TO_OTHER_NEIGHBOURING_STRESSED_SYLS
    = [
1661     POSITION_OF_CURRENT_SYL_IN_RELATION_TO_NEXT_STRESSED_SYL ,

```

```

1662     POSITION_OF_CURRENT_SYL_IN_RELATION_TO_PREV_STRESSED_SYL ,
1663 ]
1664 POSITION_OF_CURRENT_SYL_IN_RELATION_TO_PREV_ACCENTED_SYL = [
1665     NUM_SYLS_FROM_PREV_ACCENTED_SYL_EQ ,
1666     NUM_SYLS_FROM_PREV_ACCENTED_SYL_LTE ,
1667 ]
1668 POSITION_OF_CURRENT_SYL_IN_RELATION_TO_NEXT_ACCENTED_SYL = [
1669     NUM_SYLS_FROM_NEXT_ACCENTED_SYL_EQ ,
1670     NUM_SYLS_FROM_NEXT_ACCENTED_SYL_LTE ,
1671 ]
1672 POSITION_OF_CURRENT_SYL_IN_RELATION_TO_OTHER_NEIGHBOURING_ACCENTED_SYLS
1673     = [
1674     POSITION_OF_CURRENT_SYL_IN_RELATION_TO_PREV_ACCENTED_SYL ,
1675     POSITION_OF_CURRENT_SYL_IN_RELATION_TO_NEXT_ACCENTED_SYL
1676 ]
1677 RELATIVE_SYLLABLE_POSITION = [
1678     POSITION_OF_CURRENT_SYL_IN_RELATION_TO_NEXT_STRESSED_SYL ,
1679     POSITION_OF_CURRENT_SYL_IN_RELATION_TO_PREV_STRESSED_SYL ,
1680     POSITION_OF_CURRENT_SYL_IN_RELATION_TO_OTHER_NEIGHBOURING_STRESSED_SYLS
1681     ,
1682     POSITION_OF_CURRENT_SYL_IN_RELATION_TO_PREV_ACCENTED_SYL ,
1683     POSITION_OF_CURRENT_SYL_IN_RELATION_TO_NEXT_ACCENTED_SYL ,
1684     POSITION_OF_CURRENT_SYL_IN_RELATION_TO_OTHER_NEIGHBOURING_ACCENTED_SYLS
1685     ,
1686 ]
1687 #####
1688 # WORD
1689 #####
1690 WORD_LEN_IN_SYLS = [
1691     WORD_LEN_IN_NO_SYLS_EQ_R ,
1692     WORD_LEN_IN_NO_SYLS_EQ_C ,
1693     WORD_LEN_IN_NO_SYLS_EQ_L ,
1694     WORD_LEN_IN_NO_SYLS_LTE_R ,
1695     WORD_LEN_IN_NO_SYLS_LTE_C ,
1696     WORD_LEN_IN_NO_SYLS_LTE_L ,
1697 ]
1698 WORD_POSITION_IN_PHRASE_FW = [
1699     WORD_POSITION_IN_PHRASE_FW_EQ ,
1700     WORD_POSITION_IN_PHRASE_FW_LTE ,
1701 ]
1702 WORD_POSITION_IN_PHRASE_BW = [
1703     WORD_POSITION_IN_PHRASE_BW_EQ ,
1704     WORD_POSITION_IN_PHRASE_BW_LTE ,
1705 ]

```

```

1706 WORD_POSITION_IN_PHRASE = [
1707     WORD_POSITION_IN_PHRASE_FW,
1708     WORD_POSITION_IN_PHRASE_BW,
1709 ]
1710 WORD_POSITION = [
1711     WORD_POSITION_IN_PHRASE_FW,
1712     WORD_POSITION_IN_PHRASE_BW,
1713     WORD_POSITION_IN_PHRASE,
1714 ]
1715 NUM_CONT_WORDS_BEFORE_CURRENT_WORD_IN_PHRASE = [
1716     NUM_CONT_WORDS_BEFORE_CURRENT_WORD_IN_PHRASE_EQ,
1717     NUM_CONT_WORDS_BEFORE_CURRENT_WORD_IN_PHRASE_LTE,
1718 ]
1719 NUM_CONT_WORDS_AFTER_CURRENT_WORD_IN_PHRASE = [
1720     NUM_CONT_WORDS_AFTER_CURRENT_WORD_IN_PHRASE_EQ,
1721     NUM_CONT_WORDS_AFTER_CURRENT_WORD_IN_PHRASE_LTE,
1722 ]
1723 WORD_SURROUNDINGS = [
1724     NUM_CONT_WORDS_BEFORE_CURRENT_WORD_IN_PHRASE,
1725     NUM_CONT_WORDS_AFTER_CURRENT_WORD_IN_PHRASE,
1726 ]
1727 NUM_WORDS_FROM_PREV_CONT_WORD = [
1728     NUM_WORDS_FROM_PREV_CONT_WORD_EQ,
1729     NUM_WORDS_FROM_PREV_CONT_WORD_LTE,
1730 ]
1731 NUM_WORDS_FROM_NEXT_CONT_WORD = [
1732     NUM_WORDS_FROM_NEXT_CONT_WORD_EQ,
1733     NUM_WORDS_FROM_NEXT_CONT_WORD_LTE,
1734 ]
1735 RELATIVE_WORD_POSITION = [
1736     NUM_WORDS_FROM_PREV_CONT_WORD,
1737     NUM_WORDS_FROM_NEXT_CONT_WORD,
1738 ]
1739
1740 #####
1741 # PHRASE
1742 #####
1743
1744 PHRASE_LEN_IN_SYLS = [
1745     PHRASE_LEN_IN_NO_SYLS_EQ_L,
1746     PHRASE_LEN_IN_NO_SYLS_LTE_L,
1747     PHRASE_LEN_IN_NO_SYLS_EQ_C,
1748     PHRASE_LEN_IN_NO_SYLS_LTE_C,
1749     PHRASE_LEN_IN_NO_SYLS_EQ_R,
1750     PHRASE_LEN_IN_NO_SYLS_LTE_R,
1751 ]
1752 PHRASE_LEN_IN_WORDS = [

```

