INFORMATICA, 2022, Vol. 33, No. 4, 795–832

DOI: https://doi.org/10.15388/22-INFOR502

An Overview of Lithuanian Intonation: A Linguistic and Modelling Perspective

Gerda Ana MELNIK-LEROY*, Jolita BERNATAVIČIENĖ,
Gražina KORVEL, Gediminas NAVICKAS,
Gintautas TAMULEVIČIUS, Povilas TREIGYS

Institute of Data Science and Digital Technologies, Vilnius University,
Akademijos str. 4, LT-08412 Vilnius, Lithuania
e-mail: gerda.melnik@mif.vu.lt

Received: May 2022; accepted: November 2022

Abstract. Intonation is a complex suprasegmental phenomenon essential for speech processing. However, it is still largely understudied, especially in the case of under-resourced languages, such as Lithuanian. The current paper focuses on intonation in Lithuanian, a Baltic pitch-accent language with free stress and tonal variations on accented heavy syllables. Due to historical circumstances, the description and analysis of Lithuanian intonation were carried out within different theoretical frameworks and in several languages, which makes them hardly accessible to the international research community. This paper is the first attempt to gather research on Lithuanian intonation from both the Lithuanian and the Western traditions, the structuralist and generativist points of view, and the linguistic and modelling perspectives. The paper identifies issues in existing research that require special attention and proposes directions for future investigations both in linguistics and modelling.

Key words: intonation, stress, Pitch accent, intonation modelling, speech recognition, Lithuanian, under-resourced language.

1. Introduction

Alongside word stress and rhythm, intonation is an essential element of linguistic prosody. Intonation has several functions, such as highlighting grammatical structure, conveying emotion and focusing attention on relevant portions of the spoken message. Importantly, ample evidence from psycholinguistic studies showed that intonation is used to identify phrasal boundaries (prosodic phrases). Hence, intonation plays a crucial role in parsing continuous speech both in infants during early language acquisition (Jusczyk *et al.*, 1992) and in adults during speech processing (Cutler *et al.*, 1997; Féry, 2017). Intonation is a complex suprasegmental linguistic phenomenon containing both universal and language-specific characteristics (Ladd, 2001). The latter depend on the phonetic and phonological features of a given language, such as the type of lexical stress or tone, the peculiarities of vowel reduction processes, the syllable structure and rhythm, etc. Although most classical

*Corresponding author.

prosodic typologies are solely based on word-level features, such as the famous three-way distinction between "stress, tone and pitch-accent" languages, growing cross-linguistic evidence points to the fact that sentence-level intonational features are not entirely predictable from word-prosodic systems (Jun, 2005; Hamlaoui *et al.*, 2019). For this reason, there is an ongoing discussion on the necessity to integrate more aspects of sentence-level intonation (Zerbian, 2010) or on the creation of an intonational typology altogether (Ladd, 2001, 2008; Arvaniti, 2016). Finally, although intonation has been studied in linguistics and psycholinguistics for decades (Bolinger, 1964; Lieberman, 1965; Selkirk, 1978; Hirst and Di Cristo, 1998; Hallé *et al.*, 1991; Snow and Balog, 2002), this research mostly examined only a few languages (e.g. English, Swedish, Japanese or French) and rarely compared intonational systems cross-linguistically (Zerbian, 2010).

In recent years, intonation has attracted attention due to the rise of modern speech technologies. It has been confirmed that the incorporation of intonational information improves keyword extraction performance (Lezhenin *et al.*, 2018). Also, intonation is useful in the identification of a speaker's dialect (Alvarez and Issa, 2020; Biadsy and Hirschberg, 2009). Turning to speech synthesis, although contemporary Text-To-Speech (TTS) systems can produce high-quality speech, they lack intonation and other prosodic features, thus not only degrading its naturalness, but also the ease of processing (Ronanki *et al.*, 2016; Hodari *et al.*, 2021). Creating formal models of intonation proved to be important in order to develop effective synthetic speech methods (de Pijper, 1983; Syrdal *et al.*, 1998; Suni *et al.*, 2013; Reddy and Rao, 2011; Ronanki *et al.*, 2016). As modelling a linguistic phenomenon allows to determine a generalized and formalized structure in the language (Chomsky, 1956) and sheds light on the cognitive mechanism of language as a whole (Yngve, 1954), intonation modelling has the potential to enhance linguistic theory. Overall, modelling the prosody and intonation of a language is and will be a crucial issue in language technology. Only by being able to model (and synthesize) the prosody (intonation) of human language we will we be able to ensure that the speech signal we synthesize sounds natural, with emotion, properly accentuated, and with levels of varying significance. Synthesized language with these properties can be helpful in various natural language-based scenarios (e.g. in interaction scenarios, dialogue scenarios, education scenarios etc.), which need to convey intention and attitude, grab the listener's attention, emphasize certain information, and improve the user's experience. A separate need is dictated by tonal languages, in which intonation is an inherent part of the lexical information (Janyoi and Seresangtakul, 2020).

Solutions based on artificial neural networks, which are now widely used in speech signal synthesis, are essentially an aggregative approach – all knowledge of the generated speech signal is obtained by examining many language examples that might strongly differ in the contextual information provided in them. Thus, the speech signal generated by such a synthesizer will be mainly focused on the phonetic lexical content, not the suprasegmental properties of the language (including intonation). Moreover, as prosodic information varies at a slower rate in comparison to other content in the acoustic signal (e.g. segmental information and background noise), modelling intonation from conventional speech datasets is problematic (Hodari *et al.*, 2021). In addition to this, such an aggregating solution does not provide the control of intonation (where one could fully control the process of assigning or changing intonation), nor the opportunity to understand and interpret the model (ideally, one would want to know which parameters to change in order to implement a specific change in intonation). Therefore, the aim has recently shifted to modelling intonation and prosody as a whole with the perspective to gain control over it. Various concepts of intonation modelling can be found. Some researchers have tried to apply ideas of classical intonation models by predicting their meanings (Chakrasali *et al.*, 2022; Kuczmarski, 2021; Marelli *et al.*, 2019) or identifying intonation segments (Alvarez *et al.*, 2022). Other intonation modelling instances include the transfer of intonation components into neutral synthesized speech (Honnet and Garner, 2016), unsupervised or supervised training of latent prosody space (Sun *et al.*, 2020; Raitio *et al.*, 2020, 2022). In Birkholz and Zhang (2020), microprosody-based intonation modelling is presented as an additional way of improving the naturalness of synthesized speech.

It is important to note that models of intonation have been mostly developed for a few most common languages, such as English (Ronanki *et al.*, 2016; Hodari *et al.*, 2019) or Chinese (Tseng *et al.*, 2005; Liu, 2017). However, these models are not directly applicable to other languages, as intonation involves both universal and language-specific components (Hirst and Di Cristo, 1998; Grabe and Karpinski, 2003). Although applications of intonation models to TTS for other languages have been scarce, over the past decade, an increase in the number of papers in this field is noticeable (Swedish and its dialects: Bruce *et al.*, 2007; Hindi, Telugu and Tamil: Rao and Yegnanarayana, 2009; Macedonian: Gerazov *et al.*, 2010; Bengali: Reddy and Rao, 2011). Thus, modelling intonation patterns at the word and sentence level for others, especially in under-resourced languages, is a key challenge for current research in linguistics and computer science.

This paper focuses on intonation in one of such languages, namely Lithuanian, an Indo-European language from the Baltic group. This language is spoken by less than 4 mil. people around the world (according to the official information from Statistics Lithuania), and can therefore be considered an under-resourced language. From the perspective of linguistic description and study, Lithuanian had a somewhat paradoxical fate, receiving major attention on its accentuation system and being understudied in most other aspects. In the wake of historical linguistics in the 19th century, Lithuanian reached the peak of its linguistic glory, becoming the reference language for F. de Saussure, who based his Saussure's Law on Lithuanian pitch accent that he considered as the missing link of Indo-European linguistic history (Joseph, 2009; de Saussure, 1879, 1894, 1896). Ever since, Lithuanian has been perceived as the most "archaic of modern Indo-European languages", attracting the attention of many imminent historical and structuralist linguists, such as Antoine Meillet and Nikolai Trubetzkoy (Michelini, 2000; Petit, 2020). Later on, prominent generative phonologists, such as Morris Halle and Juliette Blevins also worked with its accentuation system (Blevins, 1993; Halle and Vergnaud, 1987). Nevertheless, due to historical circumstances in the region in the second half of the 20th century, international scientific collaborations became impossible, and most of the research on this language's prosody was published in Lithuanian or Russian and followed other theoretical approaches than the ones developing in the Western world. This resulted in a major linguistic and theoretical gap between prosodic research on Lithuanian in Lithuania and that carried out abroad. Prosodic phenomena other than the pitch accent, such as sentence intonation, received little attention and were often studied indirectly. A new opportunity to continue the study of Lithuanian intonation has been brought by the surge of speech and language technologies, which accentuated the urgent need to enhance linguistic knowledge of Lithuanian prosody. As is the case for other languages, these technologies necessitate integrated theoretical and applied linguistic knowledge, which has the potential to boost research in linguistics (Botinis *et al.*, 2001).

Therefore, this paper is the first attempt to gather research on Lithuanian intonation from both the Lithuanian and the Western traditions, the structuralist and generativist points of view, and the linguistic and modelling perspectives. We included in this overview all papers, books and conference proceedings published in English, Lithuanian and French dealing directly or indirectly with intonation in Lithuanian. Several unpublished but relevant theses on the topic are also discussed in this paper. We hope that this critical multidisciplinary overview will help summarize the somewhat fragmented research in the field in order to gain a better understanding of Lithuanian intonation and identify potential directions for future research both in linguistics and modelling.

The remainder of the paper is organized as follows: Section 2 provides a clarification of the terminology describing intonation. Section 3 briefly introduces intonation in terms of prosodic typology. Section 4 reviews Lithuanian intonation from the linguistic perspective by first situating the Lithuanian language within the word-prosodic typology (4.1), then discussing findings on sentence-level intonation in phonetic (4.2) and phonological (4.3) research. Section 5 reviews various techniques and models for prosody modelling. Two main approaches to intonation modelling are overviewed: directly modelling pitch and trying to simulate the pitch production process. Section 6 presents a critical discussion of the reviewed literature and provides suggestions of the major directions and further research in this field, with an emphasis on the role of intonation modelling for the development of linguistic theory and speech technologies.

2. What is Intonation? – Clearing up Terminological Ambiguities

Variations of pitch are used across languages and can thus be considered a universal feature. Nevertheless, this prosodic phenomenon is highly complex, as, despite its universal use, it has language-specific characteristics linked to the whole phonetic-phonological system of a given language. Moreover, it involves several levels of language, from lexical, to post-lexical and extralinguistic ones.

Before turning to the overview of the main typological distinctions of pitch and intonation, we would like to make a brief point about the somewhat confusing terminology used in this field of research. Depending on the authors and their theoretical background, words such as “intonation”, “prosody”, “tone”, and “pitch variations” either denote distinct phenomena or are used as synonyms (Botinis *et al.*, 2001). However, the exact meanings of these terms are only rarely explicitly defined. These differences in usage are explainable by the multiple ambiguities of the term “intonation” and the adopted angles of its study (see Fig. 1 for a schematic representation of the different aspects of intonation).

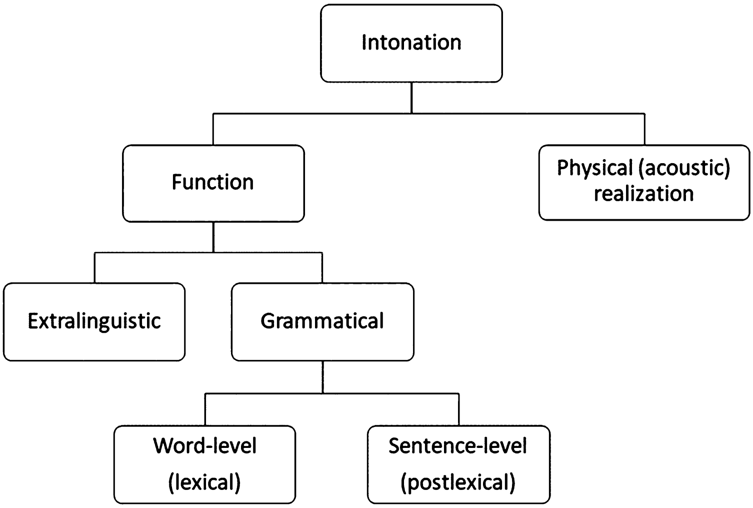


Fig. 1. A schematic representation of the different aspects of intonation examined in the literature.