```

1753     PHRASE_LEN_IN_NO_WORDS_EQ_L ,
1754     PHRASE_LEN_IN_NO_WORDS_EQ_C ,
1755     PHRASE_LEN_IN_NO_WORDS_EQ_R ,
1756     PHRASE_LEN_IN_NO_WORDS_LTE_L ,
1757     PHRASE_LEN_IN_NO_WORDS_LTE_C ,
1758     PHRASE_LEN_IN_NO_WORDS_LTE_R ,
1759 ]
1760 PHRASE_LEN = [
1761     PHRASE_LEN_IN_SYLS ,
1762     PHRASE_LEN_IN_WORDS ,
1763 ]
1764 PHRASE_POSITION_IN_UTTERANCE_FW = [
1765     PHRASE_POSITION_IN_UTTERANCE_FW_EQ_C ,
1766     PHRASE_POSITION_IN_UTTERANCE_FW_LTE_C ,
1767 ]
1768 PHRASE_POSITION_IN_UTTERANCE_BW = [
1769     PHRASE_POSITION_IN_UTTERANCE_BW_EQ_C ,
1770     PHRASE_POSITION_IN_UTTERANCE_BW_LTE_C ,
1771 ]
1772 PHRASE_POSITION = [
1773     PHRASE_POSITION_IN_UTTERANCE_FW ,
1774     PHRASE_POSITION_IN_UTTERANCE_BW ,
1775 ]
1776
1777 #####
1778 # UTTERANCE
1779 #####
1780
1781 UTTERANCE_LEN_IN_SYLS = [
1782     UTTERANCE_LEN_IN_NO_SYLS_EQ ,
1783     UTTERANCE_LEN_IN_NO_SYLS_LTE ,
1784 ]
1785 UTTERANCE_LEN_IN_WORDS = [
1786     UTTERANCE_LEN_IN_NO_WORDS_EQ ,
1787     UTTERANCE_LEN_IN_NO_WORDS_LTE ,
1788 ]
1789 UTTERANCE_LEN_IN_PHRASES = [
1790     UTTERANCE_LEN_IN_NO_PHRASES_EQ ,
1791     UTTERANCE_LEN_IN_NO_PHRASES_LTE ,
1792 ]
1793 UTTERANCE_LEN = [
1794     UTTERANCE_LEN_IN_SYLS ,
1795     UTTERANCE_LEN_IN_WORDS ,
1796     UTTERANCE_LEN_IN_PHRASES ,
1797 ]
1798
1799 # SEGMENTAL

```

```

1800
1801 SEGMENTAL_POSITIONAL_ABSOLUTE = [
1802     NUM_SEG_IN_SYL_FW_EQ ,
1803     NUM_SEG_IN_SYL_FW_LTE ,
1804     NUM_SEG_IN_SYL_BW_EQ ,
1805     NUM_SEG_IN_SYL_BW_LTE ,
1806 ]
1807
1808 SEGMENTAL_QUALITATIVE = [
1809     QUINTPHONE_SEGMENT_VC_LL ,
1810     QUINTPHONE_SEGMENT_VC_L ,
1811     QUINTPHONE_SEGMENT_VC_C ,
1812     QUINTPHONE_SEGMENT_VC_R ,
1813     QUINTPHONE_SEGMENT_VC_RR ,
1814     QUINTPHONE_SEGMENT_ARTICULATORY_TYPE_LL ,
1815     QUINTPHONE_SEGMENT_ARTICULATORY_TYPE_L ,
1816     QUINTPHONE_SEGMENT_ARTICULATORY_TYPE_C ,
1817     QUINTPHONE_SEGMENT_ARTICULATORY_TYPE_R ,
1818     QUINTPHONE_SEGMENT_ARTICULATORY_TYPE_RR ,
1819 ]
1820
1821 # SYLLABIC
1822
1823 SYLLABIC_POSITIONAL_ABSOLUTE = [
1824     SYL_POSITION_IN_WORD_FW_EQ ,
1825     SYL_POSITION_IN_WORD_FW_LTE ,
1826     SYL_POSITION_IN_WORD_BW_EQ ,
1827     SYL_POSITION_IN_WORD_BW_LTE ,
1828     SYL_POSITION_IN_PHRASE_FW_EQ ,
1829     SYL_POSITION_IN_PHRASE_FW_LTE ,
1830     SYL_POSITION_IN_PHRASE_BW_EQ ,
1831     SYL_POSITION_IN_PHRASE_BW_LTE ,
1832 ]
1833
1834 SYLLABIC_POSITIONAL_RELATIVE = [
1835     NUM_SYLS_FROM_PREV_STRESSED_SYL_EQ ,
1836     NUM_SYLS_FROM_PREV_STRESSED_SYL_LTE ,
1837     NUM_SYLS_FROM_NEXT_STRESSED_SYL_EQ ,
1838     NUM_SYLS_FROM_NEXT_STRESSED_SYL_LTE ,
1839     NUM_SYLS_FROM_PREV_ACCENTED_SYL_EQ ,
1840     NUM_SYLS_FROM_PREV_ACCENTED_SYL_LTE ,
1841     NUM_SYLS_FROM_NEXT_ACCENTED_SYL_EQ ,
1842     NUM_SYLS_FROM_NEXT_ACCENTED_SYL_LTE ,
1843 ]
1844
1845 SYLLABIC_QUALITATIVE = [
1846     SYL_STRESSED_L ,

```