Historically, intonation has received little attention and has long been seen merely as a means of expressing the speaker's attitudes and emotions. For this reason, it was considered to be a non-grammatical "form of animal communication" (Gussenhoven, 2004) which was "around the edge of language" (Bolinger, 1964). However, later studies and evidence showed that besides this extra- or para-linguistic function, intonation has a much more structured grammatical aspect which plays an important role in the world languages (see the autosegmental-metrical approach, Ladd, 1996; Pierrehumbert, 1980). Within the grammatical realm, intonation is still interpreted in two different ways (Hirst and Di Cristo, 1998). Under one interpretation, intonation is used interchangeably with the term "prosody" to define a wide range of suprasegmental phenomena, describing both the lexical identity of words (stress, tone) and the post-lexical or sentence-level concepts (intonational phrases etc.). Under the second interpretation, the term intonation only denotes the second, post-lexical aspect. The final source of ambiguity comes from the distinction between the physical/formal level of description and the linguistic/abstract one. While the physical approach attempts to identify and measure the acoustic and perceptual characteristics of intonation, the linguistic approach aims at examining the speakers' abstract cognitive representations of this physical manifestation (Hirst and Di Cristo, 1998). Although attempts have been made to find a one-to-one correspondence between these abstract prosodic phenomena and their acoustic realizations, it has become clear that the same acoustic means are used to express different abstract linguistic elements (Féry, 2017). Thus, known markers of sentence-level intonation, such as variations of fundamental frequency (F0), intensity, duration and rhythm, also play a role in the realization of word-level intonation, stress, as well as the expression of extralinguistic intonation. For this reason, references to F0 or "tone" might denote a variety of distinct prosodic phenomena. Finally, according to the autosegmental-metrical approach, the same abstract phonological melody can have different realizations (Arvaniti, 2022). Hence, the need to decompose the observable continuous pitch contours into a series of primitives in order to accurately describe and model intonation across languages (Arvaniti and Ladd, 2009). The main focus of this paper is grammatical intonation in Lithuanian. Although our primary concern is sentence-level intonation, work on other relevant lexical prosodic phenomena, especially pitch accent, will be discussed, as it is a necessary prerequisite for post-lexical intonation description and modelling.

3. Prosodic Typology and Intonation

As discussed in the previous section, melodic rise or fall can be characteristic of different grammatical features (Féry, 2017). At the word-level, intonation in some languages is specified lexically. In these languages (e.g. Mandarin, Thai or Vietnamese), variations of pitch indicate changes in the lexical meaning of a word. At the sentence level, intonation is understood as the use of tone for non-lexical purposes, such as syntactic demarcation, differentiation of sentence types, and signalling of information structure (Zerbian, 2010). As on the physical level, word-intonation and sentence-intonation, as well as other prosodic phenomena, share many characteristics, they can hardly be described in isolation, and their interaction is of typological interest (Hirst and Di Cristo, 1998; Ladd, 2001). Moreover, word- and sentence-level prosodic phenomena are related from the point of view of phonological structure. Specifically, sentence intonation is associated with feet and, thus, with stressed syllables (Arvaniti, 2022).

As no clear intonation typology has been established so far, classical prosodic typologies based on word-prosodic characteristics are still widely used for the description of cross-linguistic distinctions of post-lexical intonation. Many typologies distinguish stress, tone and pitch-accent languages (see Fig. 2) (Jun, 2005). Stress languages, such as English, German or Dutch, have lexical stress but do not use tone at the lexical level. This means that tones do not affect word meaning in these languages, as it is determined by the segmental content and stress. In such languages, lexical stress is obligatory and marks the most prominent syllable, while different tones or melodies are assigned post-lexically

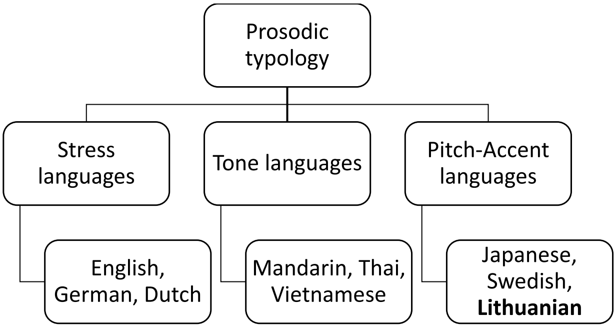


Fig. 2. A schematic representation of the prosodic typology with examples of languages belonging to the three classes.

(only at the sentence level) (Hyman, 2006). Stressed syllables serve as the location in the phonological structure where the intonation changes (shaping the intonation contour) occur (Gussenhoven, 2004). On the other hand, in tone languages, such as Mandarin, Thai or Vietnamese, the variations of pitch are specified lexically and are phonologically contrastive. However, sentence intonation is also used in these languages to a certain extent (Abramson, 1962; Connell *et al.*, 1983). The third class of languages, the pitch-accent languages, are somewhere in between stress and tone ones (Hirst and Di Cristo, 1998). Such languages (Japanese, Swedish) combine both stress and tone at the word-level but have restricted use of tone, whereby only the stressed syllable can have a tonal specification. In tone languages and pitch accent languages, the tone is lexically defined. Thus, sentence intonation comes additionally, and it has to accommodate the lexical tones (Féry, 2017). Similarly to stress languages, in pitch-accent languages, stressed syllables are positions that attract both lexical and post-lexical tones (Gussenhoven, 2004). Lithuanian belongs to this last class of pitch accent languages and thus uses intonation both at the lexical and post-lexical levels. The following section will introduce the specificities of Lithuanian prosody.

4. Intonation in Lithuanian

4.1. *Word-Level Intonation in Lithuanian*

As mentioned before, a number of studies have shown that Lithuanian is typologically a pitch-accent language, with tonal variations on heavy (i.e. bimoraic) syllables (Blevins, 1993; Girdenis, 2003). Fig. 3 is a schema of the prosodic syllable types in Lithuanian. Stress in Lithuanian is free (its position varies across words and forms of the same word). It is important to note that Lithuanian is one of the most highly inflected Indo-European languages, in which the location of stress depends on the underlying specifications of both the stem and the inflectional affix of every inflected nominal (Savičiūtė *et al.*, 2018; Kushnir, 2019). In terms of stress realization, the most important acoustic correlates are duration and intensity (Laigonaitė, 1958; Hock, 2015), as well as vowel quality (Girdenis, 2003; Hualde and Riad, 2014). Although traditional grammars posit a three-way distinction for Lithuanian pitch accents (Ulvydas, 1965; Mikalauskaitė, 1975), most research shows that only the acute accent [´] and the circumflex accent [˜] are contrastive (Girdenis, 2003). These pitch accents only occur in three types of heavy syllables: those with long vowels (e.g. rū́gsta ‘turn sour-3PRES’[1] vs rū̃ksta ‘smoke-3PRES’), those with diphthongs (šáukite ‘shoot-2PL-IMP’ vs šaũkite ‘shout-2PL-IMP’) and those that contain a short vowel followed by a sonorant (káltas ‘chisel’ vs kal̃tas ‘guilty’) (Dogil and Möhler, 1998). The circumflex pitch accent is steady or rising, while the acute is realized as a falling contour (Kushnir, 2019). There is an ongoing discussion about other possible acoustic correlates of the acute-circumflex accent opposition, but it has not produced conclusive results yet (for more information, see Girdenis, 2003; Dogil and Williams, 1999).

[1] Glosses are provided according to The Leipzig Glossing Rules (Haspelmath and Bickel, 2008).

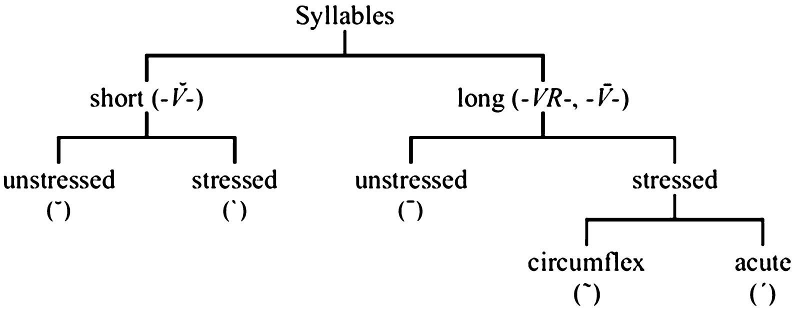


Fig. 3. Prosodic syllable types of standard Lithuanian, from Girdenis (2003).

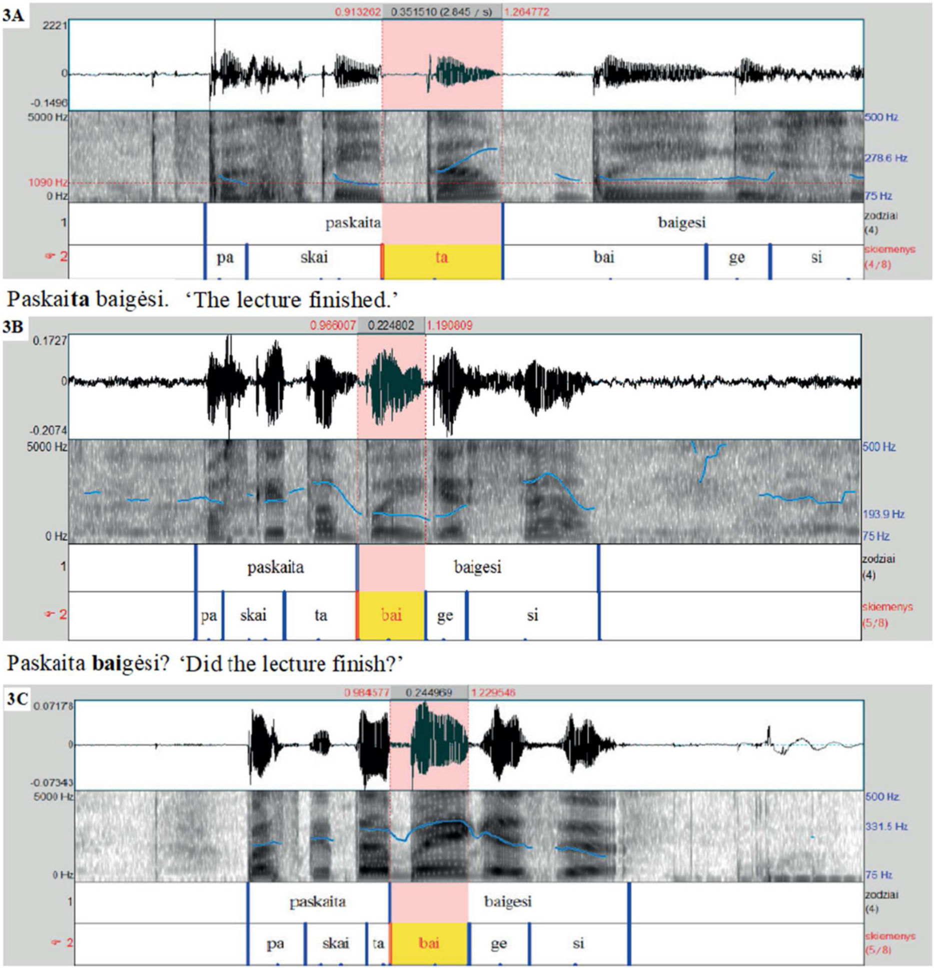
4.2. *Works on Sentence-Level Intonation from the Phonetic Approach*

Due to historical circumstances in Eastern Europe in the second half of the 20th century, Lithuanian prosody, especially at the sentence level, has not received sufficient attention in linguistics. This resulted in major gaps in the description of intonational patterns not only of Baltic (Kundrotas, 2017) but also larger, Slavic languages (Malisz and Żygis, 2017). This lack of investigations also explains why Lithuanian has not been described in the context of broader intonation typologies (Hirst and Di Cristo, 1998).

Lithuanian intonation at the sentence level was mostly studied by Lithuanian linguists from two standpoints: the syntactic and the phonetic. In his historical overview of research on Lithuanian intonation, Kundrotas (2020) notes that Lithuanian intonation at the sentence level was approached for the first time and in the most extensive manner by syntacticians (Balkevičius, 1963, 1998; Talandienė, 1970). As this research focused on the functional role of intonation rather than on its phonology and phonetics, its overview is out of the scope of the current paper. Phonological research was, for the most part, carried out by a single author, Kundrotas (2008, 2009, 2017, 2020).

Following the papers in syntax, the first, also rather sporadic, attempts to provide a phonetic description of Lithuanian intonation were published. Most of this research dealt with the phonetic description of sentence types. This direction of research was probably chosen due to the preliminary observations provided by studies on this topic in syntax. First, we can distinguish works aimed at identifying the acoustic correlates of sentence intonation in Lithuanian. Krapikaitė (2009, 2011) examined three possible phonetic correlates of intonation (F0, duration and intensity) in three sentence types (statements, questions and exclamations). She concluded that F0 is the main distinctive marker of sentence type in Lithuanian. Preliminary evidence from these studies shows that intensity modulations can signal focus-marking, while duration could be considered as a secondary marker of sentence intonation. The findings on the role of F0 have been confirmed in a pilot study by Kazlauskienė and Sabonytė (2018), who found that pitch variations in Lithuanian best reflect sentence intonation, i.e. sentence type and focus but are not a marker of stress.

The majority of papers dedicated to sentence types in Lithuanian aimed at providing a phonetic description of pitch variations in these sentences. For instance, Krapikaitė (2009)



Paskai**ta** baigėsi. ‘The lecture finished.’

Paskaita **bai**gėsi? ‘Did the lecture finish?’

Paskaita **bai**gėsi! ‘The lecture finished!’

Fig. 4. (A; B; C). Annotated example of: A. a statement phrase; B. a question; C. an exclamation, from Krapikaitė (2009). The stressed syllables of focused words are in bold.

recorded statements, questions and exclamations produced by multiple speakers of Standard Lithuanian. As can be seen in an annotated example of a statement phrase (Fig. 4A), intonation starts low at the beginning of the phrase, then rises and peaks on the third and stressed syllable of the first word. It gradually decreases in the second word and reaches a low plateau.

In the question phrase (Fig. 4B) the intonation of the first word increases slightly until the third stressed syllable of the first word. It then drops significantly on the first stressed syllable of the second word and rises again to reach the intonational boundary peak. As noted by the author, the suffix –si is a reflexive marker which often undergoes reduction to –s. Therefore, the final fall of intonation in this particular phrase should not be considered as a trait of questions in Lithuanian. Thus, questions seem to follow the raising-falling-raising patterns. In the exclamation (Fig. 4C), intonation rises to its peak on the first, stressed, syllable of the second word and gradually falls towards the end of the word. The amplitude of this fall is the highest of all three sentence types. In addition to this, the overall pitch in the exclamatory sentence is higher than in declarative sentences and questions (this has also been found by Kazlauskienė and Sabonytė, 2018). Note that the melodic contour is opposite in questions vs exclamations: while in the question intonation is at its lowest point on the first syllable of the second word, it reaches its peak on the same syllable in the exclamation. Krapikaitė (2009) concludes that this different pattern of F0 on the same stressed syllables in different sentence types are an indication that in Lithuanian, the melodic contour of intonation depends on sentence focus rather than on lexical stress. Specifically, the intonational peak is located on the focused words, and distinctive pitch events happen in or after the stressed syllables of these words. For instance, in the statement phrase ***Paskaita*** *baigėsi*, the focus is on the first word, where the peak is on the stressed syllable –ta and is followed by a fall in intonation. Similarly, in the exclamation *Paskaita* ***baigėsi!*** The focus, as well as the peak, is on the stressed syllable –bai of the second word which is followed by a sharp intonation fall. In the question *Paskaita* ***baigėsi?*** The focus is also on the second word, but this time the stressed syllable –bai has a low tone, which is followed by a strong rise. Overall, this evidence suggests that phrase type in Lithuanian is defined by the movement of F0 in the intonational centre, together with the tone level in the post-nuclear part of the contour (higher or lower relative to the pre-accented part) (Kundrotas, 2008).

Although most of this work was of phonetic nature, in Krapikaitė (2015), an attempt to provide some phonological generalizations using ToBI (Tones and Break Indices, Beckman and Ayers, 1994) was made. The author identifies possible pitch accent tones in different sentence types, using the same sentence examples as in her previous work: a rising peak accent L+H* was identified on the last syllable –ta of the first word in the statement sentence (Fig. 4A); a low tone L* on the first syllable of the second word in the question (Fig. 4B); and a high pitch accent H* on the same syllable in the exclamation phrase (Fig. 4C). Krapikaitė suggests that each sentence type contains a single pitch accent and can thus be distinguished based on them. However, this paper uses a very limited set of tonal events proposed by ToBI, making the analysis incomplete in at least two aspects. First, it seems necessary to include additional pitch accents to account for the melodic contours of the given sentence types. Second, in order to make generalizations on markers of sentence type, it is important to address phrasal tones. For instance, we propose that the statement type should include L-L% final phrase boundaries, which have been widely attested cross-linguistically in declarative sentences (Gussenhoven, 2016). This would result in a L+H* L-L% melody for the statement phrase given in the examples.

*Paskai**ta***	*baigėsi.*
Lecture-NOM	finish-PST
L+H*	L-L%

‘The lecture finished.’

In order to obtain a high-low-high melodic contour attested in Krapikaitė (2009, 2011) in the question phrase, the description should include a H* pitch accent in the first word (on the last syllable –ta). This would lead to H* L*, and final tones H-H%, which, again, are cross-linguistically attested as markers of yes-no questions (Hedberg *et al.*, 2017).

Paskaita	***bai**gėsi?*	
Lecture-NOM	finish-PST	
H*	L*	H-H%

‘Did the lecture finish?’

Finally, the exclamation phrase in the given example could be analysed as a sequence of a H* pitch accent in the first word, followed by the H* in the second and a L-L% final tone combination.

Paskaita	***bai**gėsi!*	
Lecture-NOM	finish-PST	
H*	H*	L-L%

‘The lecture finished!’

A more detailed analysis of question intonation was proposed by Kazlauskienė and Dereškevičiūtė (2018). They studied three types of question sentences and simple statements in order to identify the main intonational patterns of these sentence types. The authors used ToBI, but, again, chose to use a simplified notation with single-tone H* and L* pitch accents and H% and L% boundary notations. The first type of questions to be analysed was the yes/no question without the particle “ar”, which has been shown to be the most used yes/no question formulation in spoken Lithuanian (Balčiūnienė and Simonavičienė, 2009). These questions are comparable to questions without morphosyntactic question markers in English (Grabe and Karpinski, 2003). Two patterns were identified:

	Galime	*eiti*	*namo?*
	Can-2pl-PRS	go-INF	home
Pattern 1	L*	L*	H* H%
Pattern 2	L*	H*	L* L%

‘Can go home?’

The first one follows the typologically common low-rise nuclear contour, which has also been observed by Krapikaitė (2009, described above). The second, the falling pattern, reflects the focus placement on the second word “eiti”, ‘can we GO home’?

The second type of questions are yes/no questions starting with the interrogative particle ‘ar’.

	Ar galime	*eiti*	*namo?*
	Int.Part can-2pl-PRS	go-INF	home
Pattern 3	H*	L*	H* H%
Pattern 4	H*	L*	L* H%
Pattern 5	L*	H*	L* L%

‘Can [we] go home?’

Similarly to pattern 1, pattern 3 ends in a low-rise nuclear contour, but the first word is assigned a high pitch accent. It is possible that the presence of the interrogative particle in the clause-initial position affects this change in pitch accent. Note also that while in pattern 1 and 2 the only marker of questions was intonation, in pattern 4 the particle "ar" unambiguously indicates the sentence type from the very beginning of the phrase. Thus, one could expect a smaller intonation rise in yes-no questions with the particle. Although the amplitude of the final rise was twice as high in pattern 2 compared to pattern 3, the rise is still clearly present in pattern 3. Kazlauskienė and Dereškevičiūtė further investigated the interaction between the particle and intonation as markers of question type. They tested the participants' perception of sentences with and without the particle "ar" by cutting it out from the recordings and asking participants to identify the sentence type. Participants were more than 90 % correct in identifying questions even when the particle was removed, pointing to the fact that it is indeed intonation that conveys most information about the sentence type. Comparable evidence comes from Polish, where a similar particle "czy" is used in yes-no questions. Mikoś (1976) found a small, although non-significant difference in questions with vs without the particle in Polish and concluded that "czy" is an optional particle, which could be treated as a redundant feature. Turning back to the last two patterns of Lithuanian yes-no questions with interrogative particles, pattern 4 describes the case with a focus on the phrase-initial word 'CAN [we] go home?', which conveys impatience. Pattern 5 mirrors pattern 2 and emphasizes the second word 'go'. Thus, this evidence points to the role of intonation in marking yes-no questions both in sentences with and without the interrogative particle "ar". More studies are needed in order to shed light on whether this particle is a redundant marker of question type, or whether it has an impact on the phrase-initial intonational contour and, possibly, on the amplitude of the final rise. Finally, the rising contour seems to best describe the simple yes-no question, while the falling contour is used to move focus away from the last to the first or second syntagma in the phrase.

Kazlauskienė & Dereškevičiūtė end their study by analysing wh-questions:

	Kada	*galime*	*eiti*	*namo?*
	When	can-2pl-PRS	go-INF	home
Pattern 6	H*	L*	L*	L* H%
Pattern 7	H*	H*	L*	L* L%
	'When can we go home?'			

In both cases, wh-questions have a falling contour, but they differ in focus marking: in pattern 2 the wh-word is marked, while in Pattern 7, the second word 'can we' is marked. Interestingly, when the wh-word was excised from the recordings, participants' accuracy in identifying the sentence type fell drastically. This result is expected for pattern 6, where the removal of the wh-word leaves the sentence with no other lexical markers of question type (in Lithuanian, there is no subject auxiliary inversion). Note, though, that the L*H% pattern at the end of this phrase could potentially be used to identify the sentence type. The result of the perceptual experiment is even more surprising for pattern 7 as even without the wh-word, the sentence has the H*L*L*L% contour similar to pattern 4, which should be indicative of question type, as High-Fall contours are attested to be a cross-linguistically common pattern for wh-questions (Hedberg *et al.*, 2010). Thus, further investigation is needed in order to clarify to what extent lexical vs intonational information is used in Lithuanian to identify the sentence type.

As can be seen, the above-cited phonetic studies target sentence type intonation, although focus marking is also addressed in a less direct manner. These studies provide the first glimpse into the peculiarities of sentence intonation in Lithuanian and raise many interesting considerations as to the presence of certain cross-linguistically attested intonational patters or the complex interaction of intonation with other syntactic markers (e.g. particles). Note, however, that these studies had small sample sizes and did not carry out (or did not report) statistical tests. Therefore, their findings are of a rather preliminary nature. Moreover, none of them discussed the interaction between the lexical pitch accent and the described post-lexical pitch variations. Finally, some of the studies made a first attempt to use an internationally recognized framework for the development of prosodic annotation, i.e. ToBI, and to adapt it to the Lithuanian language. In these papers, only some fragments of the ToBI system were used, making the results rather incomplete and lacking a deeper phonological analysis. The following subsection will review studies that targeted explicitly the phonology of sentence-level intonation in Lithuanian.

4.3. *Works on Sentence-Level Intonation from the Phonological Approach*

Intonation in Lithuanian has so far been almost overlooked in the field of phonology. Some researchers working on Lithuanian phonology and accentology mention the major functions of sentence intonation but consider this phenomenon to be out of the scope of their research (Pakerys, 2003; Kazlauskienė, 2012). Others, such as the most prominent Lithuanian phonologist Aleksas Girdenis, acknowledge the importance of intonation but do not study this phenomenon because of its complexity. For this reason, intonation is only fragmentally studied when it is necessary in order to explain some suprasegmental elements of a word (Girdenis, 2003). Finally, a prevailing view in some research circles is that more phonetic description is needed in order to be able to make generalizations about the more abstract phonological level (Kundrotas, 2020).

Perhaps the most extensive phonological analysis of Lithuanian intonation was conducted by G. Kundrotas. It is important to note that Kundrotas carried out significant research in documenting intonational patterns in varying contexts and across speech styles, as well as testing their validity experimentally. The author follows Trubetzkoy's (1969) method and seeks to identify meaningful phonemic oppositions in Lithuanian intonation. He distinguishes three such types of oppositions (the stressed syllables of focused words are in bold):

a) Different tone contours

- ***Šian***_dien_ *jis* *išvažiuoja?*
  Today he-NOM leave-3sg-PRS
  'Does he leave today?'
- ***Šian***_dien_ *jis* *išvažiuoja.*
  'He leaves today'.

b) Different locations of prosodic focus

- ***Šian**dien jis išvažiuoja?*
  ‘Does he leave TODAY?’
- *Šiandien jis išva**žiuo**ja?*
  ‘Does he LEAVE today?’

c) Different phrasing (different prosodic constituents marked by boundary tones and pauses)

- *Senelis **pa**sakojo: jie ten seniai*
  Grandfather-NOM tell-3SG-PAST they-NOM there for a long time
  *gy**ve**no.*
  live-3SG-PAST
  ‘The grandfather told: they lived there for a long time’
- *Se**ne**lis, pasakojo jie, ten seniai*
  Grandfather-NOM tell-3PL-PAST: they-NOM there for a long time
  *gy**ve**no.*
  live-3PL-PAST
  ‘The grandfather, they told, lived there for a long time.

It is important to mention that this researcher follows the theoretical framework of the Prague structuralist school and posits a holistic view of intonation (Bolinger, 1964; Liberman, 1975; Liberman and Sag, 1974). Therefore, Kundrotas argues that the intonation contour expressed by F0 is the main and indecomposable intonation unit in a language (Kundrotas, 2008, 2009). According to the author, the contour varies according to two phonologically meaningful dimensions: the tone of the intonational nucleus (focus or sentence stress) can be either rising or falling; the height of the contour following a nuclear syllable can be higher or lower than its level before the nuclear syllable. Based on these assumptions and experimental studies, Kundrotas (2008, 2009) identifies seven intonational contours in Lithuanian, which will be presented below.

The first intonation contour (IC1) is characterized by a gradual lowering of the tone in the nucleus, the post-nuclear part of the contour being lower than the pre-nuclear part. This intonation contour is mostly used in declarative sentences:

*Lauke seniai **sute***^{1}*mo.*
Outside for a while be dark-3SG-PAST
‘It has been dark outside for a while already’

IC2 is similar to the first one, as it also starts at the normal pitch level of the speaker, then gradually falls in the nucleus and further falls in the post-nuclear part of the contour. However, the pitch in the nucleus is slightly higher than in IC1. IC2 is used in:

- Wh-questions:

*Kaip jis **mo***^{2}*kosi?*
How he-NOM study-3SG-PRES
‘How does he study?’ (How good his results are?)

- Exclamations or vocatives.

_**Vy**_^{2}*tai!*	_Pa**lauk**_^{2}*!*
Vytas-VOC	Wait-2SG-IMP

‘Vytas! Wait!’

Jis	_**ma**_^{2}*no*	*draugas.*
He-NOM	my	friend-NOM

‘He is MY friend’ (not yours)

The main distinctive feature of IC3 is the gradually rising or the rising-falling tone in the nucleus. In the post-nuclear position, the tone gradually falls and reaches a lower position than in the pre-nuclear part of the contour. According to Siniova and Kundrotas (2014), this contour is widely used in Lithuanian in a variety of contexts:

- In questions without interrogative words

Jis	*čia*	_**mo**_^{3}*kosi?*
He-NOM	here	study-3SG-PRES

‘Does he study here?’