```

1847     SYL_ACCENTED_L ,
1848     SYL_ACCENT_TYPE_L ,
1849     SYL_STRESSED_C ,
1850     SYL_ACCENTED_C ,
1851     SYL_ACCENT_TYPE_C ,
1852     SYL_VOWEL_TYPE ,
1853     SYL_STRESSED_R ,
1854     SYL_ACCENTED_R ,
1855     SYL_ACCENT_TYPE_R ,
1856 ]
1857
1858 SYLLABIC_COMPOSITIONAL = [
1859     SYL_LEN_IN_NO_SEG_EQ_L ,
1860     SYL_LEN_IN_NO_SEG_LTE_L ,
1861     SYL_LEN_IN_NO_SEG_EQ_C ,
1862     SYL_LEN_IN_NO_SEG_LTE_C ,
1863     SYL_LEN_IN_NO_SEG_EQ_R ,
1864     SYL_LEN_IN_NO_SEG_LTE_R ,
1865 ]
1866
1867 SYLLABIC_PARENTAL_COMPOSITION = [
1868     NUM_STRESSED_SYLS_BEFORE_CURRENT_SYL_IN_CURRENT_PHRASE_EQ ,
1869     NUM_STRESSED_SYLS_BEFORE_CURRENT_SYL_IN_CURRENT_PHRASE_LTE ,
1870     NUM_STRESSED_SYLS_AFTER_CURRENT_SYL_IN_CURRENT_PHRASE_EQ ,
1871     NUM_STRESSED_SYLS_AFTER_CURRENT_SYL_IN_CURRENT_PHRASE_LTE ,
1872     NUM_ACCENTED_SYLS_BEFORE_CURRENT_SYL_IN_CURRENT_PHRASE_EQ ,
1873     NUM_ACCENTED_SYLS_BEFORE_CURRENT_SYL_IN_CURRENT_PHRASE_LTE ,
1874     NUM_ACCENTED_SYLS_AFTER_CURRENT_SYL_IN_CURRENT_PHRASE_EQ ,
1875     NUM_ACCENTED_SYLS_AFTER_CURRENT_SYL_IN_CURRENT_PHRASE_LTE ,
1876 ]
1877
1878 # WORD
1879
1880 WORD_POSITIONAL_ABSOLUTE = [
1881     WORD_POSITION_IN_PHRASE_FW_EQ ,
1882     WORD_POSITION_IN_PHRASE_FW_LTE ,
1883     WORD_POSITION_IN_PHRASE_BW_EQ ,
1884     WORD_POSITION_IN_PHRASE_BW_LTE ,
1885 ]
1886
1887 WORD_POSITIONAL_RELATIVE = [
1888     NUM_WORDS_FROM_PREV_CONT_WORD_EQ ,
1889     NUM_WORDS_FROM_PREV_CONT_WORD_LTE ,
1890     NUM_WORDS_FROM_NEXT_CONT_WORD_EQ ,
1891     NUM_WORDS_FROM_NEXT_CONT_WORD_LTE ,
1892 ]
1893

```

```

1894 WORD_COMPOSITIONAL = [
1895     WORD_LEN_IN_NO_SYLS_EQ_L ,
1896     WORD_LEN_IN_NO_SYLS_LTE_L ,
1897     WORD_LEN_IN_NO_SYLS_EQ_C ,
1898     WORD_LEN_IN_NO_SYLS_LTE_C ,
1899     WORD_LEN_IN_NO_SYLS_EQ_R ,
1900     WORD_LEN_IN_NO_SYLS_LTE_R ,
1901 ]
1902
1903 WORD_PARENTAL_COMPOSITION = [
1904     NUM_CONT_WORDS_BEFORE_CURRENT_WORD_IN_PHRASE_EQ ,
1905     NUM_CONT_WORDS_BEFORE_CURRENT_WORD_IN_PHRASE_LTE ,
1906     NUM_CONT_WORDS_AFTER_CURRENT_WORD_IN_PHRASE_EQ ,
1907     NUM_CONT_WORDS_AFTER_CURRENT_WORD_IN_PHRASE_LTE ,
1908 ]
1909
1910 # PHRASAL
1911
1912 PHRASAL_POSITIONAL_ABSOLUTE = [
1913     PHRASE_POSITION_IN_UTTERANCE_FW_EQ_C ,
1914     PHRASE_POSITION_IN_UTTERANCE_FW_LTE_C ,
1915     PHRASE_POSITION_IN_UTTERANCE_BW_EQ_C ,
1916     PHRASE_POSITION_IN_UTTERANCE_BW_LTE_C ,
1917 ]
1918
1919 PHRASAL_COMPOSITIONAL = [
1920     PHRASE_LEN_IN_NO_SYLS_EQ_L ,
1921     PHRASE_LEN_IN_NO_SYLS_LTE_L ,
1922     PHRASE_LEN_IN_NO_WORDS_EQ_L ,
1923     PHRASE_LEN_IN_NO_WORDS_LTE_L ,
1924     PHRASE_LEN_IN_NO_SYLS_EQ_C ,
1925     PHRASE_LEN_IN_NO_SYLS_LTE_C ,
1926     PHRASE_LEN_IN_NO_WORDS_EQ_C ,
1927     PHRASE_LEN_IN_NO_WORDS_LTE_C ,
1928     PHRASE_LEN_IN_NO_SYLS_EQ_R ,
1929     PHRASE_LEN_IN_NO_SYLS_LTE_R ,
1930     PHRASE_LEN_IN_NO_WORDS_EQ_R ,
1931     PHRASE_LEN_IN_NO_WORDS_LTE_R ,
1932 ]
1933
1934 # UTTERANCE
1935
1936 UTTERANCE_COMPOSITIONAL = [
1937     UTTERANCE_LEN_IN_NO_SYLS_EQ ,
1938     UTTERANCE_LEN_IN_NO_SYLS_LTE ,
1939     UTTERANCE_LEN_IN_NO_WORDS_EQ ,
1940     UTTERANCE_LEN_IN_NO_WORDS_LTE ,

```

```

1941     UTTERANCE_LEN_IN_NO_PHRASES_EQ ,
1942     UTTERANCE_LEN_IN_NO_PHRASES_LTE ,
1943 ]
1944
1945 POSITIONAL_ABSOLUTE = [
1946     SEGMENTAL_POSITIONAL_ABSOLUTE ,
1947     SYLLABIC_POSITIONAL_ABSOLUTE ,
1948     WORD_POSITIONAL_ABSOLUTE ,
1949     PHRASAL_POSITIONAL_ABSOLUTE ,
1950 ]
1951
1952 POSITIONAL_RELATIVE = [
1953     SYLLABIC_POSITIONAL_RELATIVE ,
1954     WORD_POSITIONAL_RELATIVE ,
1955 ]
1956
1957 QUALITATIVE = [
1958     SEGMENTAL_QUALITATIVE ,
1959     SYLLABIC_QUALITATIVE ,
1960 ]
1961
1962 COMPOSITIONAL = [
1963     SYLLABIC_COMPOSITIONAL ,
1964     WORD_COMPOSITIONAL ,
1965     PHRASAL_COMPOSITIONAL ,
1966     UTTERANCE_COMPOSITIONAL ,
1967 ]
1968
1969 PARENTAL_COMPOSITION = [
1970     SYLLABIC_PARENTAL_COMPOSITION ,
1971     WORD_PARENTAL_COMPOSITION ,
1972 ]
1973
1974 SEGMENTAL = [
1975     SEGMENTAL_POSITIONAL_ABSOLUTE ,
1976     SEGMENTAL_QUALITATIVE ,
1977 ]
1978
1979 SYLLABIC = [
1980     SYLLABIC_POSITIONAL_ABSOLUTE ,
1981     SYLLABIC_POSITIONAL_RELATIVE ,
1982     SYLLABIC_QUALITATIVE ,
1983     SYLLABIC_COMPOSITIONAL ,
1984     SYLLABIC_PARENTAL_COMPOSITION ,
1985 ]
1986
1987 WORD = [

```