- In statements to express continuation of a thought

Jis	*mokėsi*	_Vokie**ti**_^{3}*joje,*	*Dabar*	*dirba*
He-NOM	study-3SG-PAST	Germany-LOC,	Now	work-3SG-PRES
Lietuvo**je**				
Lithuania-LOC				

‘He studied in Germany, [and] now works in Lithuania’.

- In polite requests

Uždary^{3}*kite*	*langą.*
Close-2PL-IMP	window-ACC

‘Could you close the window?’

IC4 is characterized by a falling-rising pitch. Specifically, the tone slightly falls at the beginning of the nucleus but gradually rises throughout the nucleus and in the post-nuclear part of the contour. The post-nuclear part of the contour is higher than the pre-nuclear part. This intonation contour is used in:

- Interrogative sentences with the particle “o”, expressing juxtaposition.

Jonas	_**jau**_^{1}	*kalba*	*lietuviškai.*	*O*	_**Pe**_^{4}*tras? //*
John-NOM	already	speak-3-SG3PRES	Lithuanian.	PART	Peter-NOM?

‘John already speaks Lithuanian. What about Peter?’

- Yes/no questions in formal speech, signalling official tone

Jūsų	_pavar**dė**_^{4}*?*
Your-GEN	surname-NOM

‘Your surname, [please]’

- Questions expressing a moralizing note

Ko**dėl**^{4}	*taip*	*vėlai*	*sugrįžai?*
Why	so	late	return-2-SG-PAST

‘Why did you return [home] so late?’

IC5 has two intonation centres, with a rising tone in the first and a falling tone in the second one. Both the intensity and duration of the syllables which constitute the intonation centres are higher compared to other stressed syllables. The two intonation centres can either be located close to each other in the utterance or can be separated by several syllables. Intonation between the two centres is higher than in the pre-nuclear part of the contour but lower in the post-nuclear part. This intonation contour is used in:

- Exclamations to express superlatives or to emphasize a trait, especially when the adjective is elided.

Ko^{5}*kia*	*die**na***^{5}*!*
What-NOM	day-NOM

‘What a day!’

- To express a wish

Kad	***ti***^{5}***k***	*jis*	***pasvei***^{5}*ktų!*
If	only	he-NOM	get better-3SG-COND

‘If only he could get better!’

IC6 is characterized by a gradually rising pitch in the nucleus. The intonation remains higher in the post-nuclear position compared to the pre-nuclear part of the contour. The duration of the nuclear syllable is longer than the duration of other stressed syllables in the phrase. This duration difference is the main distinctive feature of IC6, compared to IC3, which has also a rising pitch in the nucleus. It is used to:

- Express continuation of a thought:

Girdėjote	*žinių*	***lai***^{6}*dą*	*„Respublika*	***šian***^{1}*dien“*
Listen to-2PL-PAST	news-GEN	program-ACC	the Republic-NOM	today

‘You were listening to the news program “The Republic today.”’

- In interrogative sentences to express surprise or doubt

Kur^{6}	*jis*	*dabar*	*gyvena?*
Where	he-NOM	now	live-3-sg-PRES

‘Where is it that he lives now?’

IC7 is characterized by a rising or rising-falling contour in the nucleus. Pitch falls in the post-nuclear position and becomes lower than in the pre-nuclear part. The distinctive feature of IC7 is the glottalization of the nucleus. It is used in exclamatory, imperative and interrogative sentences as a means of expressive negation by using an opposing meaning to what has been said:

*Ka**da***^{7}	*jis*	*atvažiuos!*
When	he-NOM	come-3sg-FUT

‘Is he ever coming!’ (I.e. he will not come).

Our review of the studies by Kundrotas reveals that this author provided the most exhaustive description and phonological analysis of Lithuanian sentence-level intonation to date. In his numerous papers, the researcher not only identifies major melodic contours used in Lithuanian but also points to the fact that the same tune might be used in different contexts, and several tunes might be used to express the same function. Unfortunately, the holistic view of intonation posited by the author makes the further development of Lithuanian intonational theory somewhat problematic. Specifically, the view that the intonation contour cannot and should not be divided into smaller units makes it hardly possible to explain the "systematic variation" observed in the different realizations of the same contour (Arvaniti and Ladd, 2009). Moreover, in order to capture the diversity of remaining intonation patterns occurring in the language and their variations, one would have to come up with a very large number of such indecomposable melodies. In addition to this, the theoretical framework followed in these studies does not allow to take into account intonation at the word level (pitch-accents) and their interaction with sentence-level intonation. Finally, it impedes the comparison of these results with studies on other languages carried out by the international research community in recent decades, which mostly follow the now widely accepted Autosegmental-metrical framework. The following section will introduce this model as well as several other theoretical and computational models of intonation. The possibilities to apply these models to the Lithuanian language will be discussed in the Discussion section.

5. Intonation Modelling

5.1. *Approaches to Intonation Modelling*

From the modelling perspective, intonation is understood as the variation of fundamental frequency F0 (in terms of speech production) or pitch (in terms of speech perception). One way of capturing intonation could be by making physiological measurements of the articulation process – the relationships between F0 and articulatory gestures and the physiological properties of the vocal tract are obvious and undisputed. However, such measurements of the articulation process are complex and difficult to implement, thus making acoustic modelling more appealing. The principles of acoustic analysis enable the exploration of frequency properties (F0 among them) by measuring the wave of the speech signal. The obtained measurements and their derivatives are much more related to perceptible manifestations of intonation. Therefore, most of the modelling techniques are based on acoustic concepts and measurable speech signal properties. Various techniques and models were proposed for prosody modelling. According to Xu (2015), a three-way division can be made across the prosodic theories: linear vs superpositional, formal vs functional, and acoustic vs articulatory. There are two main approaches to intonation modelling: directly modelling the pitch and trying to simulate the pitch production process (Honnet, 2017). Nevertheless, as noted above, certain F0 variations are language-dependent, and this should be considered when employing various intonational models.

5.1.1. *Fujisaki Model*

The Fujisaki model (Fujisaki and Hirose, 1984) was proposed considering that the modelling of the F0 contour (which is the primary indicator of the intonation process) could not be described using a simple combination of straight, declining and rising approximations of the F0 segments. The authors claimed that natural and high-quality prosody in synthesized speech could be obtained only with complex functional dependencies.

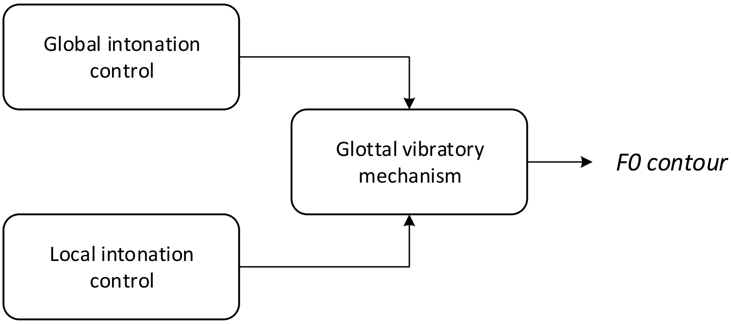


Fig. 5. Graphical representation of the Fujisaki model.

Their argument was based on the continuous nature of F0 contour despite discrete prosodic information elements in human speech.

The main assumption of the model was the presence of two components in the analysed contour: the phrase-level (global, sentence level) element and the accent-level (local, word level) element (Fig. 5). Both components are modelled using a second-order linear system with different excitation signals. In the case of the phrase-level component, an impulse is proposed as the system input. For intra-phrase variations of F0, positive impulses are employed, while for rapid fallings (e.g. at the end of a phrase), a negative impulse is suggested. In the case of an accent-level component, the system is excited with a stepwise unit. The sum response of two linear system combinations describes the F0 contour as the continuous variation of the F0 with baseline, maximal level and other characteristics making the intonation model smooth and close to natural intonation.

It is important to note that the Fujisaki model was initially designed for Japanese, which is considered a pitch-accented language, i.e. the F0 (tone) can convey both lexical meaning and phrase-level patterns. Therefore, the assumptions for the Fujisaki model were inspired by the F0 variation-rich Japanese language, and this fact should be taken with care in the intonational analysis of non-tonal, stress-accent languages. In the case of the Lithuanian, though, this characteristic can be useful.

5.1.2. *Tilt Intonation Model*

The main idea of the Tilt intonation model (Taylor, 1998) is to analyse intonation as a sequence of intonation events. The author defines two kinds of acoustic events: pitch-accents (segmental level) and boundary tones (suprasegmental level). Each event is characterized by the rise, fall and other varying F0 shape properties (so-called tilt parameters). Some events can consist of the rise part only, and some can consist of the fall part only. Every event is described with a particular parameter set: the amplitude of the tilt, the duration, the tilt (the value $+1$ defines the rise, -1 defines the fall), and the F0 position.

The Tilt model closely relates to the Rise/Fall/Connection model (Taylor, 1994). In addition to rises and falls, the later model includes connections for non-intonational segments with a constant value of F0. There the intonation is also modelled as a linear combination (sequence) of rises, falls, and connections. For example, the pitch accent (stressed

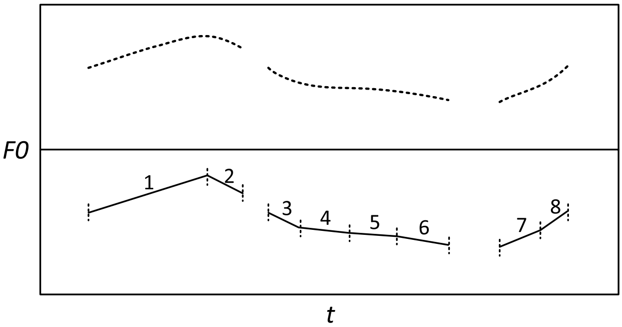


Fig. 6. Linear approximation of F0 contour.

syllable) is modelled as a set of a single rise and a single fall. Similarly, like in the Tilt intonational model, parameters such as rise amplitude or duration are also considered in the RFC model. The Tilt intonation model is similar to Fujisaki model in the way that both are used to describe the variation (rises and falls) of F0.

5.1.3. *IPO Model*

Again, this model is based on acoustic and perceptual aspects of the speech signal and relies on the approximation of the F0 contour (Hart *et al.*, 1990). The intonation (the melody) of the speech signal is considered as containing micro-intonation and macro-intonation events. The reason for micro-intonation is purely physiological: inertia and non-linear properties of the articulatory system, F0 fluctuations of unconscious origin. This type of intonation is considered to be irrelevant to the intonation analysis and, therefore, should be ignored.

The F0 contour is modelled in the following steps. First of all, the F0 contour is extracted from the speech. All the detected phenomena of micro-intonations (i.e. micro fluctuations) are removed in order to get relevant and meaningful intonation patterns. After preprocessing, the F0 contour is "stylized", i.e. F0 contour is analysed in segments and approximated by lines (Fig. 6). The obtained piecewise linear approximation of the contour is described using F0 patterns such as falls and rises. Although the model was developed using the Dutch language and tested for English, the model itself is based on F0 estimation and approximation and ignores language-specific presumptions and aspects.

5.1.4. *PENTA Model*

The lexical and intonational nature of F0 variations is addressed in the Parallel Encoding and Target Approximation (PENTA) model (Fig. 7). Two assumptions are at the basis of this model (a detailed figure and description can be found in Xu, 2004). First of all, during speech production, parallel communicative functions (lexical, sentential, topical, grouping emotional, etc.) are conveyed through individual encoding schemes. These schemes can be language-specific or universal and determine the target parameters (pitch target, pitch range, etc.) controlling the approximating articulatory process. This parallel encoding part is the PEN part of the model, and target parameters are supposed to define the

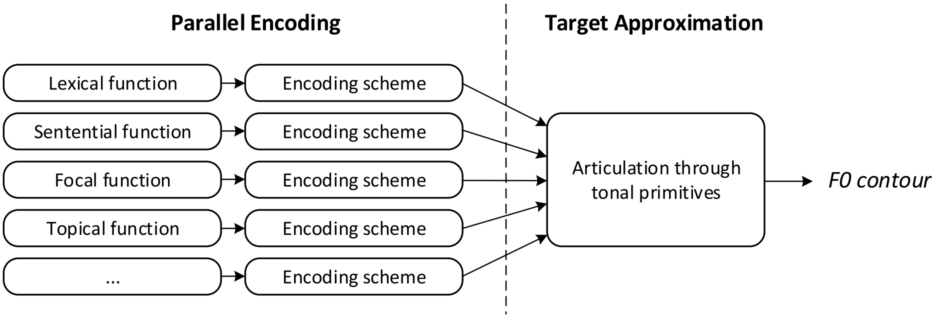


Fig. 7. PENTA model.

so-called melodic (tonal) primitives. The second assumption is related to the approximation process: various physical and neurological properties limit the generation of the F0 contour, resulting in F0 target approximation (the TA part of the model). Taking defined tonal primitives, the articulatory system approximates the F0 targets, each synchronized with a syllable.

These assumptions differ from the aforementioned intonation models as these models contain nucleus F0 elements to shape complex F0 contour. In comparison, the PENTA model is more abstract. Nevertheless, its complex nature enables the inclusion of specific language properties, phonological factors, etc.