```

1988     WORD_POSITIONAL_ABSOLUTE ,
1989     WORD_POSITIONAL_RELATIVE ,
1990     WORD_COMPOSITIONAL ,
1991     WORD_PARENTAL_COMPOSITION ,
1992 ]
1993
1994 PHRASAL = [
1995     PHRASAL_POSITIONAL_ABSOLUTE ,
1996     PHRASAL_COMPOSITIONAL ,
1997 ]
1998
1999 UTTERANCE = [
2000     UTTERANCE_COMPOSITIONAL ,
2001 ]
2002
2003 #####
2004 # TARGET FEATURE GROUPS
2005 #####
2006 FEATURE_GROUPS = {
2007     'VUV': VUV ,
2008     'QUINTPHONE_SEGMENT_VC': QUINTPHONE_SEGMENT_VC ,
2009     'QUINTPHONE_SEGMENT_ARTICULATORY_TYPE':
2010         QUINTPHONE_SEGMENT_ARTICULATORY_TYPE ,
2011     'SEG_POS_IN_SYL': SEG_POS_IN_SYL ,
2012     'QUINTPHONE_SEGMENTAL_FEATURES': QUINTPHONE_SEGMENTAL_FEATURES ,
2013     'SYLLABLE_STRESS': SYLLABLE_STRESS ,
2014     'SYLLABLE_ACCENTED': SYLLABLE_ACCENTED ,
2015     'SYLLABLE_ACCENT_TYPE': SYLLABLE_ACCENT_TYPE ,
2016     'SYLLABLE_LEN_IN_SEGMENTS': SYLLABLE_LEN_IN_SEGMENTS ,
2017     'SYLLABLE_NUCLEUS': SYLLABLE_NUCLEUS ,
2018     'SYLLABLE_POSITION': SYLLABLE_POSITION ,
2019     'RELATIVE_SYLLABLE_POSITION': RELATIVE_SYLLABLE_POSITION ,
2020     'SYLLABLE_NEIGHBOURHOOD': SYLLABLE_NEIGHBOURHOOD ,
2021     'WORD_LEN_IN_SYLS': WORD_LEN_IN_SYLS ,
2022     'WORD_POSITION': WORD_POSITION ,
2023     'WORD_SURROUNDINGS': WORD_SURROUNDINGS ,
2024     'RELATIVE_WORD_POSITION': RELATIVE_WORD_POSITION ,
2025     'PHRASE_LEN': PHRASE_LEN ,
2026     'PHRASE_POSITION': PHRASE_POSITION ,
2027     'UTTERANCE_LEN': UTTERANCE_LEN ,
2028 }
2029
2030 DETAILED_GROUPS = {
2031     'VUV': VUV ,
2032     'SYL_VOWEL_TYPE': SYL_VOWEL_TYPE ,
2033     'QUINTPHONE_SEGMENT_VC': QUINTPHONE_SEGMENT_VC ,

```

```

2033 'QUINTPHONE_SEGMENT_ARTICULATORY_TYPE':
      QUINTPHONE_SEGMENT_ARTICULATORY_TYPE,
2034 'SEG_POS_IN_SYL': SEG_POS_IN_SYL,
2035 'QUINTPHONE_SEGMENTAL_FEATURES': QUINTPHONE_SEGMENTAL_FEATURES,
2036 'SYLLABLE_STRESS': SYLLABLE_STRESS,
2037 'SYLLABLE_ACCENTED': SYLLABLE_ACCENTED,
2038 'SYLLABLE_ACCENT_TYPE': SYLLABLE_ACCENT_TYPE,
2039 'SYLLABLE_LEN_IN_SEGMENTS': SYLLABLE_LEN_IN_SEGMENTS,
2040 'SYLLABLE_NUCLEUS': SYLLABLE_NUCLEUS,
2041 'SYLLABLE_POSITION_IN_WORD_FW': SYLLABLE_POSITION_IN_WORD_FW,
2042 'SYLLABLE_POSITION_IN_WORD_BW': SYLLABLE_POSITION_IN_WORD_BW,
2043 'SYLLABLE_POSITION_IN_PHRASE_FW':
      SYLLABLE_POSITION_IN_PHRASE_FW,
2044 'SYLLABLE_POSITION_IN_PHRASE_BW':
      SYLLABLE_POSITION_IN_PHRASE_BW,
2045 'SYLLABLE_POSITION_IN_WORD': SYLLABLE_POSITION_IN_WORD,
2046 'SYLLABLE_POSITION_IN_PHRASE': SYLLABLE_POSITION_IN_PHRASE,
2047 'NUM_STRESSED_SYLS_IN_CURRENT_PHRASE_BEFORE_CURRENT_SYL':
      NUM_STRESSED_SYLS_IN_CURRENT_PHRASE_BEFORE_CURRENT_SYL,
2048 'NUM_STRESSED_SYLS_IN_CURRENT_PHRASE_AFTER_CURRENT_SYL':
      NUM_STRESSED_SYLS_IN_CURRENT_PHRASE_AFTER_CURRENT_SYL,
2049 '
      NUM_STRESSED_SYLS_IN_CURRENT_PHRASE_IN_RELATION_TO_CURRENT_SYL_POSITION
      ':
      NUM_STRESSED_SYLS_IN_CURRENT_PHRASE_IN_RELATION_TO_CURRENT_SYL_POSITION
      ',
2050 'NUM_ACCENTED_SYLS_IN_CURRENT_PHRASE_BEFORE_CURRENT_SYL':
      NUM_ACCENTED_SYLS_IN_CURRENT_PHRASE_BEFORE_CURRENT_SYL,
2051 'NUM_ACCENTED_SYLS_IN_CURRENT_PHRASE_AFTER_CURRENT_SYL':
      NUM_ACCENTED_SYLS_IN_CURRENT_PHRASE_AFTER_CURRENT_SYL,
2052 '
      NUM_ACCENTED_SYLS_IN_CURRENT_PHRASE_IN_RELATION_TO_CURRENT_SYL_POSITION
      ':
      NUM_ACCENTED_SYLS_IN_CURRENT_PHRASE_IN_RELATION_TO_CURRENT_SYL_POSITION
      ',
2053 'POSITION_OF_CURRENT_SYL_IN_RELATION_TO_NEXT_STRESSED_SYL':
      POSITION_OF_CURRENT_SYL_IN_RELATION_TO_NEXT_STRESSED_SYL,
2054 'POSITION_OF_CURRENT_SYL_IN_RELATION_TO_PREV_STRESSED_SYL':
      POSITION_OF_CURRENT_SYL_IN_RELATION_TO_PREV_STRESSED_SYL,
2055 '
      POSITION_OF_CURRENT_SYL_IN_RELATION_TO_OTHER_NEIGHBOURING_STRESSED_SYLS
      ':
      POSITION_OF_CURRENT_SYL_IN_RELATION_TO_OTHER_NEIGHBOURING_STRESSED_SYLS
      ',
2056 'POSITION_OF_CURRENT_SYL_IN_RELATION_TO_PREV_ACCENTED_SYL':
      POSITION_OF_CURRENT_SYL_IN_RELATION_TO_PREV_ACCENTED_SYL,

```