5.1.5. *Autosegmental-Metrical (AM) Model*

The Autosegmental-Metrical (AM) approach has originated as a result of a few influential studies and doctoral theses in intonational phonology, including works by Liberman (1975), Bruce (1977), Pierrehumbert (1980), Ladd (1996). This phonological model represents intonation as a combination (sequence) of low and high tones – only two different tonal levels are considered. Between these tonal events, the F0 (pitch) contour is undefined and can be modelled as an interpolation between these events. The AM model relates these events with particular points in the utterance: prominent syllables in the segmental string and utterance boundaries. The tones associated with the syllables are called pitch accents, and the boundary-associated points are called edge tones. Edge tones mark the boundaries of the intonational phrase and are independent of pitch accents. Pierrehumbert (1980) defined two types of edge tones: final boundary tones and phrase accents. The distinguishing property of the AM model is that it considers tonal events as the linguistically important part and the remaining part of the F0 (pitch) contour only as a transition between events. This is the main difference from the IPO model.

Intensive research on AM model ideas has led to the development of ToBI (Tone and Break Index), a framework for the development of prosodic annotation systems (Beckman *et al.*, 2005). The latter has been applied to a variety of languages, such as ToDI – Transcription of Dutch Intonation (Gussenhoven, 2004), IViE – Intonational Variation in English (Grabe *et al.*, 1998), J-ToBI for Japanese (Venditti, 2005).

5.1.6. *Alternative Techniques*

For application purposes, generative frameworks were applied to synthesize intonational and emotional speech signals. These techniques focus on the acoustic realization of the prosody (intonation), avoiding formal phonological analysis or explicit theoretical justification. Their relation with formal intonation models is weak, and it is limited to the use of certain terms and principles (such as the rise and fall of the F0).

Two different paradigms are followed here: the modification of F0 contour in synthesized speech (i.e. re-synthesis in order to provide intonation) or the generation of an intonational contour "from scratch". In formant-based synthesis, a rule-based modification of fundamental and formant frequencies is applied (Kohler, 1991; Carlson *et al.*, 1992; Pyž *et al.*, 2014). Frequency information is also supplemented with temporal indicators of intonation (speaking rate, pauses, phoneme duration) (Cahn, 1990). The rigorous and formalized modelling of the F0 contour is continued by applying mathematical approximation methods (Hirst and Espesser, 1993; Andruski and Costello, 2004; Grabe *et al.*, 2007).

Later, data-driven techniques emerged and began to dominate. In concatenative synthesis, the Hidden Markov Model framework was the main intonation modelling technique. Here, corpus data-based intonation is modelled both at the phonetic unit level (e.g. syllable, phoneme) (Boidin and Boeffard, 2008) and at the utterance level (Ni *et al.*, 2016). The modelling required additional information like context, prosody and stress events, and dependence on intonational class. Nevertheless, corpus and HMM-based intonation modelling had issues with data sparsity and over-smoothing.

The proliferation of artificial neural network-based techniques in speech synthesis has brought new opportunities. Neural networks are well known for their ability to capture hidden and very complex dependencies, which is an excellent assumption for intonation modelling. During the past decade, numerous studies were published on neural network-based modelling of intonation (Fan *et al.*, 2014; Zen and Sak, 2015; Ming *et al.*, 2016; Stehwien *et al.*, 2020). Again, neural networks were employed both to model intonation separately and in combination with other speech synthesis features (a more common modelling case).

5.2. *Lithuanian Intonation Modelling*

An overview of the development of intonation models for Lithuanian speech is given in this section. The works related to both word- and sentence-level phenomena are considered.

5.2.1. *Sentence-Level Intonation Modelling*

Regarding the issue of sentence-level intonation, a computational model of the fundamental frequency for the Lithuanian language was proposed by Vaičiūnas *et al.* (2016). The model was obtained by training the Hidden Markov Model (HMM)-based speech synthesis system HTS on six hours of speech coming from multiple speakers. Fundamental frequency and its derivatives were employed in the training process. As the phoneme was chosen to be the modelling unit, a set of contextual factors was used to describe its context. A list of factors (originally taken from English, Tokuda *et al.*, 2002) was related to the phoneme, the utterance, as well as previous, current and next syllable, word and phrase. The model takes text labels, part of speech tags, stress information, number of syllables, words, phrases and other linguistic and non-linguistic features. In the experimental part of the paper, the authors investigated speaker-independent but gender-dependent models and stated that the HMM-based models showed an improvement over the baseline mean-based model. They also indicated that the set of contextual factors that was good for English appeared to be useful for Lithuanian as well. As one of the directions of future research, the authors point to a detailed investigation of additional Lithuanian-specific contextual factors that could improve over the present contextual factor set. Raškinis and Kazlauskienė (2013) aimed to discover characteristic patterns of Lithuanian intonation through the clustering of pitch contours in intermediate intonation phrases. These computational procedures were applied to 23 hours of read speech containing 41417 phrases, on which orthographic word-level transcriptions were made. The authors concluded that language independence is an important feature of computational procedures covered by their paper. Also, it is pointed out that this type of research could be useful for experts as a means of creating an inventory of intonation patterns. The authors state that the paper presents research carried out in preparation to compile a Lithuanian intonation corpus, but to our knowledge, such a corpus has unfortunately not yet been compiled.

Among attempts to create a model of Lithuanian sentence-level intonation, a work of Leonavičius (2006) can be included, in which the authors modelled pitch variations of melismas. In principle, variations of pitch are considered as melismas in musical notation and as intonation in speech signal processing. The aim of this work was to synthesize melismas met in Lithuanian folk songs by applying Artificial Neural Networks. More than 500 melismas were used in the experiment. As a result, the original mathematical models of all four kinds of melismas have been created. Unfortunately, no subsequent publications by the author can be found that continue this promising research.

Finally, articles that focus on phone duration models of Lithuanian, which is one of the possible phonetic correlates of sentence-level intonation, could be mentioned. A review of existing models of duration prediction is given by Norkevičius and his colleagues (Norkevičius *et al.*, 2006; Norkevičius and Raškinis, 2008). In the experimental part of these papers, a classification and regression tree (CART) method was used for creating some preliminary phone duration models. The conclusions briefly discuss the possibilities of using this method for the Lithuanian language. Comparable results of Lithuanian phoneme duration models were obtained by Kasparaitis and Beniušė (2016). For duration modelling, the rule-based model proposed by Klatt (1979) was used. The authors proposed a new iterative algorithm for the automatic estimation of the factors. The investigation showed that the contextual effects with the strongest influence on phoneme duration in Lithuanian appeared to be: the end of the sentence, the beginning of the sentence for the consonants and a consonant in a group of consonants.

With regard to intonation annotation, there have been studies related to ToBI, an international standard for annotation and transcription of prosodic events. An investigation of whether this transcription system is applicable to the prosodic annotation of Lithuanian is presented in the works of Krapikaitė (2014, 2015). The papers presented the use of ToBI transcription in the Lithuanian Prosodic Corpus, which is being built at Vytautas Magnus University. The results of the analysis show that the intonational contour in Lithuanian can be transcribed separately in individual tiers representing independent structural types.

5.2.2. *Word-Level Intonation Modelling*

There have also been some attempts to model intonation at the word level. In this context, prosodic phenomena, such as the pitch-accent, are considered. Lithuanian pitch accents were modelled by Dogil and Möhler (1998) using a parametric phonetic description of characteristic F0 shapes. The parametric model was represented by a polynomial function. Two contrastive accents of Lithuanian were investigated: the acute accent [´] and the circumflex accent [˜]. The aim of the study was to check which of the pitch-accent parameters is responsible for the prosodic salience of Lithuanian accents. The results of the approximation of the polynomial function on Lithuanian pitch-accents showed that the acute accent could be quite precisely approximated given the set of five parameters, whereas the circumflex accent defies such an approximation (Dogil and Möhler, 1998). The acute accent is a highly invariant pitch-accent with a clearly definable form, alignment point within a mora and a precisely defined slope, and it is characterized by a large F0 amplitude. The situation is different with the circumflex accent. The low F0 amplitude and the highly variant contour and alignment of the circumflex accent make it a very indeterminate representative of pitch-accents as a phonetic category. Another paper that deals with Lithuanian pitch accent modelling was published by Paulikas and Navakauskas (2005). They aim at creating a general model that could represent the voiced speech signal with the pitch accent. The term "general" means that the authors do not distinguish between models according to the actual accent types but rather according to how their parameters were estimated. A proposed model is incorporated into the restoration algorithm and used in the homograph restoration process of Lithuanian words. Several accent models were developed, and their performance was compared in the experimental part of the paper. The third-order polynomial model that includes the polynomial approximation of normalized intensity and the period of the fundamental frequency as parameters was shown to be the best. According to the authors, thorough experimentation on a wider set of homographs is necessary to evaluate the proposed polynomial model and select optimal approximation orders. It should be noted that the model does not incorporate the duration of a missing accent as a parameter. It was assumed that the duration was known.

Interestingly, literature analysis showed that, although some works state that they only deal with the stress of Lithuanian words, they, in fact, also address the issue of pitch-accents. An example of such works may include studies on the automatic stressing of Lithuanian texts. The rules of stressing are given in many grammars of the Lithuanian language (e.g. Ambrazas *et al.*, 1996). However, these rules are inappropriate for computer use. Several articles have addressed this problem. The first algorithm for the automatic stressing of Lithuanian texts is presented by Kasparaitis (2000). This paper deals with the stressed parts of speech on the basis of a dictionary. Later, the aforementioned algorithm was improved by including formal stressing rules for nouns and adjectives (Kasparaitis, 2001). By means of these rules, such words as diminutives, names and degrees of adjectives that are not present in the dictionary may be stressed. The automatic stressing methods discussed are based on morphological analysis only. According to Anbinderis (2010b), such methods are complex and difficult to modify. The author has proposed a method that uses only sequences of letters, does not require any knowledge of the language, and is, therefore, very simple, fast, and easy to adapt to other languages (Anbinderis, 2010a, 2010b). This method has been supplemented with clitic (Anbinderis and Kasparaitis, 2007) and homograph (Anbinderis and Kasparaitis, 2009) stressing rules.

In several studies, researchers have attempted to synthesize text with stress. These previous studies can be divided into two parts: the stressed sounds are created during the synthesis process by modifying their duration and fundamental frequency, or the stressed sounds are stored separately, and then their concatenation with unstressed sounds was performed in order to obtain the synthesized text with the stress. An example of stressed text synthesis employing modifications during the synthesis process is given in the paper by Pyž *et al.* (2011). The authors created a formant synthesizer version of the Lithuanian word "laimė" ('happiness') and its cases. In order to make the synthesized signal sound more natural, the authors used impulses of different amplitudes and periods as inputs instead of the unit impulses with a constant period. In a later publication by the same authors (Pyž *et al.*, 2014), models of stressed and unstressed Lithuanian speech vowels and semivowels were developed. The change of fundamental frequencies was also modelled by dynamic inputs of multiple-input and single-output systems. However, a formant synthesizer for Lithuanian has not been developed yet. The method proposed by the authors requires more extensive research to check whether it is suitable for other Lithuanian words. In the case of concatenation synthesis, significant results have been achieved by Kasparaitis in collaboration with experts on the Lithuanian language, mainly Girdenis. The synthesizer uses Lithuanian speech corpora developed in advance. The Fujisaki model and Lithuanian diphone databases in the MBROLA format are used (Kasparaitis, 2005). This synthesizer was used by the website Text-Talk (Text-Talk, 2006) until 2010. Recently, with the growing interest in neural networks, some research related to neural network-based speech synthesis on the Lithuanian language with stress was published. Nvidias implementation of the Tacotron 2 system for Lithuanian speech synthesis is used in the paper by Radzevičius *et al.* (2021). The stressed-text speech corpus was used to train a model. The results show the advantages of using the stressed text as sample labels in comparison to models trained on a dataset with raw (not stressed) sample labels. The possibilities of using Long Short-Term Memory (LSTM) networks for Lithuanian speech synthesis are discussed in a paper by Navickas *et al.* (2019). The authors concluded that LSTM networks are suitable for the synthesis of Lithuanian speech, and they are going to use this method for speech synthesis and later for adding intonation to synthesized speech.

Based on analysed papers, it can be concluded that both word- and sentence-level phenomena received some but not much attention from researchers. The reviewed studies cover the topic in quite a fragmented manner. In order to draw a full picture of Lithuanian intonation modelling, papers related to the stress of words and phone duration models that address the issue of pitch-accents and sentence-level intonation, respectively, were also analysed in this section. A summary of the work on intonation modelling, in terms of prosodic phenomena and methods used, is given in Table 1.

Table 1
Lithuanian Intonation modelling.