```

2057     'POSITION_OF_CURRENT_SYL_IN_RELATION_TO_NEXT_ACCENTED_SYL':
2058         POSITION_OF_CURRENT_SYL_IN_RELATION_TO_NEXT_ACCENTED_SYL,
2059     'POSITION_OF_CURRENT_SYL_IN_RELATION_TO_OTHER_NEIGHBOURING_ACCENTED_SYLS':
2060         POSITION_OF_CURRENT_SYL_IN_RELATION_TO_OTHER_NEIGHBOURING_ACCENTED_SYLS,
2061     'WORD_LEN_IN_SYLS': WORD_LEN_IN_SYLS,
2062     'WORD_POSITION_IN_PHRASE_FW': WORD_POSITION_IN_PHRASE_FW,
2063     'WORD_POSITION_IN_PHRASE_BW': WORD_POSITION_IN_PHRASE_BW,
2064     'WORD_POSITION_IN_PHRASE': WORD_POSITION_IN_PHRASE,
2065     'NUM_CONT_WORDS_BEFORE_CURRENT_WORD_IN_PHRASE':
2066         NUM_CONT_WORDS_BEFORE_CURRENT_WORD_IN_PHRASE,
2067     'NUM_CONT_WORDS_AFTER_CURRENT_WORD_IN_PHRASE':
2068         NUM_CONT_WORDS_AFTER_CURRENT_WORD_IN_PHRASE,
2069     'NUM_WORDS_FROM_PREV_CONT_WORD': NUM_WORDS_FROM_PREV_CONT_WORD,
2070     'NUM_WORDS_FROM_NEXT_CONT_WORD': NUM_WORDS_FROM_NEXT_CONT_WORD,
2071     'PHRASE_LEN_IN_SYLS': PHRASE_LEN_IN_SYLS,
2072     'PHRASE_LEN_IN_WORDS': PHRASE_LEN_IN_WORDS,
2073     'PHRASE_POSITION_IN_UTTERANCE_FW':
2074         PHRASE_POSITION_IN_UTTERANCE_FW,
2075     'PHRASE_POSITION_IN_UTTERANCE_BW':
2076         PHRASE_POSITION_IN_UTTERANCE_BW,
2077     'UTTERANCE_LEN_IN_SYLS': UTTERANCE_LEN_IN_SYLS,
2078     'UTTERANCE_LEN_IN_WORDS': UTTERANCE_LEN_IN_WORDS,
2079     'UTTERANCE_LEN_IN_PHRASES': UTTERANCE_LEN_IN_PHRASES,
2080 }
2081
2082 ALL_GROUPS = {
2083     'VUV': VUV,
2084     'QUINTPHONE_SEGMENT_VC_LL': QUINTPHONE_SEGMENT_VC_LL,
2085     'QUINTPHONE_SEGMENT_VC_L': QUINTPHONE_SEGMENT_VC_L,
2086     'QUINTPHONE_SEGMENT_VC_C': QUINTPHONE_SEGMENT_VC_C,
2087     'QUINTPHONE_SEGMENT_VC_R': QUINTPHONE_SEGMENT_VC_R,
2088     'QUINTPHONE_SEGMENT_VC_RR': QUINTPHONE_SEGMENT_VC_RR,
2089     'QUINTPHONE_SEGMENT_ARTICULATORY_TYPE_LL':
2090         QUINTPHONE_SEGMENT_ARTICULATORY_TYPE_LL,
2091     'QUINTPHONE_SEGMENT_ARTICULATORY_TYPE_L':
2092         QUINTPHONE_SEGMENT_ARTICULATORY_TYPE_L,
2093     'QUINTPHONE_SEGMENT_ARTICULATORY_TYPE_C':
2094         QUINTPHONE_SEGMENT_ARTICULATORY_TYPE_C,
2095     'QUINTPHONE_SEGMENT_ARTICULATORY_TYPE_R':
2096         QUINTPHONE_SEGMENT_ARTICULATORY_TYPE_R,
2097     'QUINTPHONE_SEGMENT_ARTICULATORY_TYPE_RR':
2098         QUINTPHONE_SEGMENT_ARTICULATORY_TYPE_RR,
2099     'NUM_SEG_IN_SYL_FW_EQ': NUM_SEG_IN_SYL_FW_EQ,
2100     'NUM_SEG_IN_SYL_FW_LTE': NUM_SEG_IN_SYL_FW_LTE,

```

2090 'NUM_SEG_IN_SYL_BW_EQ': NUM_SEG_IN_SYL_BW_EQ,
 2091 'NUM_SEG_IN_SYL_BW_LTE': NUM_SEG_IN_SYL_BW_LTE,
 2092 'SYL_STRESSED_L': SYL_STRESSED_L,
 2093 'SYL_ACCENTED_L': SYL_ACCENTED_L,
 2094 'SYL_ACCENT_TYPE_L': SYL_ACCENT_TYPE_L,
 2095 'SYL_LEN_IN_NO_SEG_EQ_L': SYL_LEN_IN_NO_SEG_EQ_L,
 2096 'SYL_LEN_IN_NO_SEG_LTE_L': SYL_LEN_IN_NO_SEG_LTE_L,
 2097 'SYL_STRESSED_C': SYL_STRESSED_C,
 2098 'SYL_ACCENTED_C': SYL_ACCENTED_C,
 2099 'SYL_ACCENT_TYPE_C': SYL_ACCENT_TYPE_C,
 2100 'SYL_LEN_IN_NO_SEG_EQ_C': SYL_LEN_IN_NO_SEG_EQ_C,
 2101 'SYL_LEN_IN_NO_SEG_LTE_C': SYL_LEN_IN_NO_SEG_LTE_C,
 2102 'SYL_POSITION_IN_WORD_FW_EQ': SYL_POSITION_IN_WORD_FW_EQ,
 2103 'SYL_POSITION_IN_WORD_FW_LTE': SYL_POSITION_IN_WORD_FW_LTE,
 2104 'SYL_POSITION_IN_WORD_BW_EQ': SYL_POSITION_IN_WORD_BW_EQ,
 2105 'SYL_POSITION_IN_WORD_BW_LTE': SYL_POSITION_IN_WORD_BW_LTE,
 2106 'SYL_POSITION_IN_PHRASE_FW_EQ': SYL_POSITION_IN_PHRASE_FW_EQ,
 2107 'SYL_POSITION_IN_PHRASE_FW_LTE': SYL_POSITION_IN_PHRASE_FW_LTE,
 2108 'SYL_POSITION_IN_PHRASE_BW_EQ': SYL_POSITION_IN_PHRASE_BW_EQ,
 2109 'SYL_POSITION_IN_PHRASE_BW_LTE': SYL_POSITION_IN_PHRASE_BW_LTE,
 2110 'NUM_STRESSED_SYLS_BEFORE_CURRENT_SYL_IN_CURRENT_PHRASE_EQ':
 NUM_STRESSED_SYLS_BEFORE_CURRENT_SYL_IN_CURRENT_PHRASE_EQ,
 2111 'NUM_STRESSED_SYLS_BEFORE_CURRENT_SYL_IN_CURRENT_PHRASE_LTE':
 NUM_STRESSED_SYLS_BEFORE_CURRENT_SYL_IN_CURRENT_PHRASE_LTE,
 2112 'NUM_STRESSED_SYLS_AFTER_CURRENT_SYL_IN_CURRENT_PHRASE_EQ':
 NUM_STRESSED_SYLS_AFTER_CURRENT_SYL_IN_CURRENT_PHRASE_EQ,
 2113 'NUM_STRESSED_SYLS_AFTER_CURRENT_SYL_IN_CURRENT_PHRASE_LTE':
 NUM_STRESSED_SYLS_AFTER_CURRENT_SYL_IN_CURRENT_PHRASE_LTE,
 2114 'NUM_ACCENTED_SYLS_BEFORE_CURRENT_SYL_IN_CURRENT_PHRASE_EQ':
 NUM_ACCENTED_SYLS_BEFORE_CURRENT_SYL_IN_CURRENT_PHRASE_EQ,
 2115 'NUM_ACCENTED_SYLS_BEFORE_CURRENT_SYL_IN_CURRENT_PHRASE_LTE':
 NUM_ACCENTED_SYLS_BEFORE_CURRENT_SYL_IN_CURRENT_PHRASE_LTE,
 2116 'NUM_ACCENTED_SYLS_AFTER_CURRENT_SYL_IN_CURRENT_PHRASE_EQ':
 NUM_ACCENTED_SYLS_AFTER_CURRENT_SYL_IN_CURRENT_PHRASE_EQ,
 2117 'NUM_ACCENTED_SYLS_AFTER_CURRENT_SYL_IN_CURRENT_PHRASE_LTE':
 NUM_ACCENTED_SYLS_AFTER_CURRENT_SYL_IN_CURRENT_PHRASE_LTE,
 2118 'NUM_SYLS_FROM_PREV_STRESSED_SYL_EQ':
 NUM_SYLS_FROM_PREV_STRESSED_SYL_EQ,
 2119 'NUM_SYLS_FROM_PREV_STRESSED_SYL_LTE':
 NUM_SYLS_FROM_PREV_STRESSED_SYL_LTE,
 2120 'NUM_SYLS_FROM_NEXT_STRESSED_SYL_EQ':
 NUM_SYLS_FROM_NEXT_STRESSED_SYL_EQ,
 2121 'NUM_SYLS_FROM_NEXT_STRESSED_SYL_LTE':
 NUM_SYLS_FROM_NEXT_STRESSED_SYL_LTE,
 2122 'NUM_SYLS_FROM_PREV_ACCENTED_SYL_EQ':
 NUM_SYLS_FROM_PREV_ACCENTED_SYL_EQ,