Reference	Prosodic phenomena	Method
Sentence-level intonation modelling		
Vaičiūnas *et al.* (2016)	Computational model of the fundamental frequency	Hidden Markov model (HMM)
Raškinis and Kazlauskienė (2013)	Characteristic patterns of Lithuanian intonation	Clustering of pitch contours
Leonavičius (2006)	Models of melismas pitch variations	Artificial neural networks
Norkevičius *et al.* (2006), Norkevičius and Raškinis (2008)	Phone duration models	Classification and regression tree (CART)
Kasparaitis and Beniušė (2016)	Phone duration models	Rule-based prediction techniques
Word-level intonation modelling		
Dogil and Möhler (1998)	Pitch-accent model	Polynomial function
Paulikas and Navakauskas (2005)	General pitch-accent model	Polynomial function
Kasparaitis (2000, 2001)	Text stressing	Morphological analysis
Anbinderis (2010a, 2010b)	Text stressing	Morphological analysis supplemented with clitic and homograph stressing rules
Kasparaitis (2005)	Text stressing	Concatenation synthesis
Pyž *et al.* (2011, 2014)	Text stressing	Formant synthesis
Radzevičius *et al.* (2021)	Text stressing	Artificial neural networks

6. Discussion and Conclusions

This paper provides a first interdisciplinary overview of existing research carried out on Lithuanian intonation. It shows that in the last three decades, several attempts have been made to tackle this topic both from the linguistic perspective and the perspective of computer-aided modelling. But perhaps unsurprisingly, the study of such a complex phenomenon with its inherent ambiguities resulted in a rather fragmented research. Moreover, it is important to note the lack of cooperation between speech engineers working on technological applications such as automatic speech synthesis and linguists. This clearly hinders the development of a common theory of Lithuanian intonation. The above-presented insights reveal that the applicability of linguistic findings to real-life modelling is quite limited.

6.1. *The Linguistic Perspective*

The above-reviewed phonetic research on Lithuanian intonation mostly concentrated on sentence types and only occasionally examined other phenomena, such as focus. These studies made the first attempt to use ToBi in order to provide a clear and more generalizable description of intonational patterns in Lithuanian that would be understandable to the international research community. However, the system is not used thoroughly, and only the major principles are employed without making use of all the functionalities that would allow accounting for more detailed aspects of Lithuanian intonation. The phonological work by Kundrotas, on the other hand, is very rich in the documentation of a variety of intonational patterns and their variations. Moreover, he provides an in-depth analysis of possible uses of these patterns across linguistic styles and Lithuanian dialects. Nevertheless, the theoretical approach adopted in this research, specifically the holistic view of indecomposable melodic contours, has a number of issues. For instance, it does not allow to capture the diversity of phonetic realizations of the same contour and complicates the cross-linguistic comparison of Lithuanian intonation to intonation patterns of other languages. Last but not least, the framework used for the phonological descriptions of Lithuanian intonation can hardly be understood and used by the community of computational linguists and speech engineers. Thus, there is an urgent need to apply well-known frameworks and models (both phonological and computational) to further describe the phonetics of Lithuanian intonation and provide consistent phonological generalizations.

The above-described evidence about Lithuanian word-level and sentence-level prosodic features brings up an important question about the place of Lithuanian in the intonational typology and the practical implications it can have on its modelling. As mentioned before, from the point of view of word-prosodic typology, Lithuanian is considered to a be a pitch-accent language and thus word-level prosodic specificities should influence the description of its prosodic characteristics at the sentence level. However, from the acoustic point of view, the main phonetic correlates of accent variations in the acute and the circumflex patterns have been shown to be intensity and duration, and not F0, which is used as a secondary marker only. Conversely, F0 variations best reflect sentence intonation. Thus, as the difference between the two contours lay in the location of the accent within the bimoraic syllable (in an acute syllable – on the left-hand mora, in a circumflex syllable – on the right) rather than in its tonal quality, the distinction in the notation could be reduced to a single accent mark with an indication of place (Kushnir, 2019). In this way, if sentence intonation is interpreted as being primarily marked by F0 variations, then Lithuanian could be considered to be closer to stress languages, where intonation is not defined at the word-level. This would certainly facilitate Lithuanian intonation modelling. Note, though, that in Lithuanian, the secondary markers of sentence intonation (duration and intensity) correspond to the primary markers of its accent. Thus, possible interactions could arise from this overlap, which should be examined in future studies. Based on these overall observations, we can return to the reflection on whether it would be preferable to think of languages as forming a continuous scale between tone and stress languages instead of defining them in terms of three distinct categories (Hirst and Di Cristo, 1998).

6.2. *The Modelling Perspective*

From the modelling perspective, there have been very few attempts to develop intonation models for Lithuanian compared to other languages, especially so for sentence-level intonation. The reviewed studies cover the topic in quite a fragmented manner. Regarding word-level intonation, there are more papers that have approached this issue. However, there are still many unresolved tasks. It is also important to note that articles on modelling were published quite a long time ago, and this reaffirms that the subject is complex and that it does not attract sufficient attention from the research community. Turning to speech synthesis, the most developed text-to-speech system for Lithuanian is based on the concatenation of natural speech segments stored in a database. However, worldwide dominant phonological intonation models have not been used by researchers working on Lithuanian speech synthesis either.

The possibilities for the modelling of Lithuanian intonation should be explored in light of the well-known models introduced above. First of all, there is one common feature to the models presented (except the AM model). Specifically, the Fujisaki, the Tilt and IPO models are formulated in terms of acoustics: intonation is described as a variation of the fundamental frequency F0, which can be expressed as a sequence of atomic variation elements (e.g. the F0 value rises and falls). On the one hand, this makes all these models purely descriptive: we can analyse and describe only the acoustic realization of intonation. The models do not have phonological justifications, which means that we do not have any mapping between acoustic events (F0 variation associated with both local and sentence-level intonation events) and linguistic structures. The absence of this mapping results in the inability of these models to synthesize intonation based on linguistic information (e.g. text).

On the other hand, model independence from phonological differences and subtleties of different languages makes these models potentially adaptable to different languages. As none of the three models (Fujisaki, IPO model or Tilt model) has a phonological part, some mapping rules can be integrated into them. The PENTA model contains a parallel encoding part, defining encoding functions as lexical or emotional, which can be considered as language-specific. The potential adaptability of the models can be illustrated by examples of the application of the Fujisaki model covering both stress languages such as English (Moberg and Parssinen, 2004), German (Mixdorff, 2000), Italian (Rossi *et al.*, 2002), and pitch-accent languages such as Japanese, and Swedish (Ljungqvist and Fujisaki, 1993). There have been fewer applications of the other models, which suggests that there is more experience with the Fujisaki model, making it a better alternative for the Lithuanian intonation modelling.

Considering more phonologically-oriented models, neither AM nor any other major phonological models of intonation have been applied to Lithuanian. It is true that in order to make phonological generalizations, it is necessary to already have a certain level of phonetic analysis of the language. This would allow researchers to identify which parts of the acoustic signal can be assigned to more abstract contrastive phonological units (Zerbian, 2010). Phonetic research on Lithuanian has started, but there are certainly much more researches to be done. Nevertheless, the existing material could already be used in applying the Autosegmental-Metrical model (AM), which has been successfully applied to many typologically different languages (Jun, 2005). The analysis of melodic contours based on smaller discrete linguistic primitives would help to solve the issue of variability in the realizations raised above. Importantly, AM would allow us to finally examine in greater depth the relationship between Lithuanian sentence-level intonation and the Lithuanian pitch accent. Specifically, in this framework, both word-level and sentence-level intonation is treated as a sequence of the same type of tones whose function is specified in each language's grammar. Thus, a single model could be used to describe both levels of intonation.

Note, though, that some researchers suggest choosing models from languages that are close to Lithuanian based on morphological rather than phonological characteristics. For example, Vaičiūnas *et al.* (2016) propose to investigate models of highly inflectional languages such as Polish, Czech or Croatian. On the other hand, considering the language-specific features of intonation, it is possible that models developed for other languages will not be suitable for Lithuanian. In this case, certain modifications to match the characteristics of the Lithuanian language ought to be made. Unfortunately, according to our knowledge, there are no comparative studies to this day that would identify language-specific and language-universal elements of Lithuanian intonation.

6.3. *Prospects of Using New Modelling Approaches*

It is important to note that the main disadvantage of the models described above is their requirement of much a priori knowledge about the phenomenon to be modelled, or "expert knowledge". Therefore, nowadays, scientists have come up with new language prosody modelling approaches, where models are able to learn the patterns from speech data. Specifically, these are models based on neural networks. The main difference from traditional ones is that the networks are able to learn patterns, i.e. the algorithms themselves construct rules when analysing the data, and then construct patterns based on the rules observed. In such a way, the algorithms learn the patterns from the raw audio data. Models based on deep learning show good results in intonation modelling and could be a solution, but it should be noted that these methods require large amounts of annotated training material. WaveNet network (Oord *et al.*, 2016) is one of the successful examples. However, this solution performs poorly with under-resourced languages and especially with languages for which intonation needs to be modelled at both word and sentence levels. A solution could be to use advanced deep transfer learning techniques (Wang and Zheng, 2015) to extract intonation patterns. The idea of the approach is to train the network for intonation modelling on the basis of one language, then drop the last trained layers and retrain those exclusively with the specific language for which the intonation model needs to be built, e.g. Lithuanian speech data, expecting the model will be capable of modelling the intonation at the word and sentence levels. As Vaičiūnas *et al.* (2016) state, the set of contextual factors that was good for English appeared to be useful for Lithuanian, as well. However, the authors' assumption that the Lithuanian intonation model can be improved by subjective factors – theme, rheme and logical stress – was not confirmed. As Lithuanian is an under-resourced language, creating an extensively annotated dataset is problematic, but it could benefit the traditional intonation modelling methods. For this reason, attempts to label prosodic constituents are automatically of the most significant importance. There is only one report published (Raškinis and Kazlauskienė, 2013) a while ago showing promising results in this type of intonation prediction.

6.4. *The Interdisciplinary Benefits of Intonational Research*

As stated, Lithuanian is spoken by less than four million speakers worldwide and can therefore be considered an under-resourced language. As a result, it is very difficult to build large enough corpora for speech recognition and synthesis purposes. Due to the small number of Lithuanian speakers, large language technology companies, such as Google, are not interested in investing in it. Therefore, it is largely up to scientists and local businesses to develop fundamental and applied research on the Lithuanian language and its modelling. This is valid for almost all under-resourced languages (Besacier *et al.*, 2014).

Intonation models could contribute to the theoretical description of Lithuanian and also widen the typological understanding of Lithuanian among other languages. It has the potential to help to identify and classify a wider range of intonational patterns. Moreover, the interrelation of acoustic correlates other than F0 that are used in the realization of intonation patterns could be uncovered.

Intonation models would also benefit Lithuanian speech technologies. First, they would fill in the current theoretical gap and provide the necessary theoretical description of the Lithuanian language, which would be directly applicable to language technologies. Second, it would increase the accuracy of speech recognition and help to identify speakers' emotions and dialects. Lithuanian has 2 dialects and 6 subdialects which have differences in their phonetics, syntax, morphology and lexis. But the main differences lie in their vocalism and accent-intonation systems (Girdenis and Zinkevičius, 1966), which could be reflected in the created intonation models. Third, intonation models would help to increase the language naturalness of synthetic speech and foster ease of processing for listeners.

However, the main problem, which must be solved in the future, is the interdisciplinary nature of speech research and modelling. People working in this field should have a background in two main fields – linguistics and computer science – while knowledge of psycholinguistics would clearly be beneficial, as well. Until now, most experts working on Lithuanian language modelling had a background in only one of the fields, and Lithuanian universities clearlty lack such an interdisciplinary curriculum. Interdisciplinary collaborations would benefit not only the theoretical description but also the application of language technologies. Finally, the use of widely recognized state-of-the-art theoretical frameworks and methods, as well as the use of the English language in publishing, would contribute to attracting international researchers to further develop this topic.

Turning back to the Lithuanian language, there are still many important issues that have not been addressed in order to generate contours of Lithuanian intonation for a given input and which can be considered as future challenges. First of all, the prominent intonation models reported in this paper, i.e. Fujisaki, IPO, tilt, PENTA, and AM models, could be applied to Lithuanian to evaluate their suitability for this language. Since there are no publications on this issue, it may be possible that these models have not been investigated at all. It is only known that the Fujisaki model was used in the speech synthesizer developed by Kasparaitis and his colleagues. Unfortunately, the synthesizer was used on the Text-Talk website until 2010, and the details of the fundamental frequency modelling are not published by the authors.

6.5. *Directions for Future Research*

Considering the above-described issue, we can formulate potential directions for Lithuanian language intonation modelling as follows:

- Further phonetic and phonological description of intonation patterns in Lithuanian. This work should be based on widely acknowledged theoretical frameworks, which leads to the following point.
- Phonology-based modelling of intonation. In this direction, phonology knowledge (intonation, stress, etc.) in the form of rules and restrictions would be integrated or joined with the selected acoustic intonation model. The possible result of such integration would be rule-based intonation models. An inevitable condition for this work is interdisciplinary collaboration.
- Data-driven intonation models. They would be obtained by applying statistical data analysis techniques. In these models, the concept of intonation patterns could be established and explored. The presence of such intonation patterns and their sets would mean the possibility to express all the intonation sequences in a time-aligned combination of patterns. The synthesis task would be to select and join particular patterns into a new intonation sequence (phrase, sentence, etc.). The formulation of data-driven models will require large labelled datasets, again hardly possible without a multidisciplinary collaboration.
- Creation of a Lithuanian intonation corpus with the help of automatic annotation techniques. This approach would necessitate a close collaboration with expert phoneticians. The topic could be further explored with the help of supervised machine-learning techniques to discover new patterns.

Overall, as linguistic research on intonation was a precursor to modelling research on the topic, descriptive linguistic literature can bring valuable evidence for modelling the main acoustic correlates of intonation, its types and functions in the language. Linguistics, on the other hand, could benefit from modelling research through the identification and classification of more types of intonation patterns. Moreover, the use of a clear, single and internationally-recognized framework would facilitate the collaboration between researchers in Lithuania and abroad, as well as the inclusion of the computational modelling community.

Ethics Statement

No ethics approval was required given that no new data were collected, nor new analyses were conducted on existing data.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Abramson, A.S. (1962). The vowels and tones of standard Thai: acoustical measurements and experiments. *International Journal of American Linguistics*, 28(2–3).

Alvarez, A.A., Issa, E.S.A. (2020). Learning Intonation Pattern Embeddings for Arabic Dialect Identification. arXiv:2008.00667.

Alvarez, A.A., Issa, E., Alshakhori, M. (2022). Computational modeling of intonation patterns in Arabic emotional speech. In: *Proceedings Speech Prosody 2022*, pp. 615–619.

Ambrazas, V., Garšva, K., Girdenis, A., Jakaitienė, E., Kniūkšta, P., Krinickaitė, S., Labutis, V., Laigonaitė, A., Oginskienė, E., Pikčilingis, J., Ružė, A., Sližienė, N., Ulvydas, K., Urbutis V., Valeckienė, A., Valiulytė, E. (1996). A Grammar of Modern Lithuanian. 2nd ed. Mokslo ir enciklopedijų leidykla, Vilnius.

Anbinderis, T. (2010a). Automatic stressing of Lithuanian text using decision trees. *Information Technology and Control*, 39(1), 61–67.

Anbinderis, T. (2010b). *Kai kurių lietuvių kalbos teksto kirčiavimo aspektų matematinis modeliavimas [Mathematical Modelling of Some Aspects of Stressing a Lithuanian Text]*. PhD thesis, Vilnius university.

Anbinderis, T., Kasparaitis, P. (2007). Algorithms for detecting clitics in the Lithuanian text. *Studies about Languages*, 10, 30–37.

Anbinderis, T., Kasparaitis, P. (2009). Disambiguation of Lithuanian homographs based on the frequencies of lexemes and morphological tags. *Studies about Languages*, 14, 25–31.

Andruski, J., Costello, J. (2004). Using polynomial equations to model pitch contour shape in lexical tones: an example from Green Mong. *Journal of the International Phonetic Association*, 34(2), 125–140.

Arvaniti, A. (2016). Analytical decisions in intonation research and the role of representations: lessons from Romani. *Laboratory Phonology. Journal of the Association for Laboratory Phonology*, 7(1), 1–43. https://doi.org/10.5334/labphon.14.

Arvaniti, A. (2022). The autosegmental-metrical model of intonational phonology. In: Shattuck-Hufnagel, S., Barnes, J. (Eds.), *Prosodic Theory and Practice*. MIT Press, Cambridge, MA.

Arvaniti, A., Ladd, D.R. (2009). Greek wh-questions and the phonology of intonation. *Phonology*, 26(1), 43–74. https://doi.org/10.1017/S0952675709001717.

Balčiūnienė, I., Simonavičienė, L. (2009). Kiekybinis klausiamųjų šnekamosios lietuvių kalbos pasakymų tyrimas [A Quantitative Study of Listening to Spoken Lithuanian Narratives]. *Lietuvių kalba*, 3, 272–277.

Balkevičius, J. (1963). *Dabartinės lietuvių kalbos sintaksė [The Syntax of Modern Lithuanian Language]*. State Publishing House of Political and Scientific Literature, Vilnius.

Balkevičius, J. (1998). *Lietuvių kalbos predikatinių konstrukcijų sintaksė [Syntax of Lithuanian Predicate Constructions]*. Science and Encyclopaedia Publishing Centre, Vilnius.

Beckman, M.E., Ayers, G. (1994). *Guidelines for ToBI Labelling*. Online MS and accompanying files. Available at http://www.ling.ohio-state.edu/phonetics/E_ToBI.

Beckman, M.E., Hirschberg, J.B., Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In: Jun, S.-A. (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford Scholarship Online. https://doi.org/10.1093/acprof:oso/9780199249633.003.0002.

Besacier, L., Barnard, E., Karpov, A., Schultz, T. (2014). Automatic speech recognition for under-resourced languages: a survey. *Speech Communication*, 56, 85–100. https://doi.org/10.1016/j.specom.2013.07.008.

Biadsy, F., Hirschberg, J.B. (2009). Using prosody and phonotactics in Arabic dialect identification. In: *Tenth Annual Conference of the International Speech Communication Association*. https://doi.org/10.7916/D8HM5HRV.

Birkholz, P., Zhang, X. (2020). Accounting for microprosody in modeling intonation. In: *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8099–8103. https://doi.org/10.1109/ICASSP40776.2020.9054149.

Blevins, J. (1993). A tonal analysis of Lithuanian nominal accent. *Language*, 69(2), 237–273.

Boidin, C., Boeffard, O. (2008). Modeling intonation variability with HMM for speech synthesis. In: *Proceedings of Speech Prosody*, Brazil, pp. 115–118.

Bolinger, D. (1964). Around the edge of language: intonation. *Harvard Educational Review*, 34(2), 282–296.

Botinis, A., Granström, B., Möbius, B. (2001). Developments and paradigms in intonation research. *Speech Communication*, 33(4), 263–296. https://doi.org/10.1016/S0167-6393(00)00060-1.

Bruce, G. (1977). *Swedish Word Accents in Sentence Perspective*. CWK Gleerup.

Bruce, G., Schötz, S., Granström, B. (2007). SIMULEKT – modelling Swedish regional intonation. In: *Proceedings of Fonetik (Swedish Phonetics Conference)/TMH-QPSR*, pp. 53–56.

Cahn, J.E. (1990). *Generating Expression in Synthesized Speech*. MIT Technical report.

Carlson, R., Granström, B., Nord, L. (1992). Experiments with emotive speech – acted utterances and synthesized replicas. In: *Proceedings of the 2nd International Conference of Spoken Language Processing (ICSLP 92)*, Canada, October 12–16, pp. 671–674.

Chakrasali, S.V., Indira, K., Narasimhaiah, S.Y. (2022). Performance analysis of different intonation models in Kannada speech synthesis. *Indonesian Journal of Electrical Engineering and Computer Science*, 26(1), 243–252. https://doi.org/10.11591/ijeecs.v26.i1.pp243-252.

Chomsky, N. (1956). Three models for the description of language. *IEEE Transactions on Information Theory*, 2(3), 113–124. https://doi.org/10.1109/TIT.1956.1056813.

Connell, B.A., Hogan, J.T., Rozsypal, A.J. (1983). Experimental evidence of interaction between tone and intonation in Mandarin Chinese. *Journal of Phonetics*, 11(4), 337–351. https://doi.org/10.1016/s0095-4470(19)30834-4.

Cutler, A., Dahan, D., Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: a literature review. *Language and Speech*, 40(2), 141–201. https://doi.org/10.1177/002383099704000203.

Dogil, G., Möhler, G. (1998). Phonetic invariance and phonological stability: Lithuanian pitch accents. In: *5th International Conference on Spoken Language Processing*, Sydney, Australia, November 30–December 4, 1998. http://www.mirlab.org/conference_papers/International_Conference/ICSLP1998/PDF/AUTHOR/SL980206.PDF.

Dogil, G., Williams, B. (1999). The phonetic manifestation of word stress. In: van der Hulst, H. (Ed.), *Word Prosodic Systems in the Languages of Europe*. de Gruyter, Berlin, pp. 273–334.

Fan, Y., Qian, Y., Xie, F., Soong, F.K. (2014). TTS synthesis with bidirectional LSTM based recurrent neural networks. In: *Interspeech*, pp. 1964–1968.

Féry, C. (2017). Intonation and prosodic structure. (Key Topics in Phonology.) Cambridge: Cambridge University Press. Pp. xi 374. *Phonology*, 36(1), 171–179 https://doi.org/10.1017/S0952675719000071.

Fujisaki, H., Hirose, K. (1984). Analysis of voice fundamental frequency contours for declarative sentences of Japanese. *Journal of the Acoustical Society of Japan (E)*, 5(4), 233–242.

Gerazov, B., Ivanovski, Z., Bilibajkić, R. (2010). Modeling Macedonian intonation for text-to-speech synthesis. In: *Proceedings of the DOGS*, pp. 16–18.

Girdenis, A. (2003). *Theoretical Foundations of Lithuanian Phonology*. Eugrimas.

Girdenis, A., Zinkevičius, Z. (1966). Dėl lietuvių kalbos tarmių klasifikacijos [Regarding the Classification of Lithuanian Dialects]. *Kalbotyra*, 14, 139–147. https://doi.org/10.15388/Knygotyra.1966.18940.

Grabe, E., Karpinski, M. (2003). Universal and language-specific aspects of intonation in English and Polish. In: *Proceedings of the 15th International Congress of Phonetic Sciences*, Vol. 39, pp. 1061–1064.

Grabe, E., Nolan, F., Farrar, K.J. (1998). IViE-A comparative transcription system for intonational variation in English. In: *Fifth International Conference on Spoken Language Processing*, paper 0099.

Grabe, E., Kochanski, G., Coleman, J. (2007). Connecting intonation labels to mathematical descriptions of fundamental frequency. *Language and Speech*, 50(3), 281–310.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge University Press.

Gussenhoven, C. (2016). Analysis of intonation: the case of MAE-ToBI. *Laboratory Phonology*, 7(1), 1–35. https://doi.org/10.5334/labphon.30.

Hamlaoui, F., Żygis, M., Engelmann, J., Wagner, M. (2019). Acoustic correlates of focus marking in Czech and Polish. *Language and Speech*, 62(2), 358–377. https://doi.org/10.1177/0023830918773536.

Halle, M., Vergnaud, J.R. (1987). *An Essay on Stress*. The MIT Press.

Hallé, P.A., De Boysson-Bardies, B., Vihman, M.M. (1991). Beginnings of prosodic organization: intonation and duration patterns of disyllables produced by Japanese and French infants. *Language and Speech*, 34(4), 299–318. https://doi.org/10.1177/002383099103400401.

Hart, J.T., Collier, R., Cohen, A. (1990). *A Perceptual Study of Intonation*. Cambridge University Press. https://doi.org/10.1017/CBO9780511627743.

Haspelmath, M.B., Bickel, B. (2008). *The Leipzig Glossing Rules: Conventions for Interlinear Morpheme-by-Morpheme Glosses*. Department of Linguistics of the Max Planck Institute for Evolutionary Anthropology & the Department of Linguistics of the University of Leipzig.

Hedberg, N., Sosa, J.M., Görgülü, E., Mameni, M. (2010). The prosody and meaning of wh-questions in American English. In: *Proceedings of the International Conference on Speech Prosody*, pp. 6–9.

Hedberg, N., Sosa, J.M., Görgülü, E. (2017). The meaning of intonation in yes-no questions in American English: a corpus study. *Corpus Linguistics and Linguistic Theory*, 13(2), 321–368. https://doi.org/10.1515/cllt-2014-0020.

Hirst, D.J., Espesser, R. (1993). Automatic modelling of fundamental frequency using a quadratic spline function. *Travaux de l'Institut de Phonétique d'Aix*, 15, 75–85.

Hirst, D., Di Cristo, A. (1998). A survey of intonation systems. In: Hirst, D., Di Cristo, A. (Eds.), *Intonation Systems: A Survey of Twenty Languages*. Cambridge University Press, New York, pp. 1–44.

Hock, H.H. (2015). Prosody and dialectology of tonal shifts in Lithuanian and their implications. In: *Contemporary Approaches to Baltic Linguistics*, Vol. 276, pp. 111–137. https://doi.org/10.1515/9783110343953-003.

Hodari, Z., Watts, O., King, S. (2019). Using generative modelling to produce varied intonation for speech synthesis. In: *Proceedings of the 10th ISCA Speech Synthesis Workshop*, pp. 239–244. https://doi.org/10.21437/SSW.2019-43.

Hodari, Z., Moinet, A., Karlapati, S., Lorenzo-Trueba, J., Merritt, T., Joly, A., Abbas, A., Karanasou, P., Drugman, T. (2021). Camp: a two-stage approach to modelling prosody in context. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6578–6582. https://doi.org/10.1109/ICASSP39728.2021.9414413.

Honnet, P.E.J.C. (2017). *Intonation Modelling for Speech Synthesis and Emphasis Preservation*. Thesis No. 7520, École Polytechnique Fédérale De Lausanne.

Honnet, P.-E., Garner, P.N. (2016). Emphasis recreation for TTS using intonation atoms. In: *Proccedings of 9th ISCA Workshop on Speech Synthesis Workshop (SSW 9)*, pp. 14–20. https://doi.org/10.21437/SSW.2016-3.

Hualde, J.I., Riad, T. (2014). Word accent and intonation in Baltic. In: *Proceedings of the International Conference on Speech Prosody*, pp. 668–672. https://doi.org/10.21437/speechprosody.2014-121.

Joseph, J.E. (2009). Why Lithuanian accentuation mattered to Saussure. *Language & History*, 52(2), 182–198. https://doi.org/10.1179/175975309X452067.

Hyman, L.M. (2006). Word-prosodic typology. *Phonology*, 23(2), 225–257. https://doi.org/10.1017/S0952675706000893.

Janyoi, P., Seresangtakul, P. (2020). Tonal contour generation for Isarn speech synthesis using deep learning and sampling-based F0 representation. *Applied Sciences*, 10(18), 6381. https://doi.org/10.3390/app10186381.

Jun, S.A. (Ed.) (2005). *Prosodic Typology: The Phonology of Intonation and Phrasing*. OUP Oxford. https://doi.org/10.1093/acprof:oso/9780199249633.001.0001.