```

2123 'NUM_SYLS_FROM_PREV_ACCENTED_SYL_LTE':
      NUM_SYLS_FROM_PREV_ACCENTED_SYL_LTE,
2124 'NUM_SYLS_FROM_NEXT_ACCENTED_SYL_EQ':
      NUM_SYLS_FROM_NEXT_ACCENTED_SYL_EQ,
2125 'NUM_SYLS_FROM_NEXT_ACCENTED_SYL_LTE':
      NUM_SYLS_FROM_NEXT_ACCENTED_SYL_LTE,
2126 'SYL_VOWEL_TYPE': SYL_VOWEL_TYPE,
2127 'SYL_STRESSED_R': SYL_STRESSED_R,
2128 'SYL_ACCENTED_R': SYL_ACCENTED_R,
2129 'SYL_ACCENT_TYPE_R': SYL_ACCENT_TYPE_R,
2130 'SYL_LEN_IN_NO_SEG_EQ_R': SYL_LEN_IN_NO_SEG_EQ_R,
2131 'SYL_LEN_IN_NO_SEG_LTE_R': SYL_LEN_IN_NO_SEG_LTE_R,
2132 'WORD_LEN_IN_NO_SYLS_EQ_L': WORD_LEN_IN_NO_SYLS_EQ_L,
2133 'WORD_LEN_IN_NO_SYLS_LTE_L': WORD_LEN_IN_NO_SYLS_LTE_L,
2134 'WORD_LEN_IN_NO_SYLS_EQ_C': WORD_LEN_IN_NO_SYLS_EQ_C,
2135 'WORD_LEN_IN_NO_SYLS_LTE_C': WORD_LEN_IN_NO_SYLS_LTE_C,
2136 'WORD_POSITION_IN_PHRASE_FW_EQ': WORD_POSITION_IN_PHRASE_FW_EQ,
2137 'WORD_POSITION_IN_PHRASE_FW_LTE':
      WORD_POSITION_IN_PHRASE_FW_LTE,
2138 'WORD_POSITION_IN_PHRASE_BW_EQ': WORD_POSITION_IN_PHRASE_BW_EQ,
2139 'WORD_POSITION_IN_PHRASE_BW_LTE':
      WORD_POSITION_IN_PHRASE_BW_LTE,
2140 'NUM_CONT_WORDS_BEFORE_CURRENT_WORD_IN_PHRASE_EQ':
      NUM_CONT_WORDS_BEFORE_CURRENT_WORD_IN_PHRASE_EQ,
2141 'NUM_CONT_WORDS_BEFORE_CURRENT_WORD_IN_PHRASE_LTE':
      NUM_CONT_WORDS_BEFORE_CURRENT_WORD_IN_PHRASE_LTE,
2142 'NUM_CONT_WORDS_AFTER_CURRENT_WORD_IN_PHRASE_EQ':
      NUM_CONT_WORDS_AFTER_CURRENT_WORD_IN_PHRASE_EQ,
2143 'NUM_CONT_WORDS_AFTER_CURRENT_WORD_IN_PHRASE_LTE':
      NUM_CONT_WORDS_AFTER_CURRENT_WORD_IN_PHRASE_LTE,
2144 'NUM_WORDS_FROM_PREV_CONT_WORD_EQ':
      NUM_WORDS_FROM_PREV_CONT_WORD_EQ,
2145 'NUM_WORDS_FROM_PREV_CONT_WORD_LTE':
      NUM_WORDS_FROM_PREV_CONT_WORD_LTE,
2146 'NUM_WORDS_FROM_NEXT_CONT_WORD_EQ':
      NUM_WORDS_FROM_NEXT_CONT_WORD_EQ,
2147 'NUM_WORDS_FROM_NEXT_CONT_WORD_LTE':
      NUM_WORDS_FROM_NEXT_CONT_WORD_LTE,
2148 'WORD_LEN_IN_NO_SYLS_EQ_R': WORD_LEN_IN_NO_SYLS_EQ_R,
2149 'WORD_LEN_IN_NO_SYLS_LTE_R': WORD_LEN_IN_NO_SYLS_LTE_R,
2150 'PHRASE_LEN_IN_NO_SYLS_EQ_L': PHRASE_LEN_IN_NO_SYLS_EQ_L,
2151 'PHRASE_LEN_IN_NO_SYLS_LTE_L': PHRASE_LEN_IN_NO_SYLS_LTE_L,
2152 'PHRASE_LEN_IN_NO_WORDS_EQ_L': PHRASE_LEN_IN_NO_WORDS_EQ_L,
2153 'PHRASE_LEN_IN_NO_WORDS_LTE_L': PHRASE_LEN_IN_NO_WORDS_LTE_L,
2154 'PHRASE_LEN_IN_NO_SYLS_EQ_C': PHRASE_LEN_IN_NO_SYLS_EQ_C,
2155 'PHRASE_LEN_IN_NO_SYLS_LTE_C': PHRASE_LEN_IN_NO_SYLS_LTE_C,
2156 'PHRASE_LEN_IN_NO_WORDS_EQ_C': PHRASE_LEN_IN_NO_WORDS_EQ_C,

```



```

2157     'PHRASE_LEN_IN_NO_WORDS_LTE_C': PHRASE_LEN_IN_NO_WORDS_LTE_C,
2158     'PHRASE_POSITION_IN_UTTERANCE_FW_EQ_C':
2159         PHRASE_POSITION_IN_UTTERANCE_FW_EQ_C,
2160     'PHRASE_POSITION_IN_UTTERANCE_FW_LTE_C':
2161         PHRASE_POSITION_IN_UTTERANCE_FW_LTE_C,
2162     'PHRASE_POSITION_IN_UTTERANCE_BW_EQ_C':
2163         PHRASE_POSITION_IN_UTTERANCE_BW_EQ_C,
2164     'PHRASE_POSITION_IN_UTTERANCE_BW_LTE_C':
2165         PHRASE_POSITION_IN_UTTERANCE_BW_LTE_C,
2166     'PHRASE_LEN_IN_NO_SYLS_EQ_R': PHRASE_LEN_IN_NO_SYLS_EQ_R,
2167     'PHRASE_LEN_IN_NO_SYLS_LTE_R': PHRASE_LEN_IN_NO_SYLS_LTE_R,
2168     'PHRASE_LEN_IN_NO_WORDS_EQ_R': PHRASE_LEN_IN_NO_WORDS_EQ_R,
2169     'PHRASE_LEN_IN_NO_WORDS_LTE_R': PHRASE_LEN_IN_NO_WORDS_LTE_R,
2170     'UTTERANCE_LEN_IN_NO_SYLS_EQ': UTTERANCE_LEN_IN_NO_SYLS_EQ,
2171     'UTTERANCE_LEN_IN_NO_SYLS_LTE': UTTERANCE_LEN_IN_NO_SYLS_LTE,
2172     'UTTERANCE_LEN_IN_NO_WORDS_EQ': UTTERANCE_LEN_IN_NO_WORDS_EQ,
2173     'UTTERANCE_LEN_IN_NO_WORDS_LTE': UTTERANCE_LEN_IN_NO_WORDS_LTE,
2174     'UTTERANCE_LEN_IN_NO_PHRASES_EQ':
2175         UTTERANCE_LEN_IN_NO_PHRASES_EQ,
2176     'UTTERANCE_LEN_IN_NO_PHRASES_LTE':
2177         UTTERANCE_LEN_IN_NO_PHRASES_LTE,
2178 }
2179
2180 SEGMENTAL_GROUPS = {
2181     'SEGMENTAL_POSITIONAL_ABSOLUTE': SEGMENTAL_POSITIONAL_ABSOLUTE,
2182     'SEGMENTAL_QUALITATIVE': SEGMENTAL_QUALITATIVE,
2183 }
2184
2185 SYLLABIC_GROUPS = {
2186     'SYLLABIC_POSITIONAL_ABSOLUTE': SYLLABIC_POSITIONAL_ABSOLUTE,
2187     'SYLLABIC_POSITIONAL_RELATIVE': SYLLABIC_POSITIONAL_RELATIVE,
2188     'SYLLABIC_QUALITATIVE': SYLLABIC_QUALITATIVE,
2189     'SYLLABIC_COMPOSITIONAL': SYLLABIC_COMPOSITIONAL,
2190     'SYLLABIC_PARENTAL_COMPOSITION': SYLLABIC_PARENTAL_COMPOSITION,
2191 }
2192
2193 WORD_GROUPS = {
2194     'WORD_POSITIONAL_ABSOLUTE': WORD_POSITIONAL_ABSOLUTE,
2195     'WORD_POSITIONAL_RELATIVE': WORD_POSITIONAL_RELATIVE,
2196     'WORD_COMPOSITIONAL': WORD_COMPOSITIONAL,
2197     'WORD_PARENTAL_COMPOSITION': WORD_PARENTAL_COMPOSITION,
2198 }
2199
2200 PHRASAL_GROUPS = {
2201     'PHRASAL_POSITIONAL_ABSOLUTE': PHRASAL_POSITIONAL_ABSOLUTE,
2202     'PHRASAL_COMPOSITIONAL': PHRASAL_COMPOSITIONAL,
2203 }

```