Jusczyk, P.W., Hirsh-Pasek, K., Kemler Nelson, D.G., Kennedy, L.J., Woodward, A., Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, 24(2), 252–293.

Kasparaitis, P. (2000). Automatic stressing of the Lithuanian text on the basis of a dictionary. *Informatica*, 11(1), 19–40. https://doi.org/10.3233/INF-2000-11103.

Kasparaitis, P. (2001). Automatic stressing of the Lithuanian nouns and adjectives on the basis of rules. *Informatica*, 12(2), 315–336. https://doi.org/10.3233/INF-2001-12210.

Kasparaitis, P. (2005). Diphone databases for Lithuanian text-to-speech synthesis. *Informatica*, 16(2), 193–202. https://doi.org/10.15388/informatica.2005.093.

Kasparaitis, P., Beniušė, M. (2016). Automatic parameters estimation of the D. Klatt phoneme duration model. *Informatica*, 27(3), 573–586. https://doi.org/10.15388/Informatica.2016.100.

Kazlauskienė, A. (2012). *Bendrinės lietuvių kalbos akcentologijos pagrindai [Fundamentals of the Standard Lithuanian Language Accentology]*. Vytautas Magnus University Press, Kaunas.

Kazlauskienė, A., Dereškevičiūtė, S. (2018). The intonational patterns of interrogative sentences in Lithuanian. In: *Salento University Linguistic Symposium*.

Kazlauskienė, A., Sabonytė, R. (2018). F0 in Lithuanian: the indicator of stress, syllable accent, or intonation? In: Muischnek, K., Müürisep, K. (Eds.), *Human Language Technologies – The Baltic Perspective*, Vol. 307, pp. 55–62. https://doi.org/10.3233/978-1-61499-912-6-55.

Klatt, D.H. (1979). Synthesis by rule of segmental durations in English sentences. In: Lindblom, B., Ohman, S. (Eds.), *Frontiers of Speech Communication Research*. Academic Press, New York, pp. 287–300.

Kohler, K.J. (1991). Prosody in speech synthesis: the interplay between basic research and TTS application. *Journal of Phonetics*, 19, 121–138.

Krapikaitė, N. (2009). *Pagrindiniai fonetiniai Lietuvių kalbos intonacijos požymiai* [*The Main Phonetic Features of Intonation in Lithuanian*]. MA thesis, Vytautas Magnus University.

Krapikaitė, N. (2011). Pagrindinis tonas – svarbiausias funkcinių frazių skiriamasis požymis [Pitch as the Main Distinguishing Feature of Functional Phrases]. In: *Laikas ir žodis: studentų mokslo darbai*. Vytautas Magnus University Press, pp. 55–60.

Krapikaitė, N. (2014). Using the ToBI transcription to analyze the intonation of Lithuanian. In: *Human Language Technologies – The Baltic Perspective: Proceedings of the 6th International Conference, Baltic HLT 2014*, pp. 202–205.

Krapikaitė, N. (2015). The adaptation of the ToBI system for Lithuanian pitch change. *Žmogus ir žodis*, 17(1), 67–75. https://doi.org/10.15823/zz.2015.5.

Kuczmarski, T. (2021). *Modeling of Polish Intonation for Statistical-Parametric Speech Synthesis*. PhD Thesis, Institute of Ethnolinguistics.

Kundrotas, G. (2008). Lietuvių kalbos intonacinių kontūrų fonetiniai požymiai (eksperimentinis-fonetinis tyrimas) [Acoustic Characteristics of Lithuanian Intonation Contours (Experimental Phonetic Research)]. *Žmogus ir žodis. Didaktinė Lingvistika*, 10(1), 43–55.

Kundrotas, G. (2009). *Lyginamoji lietuvių ir rusų kalbų intonacinių sistemų analizė [Comparative Analysis of Lithuanian and Russian Intonation Systems]*. Vilniaus pedagoginio universiteto leidykla, Vilnius.

Kundrotas, G. (2017). Lietuvių kalbos intonacijos tyrimo apžvalga [A Review of Research into the Intonation of the Lithuanian Language]. *Lituanistica*, 4(4), 245–254.

Kundrotas, G. (2020). Lithuanian language intonation: history of research, in the context of language intonology. *Językoznawstwo*, 1(14), 195–204. https://doi.org/10.25312/2391-5137.14/2020_12gk.

Kushnir, Y. (2019). *Prosodic Patterns in Lithuanian Morphology*. PhD thesis, Universität Leipzig.

Ladd, D.R. (1996). *Intonational Phonology*. Cambridge University Press, Cambridge.

Ladd, D.R. (2001). Intonational universals and intonational typology. In: *Language Typology and Language Universals: An International Handbook*, pp. 1380–1390.

Ladd, D.R. (2008). *Intonational Phonology*. 2nd ed. Cambridge University Press, Cambridge.

Laigonaitė, A. (1958). Dėl lietuvių kalbos kirčio ir priegaidės supratimo [On the Understanding of the Stress and Pitch-Accent of the Lithuanian Language]. *Kalbotyra I*, 23–27.

Leonavičius, R. (2006). *Melisma Synthesis Using Artificial Neural Networks*. PhD thesis, Vilnius Gediminas Technical University.

Lezhenin, I., Diachkov, V., Lamtev, A., Zhuikov, A., Bogach, N., Boitsova, E., Pyshkin, E. (2018). Automatic intonation-based keyword extraction from academic discourse. In: *2018 Federated Conference on Computer Science and Information Systems (FedCSIS)*. IEEE, pp. 165–168.

Ljungqvist, M., Fujisaki, H. (1993). Generating intonation for Swedish text-to-speech conversion using a quantitative model for the F0 contour. In: *Proceedings of the 3rd European Conference on Speech Communication and Technology EUROSPEECH*, pp. 873–876.

Liberman, M. (1975). *The Intonational System of English*. PhD thesis, MIT, Cambridge/MA.

Lieberman, P. (1965). On the acoustic basis of the perception of intonation by linguists. *WORD*, 21(1), 40–54. https://doi.org/10.1080/00437956.1965.11435417.

Liberman, M., Sag, I. (1974). Prosodic form and discourse function. In:*Papers from the Tenth Regional Meeting, CLS 10*, pp. 416–427.

Liu, H. (2017) *Fundamental Frequency Modelling: An Articulatory Perspective with Target Approximation and Deep Learning*. PhD thesis, University College London.

Selkirk, E.O. (1978). On prosodic structure and its relation to syntactic structure. In: Fretheim, T. (Ed.), *Nordic Prosody II*. TAPIR, Trondheim, pp. 268–271.

Malisz, Z., Żygis, M. (2017). Special issue: slavic perspectives on prosody. *Phonetica*, 73(3–4), 155–162. https://doi.org/10.1159/000449430.

Marelli, F., Schnell, B., Bourlard, H., Dutoit, T., Garner, P.N. (2019). An end-to-end network to synthesize intonation using a generalized command response model. In: *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7040–7044. https://doi.org/10.1109/ICASSP.2019.8683815.

Michelini, G. (2000). Le lituanien, la plus archaïque des langues indo-européennes modernes. *Cahiers Lituaniens*, 2, 28–36.

Mikalauskaitė, E. (1975). *Lietuvių kalbos fonetikos darbai [Studies in Lithuanian Phonetics]*. Mokslas, Vilnius.

Mikoś, M.J. (1976). Intonation of questions in Polish. *Journal of Phonetics*, 4(3), 247–253. https://doi.org/10.1016/s0095-4470(19)31247-1.

Ming, H., Huang, D.-Y., Xie, L., Wu, J., Dong, M., Li, H. (2016). Deep bidirectional LSTM modeling of timbre and prosody for emotional voice conversion. In: *Interspeech*, pp. 2453–2457.

Mixdorff, H. (2000). A novel approach to the fully automatic extraction of Fujisaki model parameters. In: *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing (Cat. No. 00CH37100)*, Vol. 3, pp. 1281–1284, IEEE. https://doi.org/10.1109/ICASSP.2000.861811.

Moberg, M., Parssinen, K. (2004). Comparing CART and Fujisaki intonation models for synthesis of US-English names. In: *Speech Prosody 2004*, pp. 439–442.

Navickas, G., Korvel, G., Bernatavičienė, J. (2019). Overview of speech synthesis using LSTM neural networks. In: *Computer Data Analysis and Modeling: Stochastics and Data Science: Proceedings of the Twelfth International Conference*, Minsk, September 18–22, 2019. BSU, Minsk, pp. 257–261.

Ni, J., Shiga, Y., Hori, C. (2016). Superpositional HMM-based intonation synthesis using a functional F0 model. *Journal of Signal Processing Systems*, 82, 273–286.

Norkevičius, G., Kazlauskienė, A., Raškinis, G. (2006). Garsų trukmės modeliavimas naudojant klasifikavimo ir regresijos medžius [Decision Trees in Phoneme's Duration Modelling]. In: *Informacinės technologijos 2006*, pp. 82–85.

Norkevičius, G., Raškinis, G. (2008). Modeling phone duration of Lithuanian by classification and regression trees, using very large speech corpus. *Informatica*, 19(2), 271–284. https://doi.org/10.15388/informatica.2008.213.

Oord, van den Dieleman S, A., Zen, H., Simonyan, K., Vinyals, O., Graves, Kalchbrenner N, A., Senior, A., Kavukcuoglu, K. (2016). Wavenet: a generative model for raw audio. arXiv:1609.03499.

Pakerys, A. (2003). *Lietuvių bendrinės kalbos fonetika [Phonetics of Lithuanian Language]*. 3rd ed. Encyclopedia, Vilnius.

Paulikas, Š., Navakauskas, D. (2005). Restoration of voiced speech signals preserving prosodic features. *Speech Communication*, 47(4), 457–468. https://doi.org/10.1016/j.specom.2005.05.002.

Petit, D. (2020). New insights on Lithuanian accentuation from the unpublished manuscripts of Ferdinand de Saussure (1857–1913). *Baltic Linguistics*, 1, 146–166. https://doi.org/10.32798/bl.438.

Pierrehumbert, J. (1980). *The Phonetics and Phonology of English Intonation*. PhD thesis, Massachusetts Institute of Technology.

de Pijper, J.R. (1983). *Modeling British English Intonation*. Foris, Dordrecht.

Pyž, G., Šimonytė, V., Slivinskas, V. (2011). Joining of vowel and semivowel models in Lithuanian speech formant-based synthesizer. In: *Proceedings of the 6th International Conference on Electrical and Control Technologies*, pp. 114–119.

Pyž, G., Šimonytė, V., Slivinskas, V. (2014). Developing models of Lithuanian speech vowels and semivowels. *Informatica*, 25(1), 55–72.

Radzevičius, A., Raudys, A., Kasparaitis, P. (2021). Speech synthesis using stressed sample labels for languages with higher degree of phonemic orthography. In: *International Conference on Information and Software Technologies*. Springer, Cham, pp. 378–387.

Raitio, T., Rasipuram, R., Castellani, D. (2020). Controllable neural text-to-speech synthesis using intuitive prosodic features. In: *Interspeech*. https://doi.org/10.48550/arXiv.2009.06775.

Raitio, T., Li, J., Seshadri, S. (2022). Hierarchical prosody modeling and control in non-autoregressive parallel neural TTS. In: *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 7587–7591. https://doi.org/10.1109/ICASSP43922.2022.9746253.

Rao, K.S., Yegnanarayana, B. (2009). Intonation modeling for Indian languages. *Computer Speech and Language*, 23(2), 240–256. https://doi.org/10.1016/j.csl.2008.06.005.

Raškinis, G., Kazlauskienė, A. (2013). From speech corpus to intonation corpus: clustering phrase pitch contours of Lithuanian. In: Oepen, S., Hagen, K., Johannesse, J.B. (Eds.), *NODALIDA 2013: Proceedings of the 19th Nordic Conference of Computational Linguistics*, May 22–24, 2013, pp. 353–363, Oslo University, Linköping University Electronic Press.

Reddy, V.R., Rao, K.S. (2011). Intonation modeling using FFNN for syllable based Bengali text to speech synthesis. In: *2nd International Conference on Computer and Communication Technology (ICCCT-2011)*, pp. 334–339. https://doi.org/10.1109/ICCCT.2011.6075155.

Ronanki, S., Henter, G.E., Wu, Z., King, S. (2016). A template-based approach for speech synthesis intonation generation using LSTMs. In: *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, pp. 2463–2467. https://doi.org/10.21437/Interspeech.2016-96.

Rossi, P.S., Palmieri, F., Cutugno, F. (2002). A method for automatic extraction of Fujisaki-model parameters. In: *Speech Prosody 2002, International Conference*, pp. 615–618.

de Saussure, F. (1879). *Mémoire sur le système primitif des voyelles dans les langues indo-européennes (Leipzig: printed Teubner)*. Teubner, Leipzig. Repr. in Saussure, 1922, 1–268.

de Saussure, F. (1894). À propos de l'accentuation lituanienne (intonations et accent proprement dit). *Mémoires de la Société de linguistique de Paris*, 8, 425–446.

de Saussure, F. (1896). Accentuation lituanienne. *Indogermanische Forschungen, Anzeiger*, 6, 157–166.

Savičiūtė, E., Ambridge, B., Pine, J.M. (2018). The roles of word-form frequency and phonological neighbourhood density in the acquisition of Lithuanian noun morphology. *Journal of Child Language*, 45(3), 641–672. https://doi.org/10.1017/S030500091700037X.

Siniova, O., Kundrotas, G. (2014). *