```

2198
2199 POSITIONAL_ABSOLUTE_GROUPS = {
2200     'SEGMENTAL_POSITIONAL_ABSOLUTE': SEGMENTAL_POSITIONAL_ABSOLUTE,
2201     'SYLLABIC_POSITIONAL_ABSOLUTE': SYLLABIC_POSITIONAL_ABSOLUTE,
2202     'WORD_POSITIONAL_ABSOLUTE': WORD_POSITIONAL_ABSOLUTE,
2203     'PHRASAL_POSITIONAL_ABSOLUTE': PHRASAL_POSITIONAL_ABSOLUTE,
2204 }
2205
2206 POSITIONAL_RELATIVE_GROUPS = {
2207     'SYLLABIC_POSITIONAL_RELATIVE': SYLLABIC_POSITIONAL_RELATIVE,
2208     'WORD_POSITIONAL_RELATIVE': WORD_POSITIONAL_RELATIVE,
2209 }
2210
2211 QUALITATIVE_GROUPS = {
2212     'SEGMENTAL_QUALITATIVE': SEGMENTAL_QUALITATIVE,
2213     'SYLLABIC_QUALITATIVE': SYLLABIC_QUALITATIVE,
2214 }
2215
2216 COMPOSITIONAL_GROUPS = {
2217     'SYLLABIC_COMPOSITIONAL': SYLLABIC_COMPOSITIONAL,
2218     'WORD_COMPOSITIONAL': WORD_COMPOSITIONAL,
2219     'PHRASAL_COMPOSITIONAL': PHRASAL_COMPOSITIONAL,
2220     'UTTERANCE_COMPOSITIONAL': UTTERANCE_COMPOSITIONAL,
2221 }
2222
2223 PARENTAL_COMPOSITION_GROUPS = {
2224     'SYLLABIC_PARENTAL_COMPOSITION': SYLLABIC_PARENTAL_COMPOSITION,
2225     'WORD_PARENTAL_COMPOSITION': WORD_PARENTAL_COMPOSITION,
2226 }
2227
2228 LINGUISTIC_LEVEL_GROUPS = {
2229     'SEGMENTAL': SEGMENTAL,
2230     'SYLLABIC': SYLLABIC,
2231     'WORD': WORD,
2232     'PHRASAL': PHRASAL,
2233     'UTTERANCE': UTTERANCE
2234 }
2235
2236 FEATURE_TYPE_GROUPS = {
2237     'POSITIONAL_ABSOLUTE': POSITIONAL_ABSOLUTE,
2238     'POSITIONAL_RELATIVE': POSITIONAL_RELATIVE,
2239     'QUALITATIVE': QUALITATIVE,
2240     'COMPOSITIONAL': COMPOSITIONAL,
2241     'PARENTAL_COMPOSITION': PARENTAL_COMPOSITION,
2242 }
2243
2244 LINGUISTIC_LEVEL_WITH_FEATURE_TYPE_GROUPS = {

```

```

2245     'SEGMENTAL_POSITIONAL_ABSOLUTE': SEGMENTAL_POSITIONAL_ABSOLUTE,
2246     'SYLLABIC_POSITIONAL_ABSOLUTE': SYLLABIC_POSITIONAL_ABSOLUTE,
2247     'WORD_POSITIONAL_ABSOLUTE': WORD_POSITIONAL_ABSOLUTE,
2248     'PHRASAL_POSITIONAL_ABSOLUTE': PHRASAL_POSITIONAL_ABSOLUTE,
2249     'SYLLABIC_POSITIONAL_RELATIVE': SYLLABIC_POSITIONAL_RELATIVE,
2250     'WORD_POSITIONAL_RELATIVE': WORD_POSITIONAL_RELATIVE,
2251     'SEGMENTAL_QUALITATIVE': SEGMENTAL_QUALITATIVE,
2252     'SYLLABIC_QUALITATIVE': SYLLABIC_QUALITATIVE,
2253     'SYLLABIC_COMPOSITIONAL': SYLLABIC_COMPOSITIONAL,
2254     'WORD_COMPOSITIONAL': WORD_COMPOSITIONAL,
2255     'PHRASAL_COMPOSITIONAL': PHRASAL_COMPOSITIONAL,
2256     'UTTERANCE_COMPOSITIONAL': UTTERANCE_COMPOSITIONAL,
2257     'SYLLABIC_PARENTAL_COMPOSITION': SYLLABIC_PARENTAL_COMPOSITION,
2258     'WORD_PARENTAL_COMPOSITION': WORD_PARENTAL_COMPOSITION,
2259 }

```

Listing B.1: Structure of features groups as defined in the Python source code of the current work.

Colophon

This thesis was typeset with $\text{\LaTeX}$ 2_ε. It uses the *Clean Thesis* style developed by Ricardo Langner. The design of the *Clean Thesis* style is inspired by user guide documents from Apple Inc.

Download the *Clean Thesis* style at <http://cleanthesis.der-ric.de/>.

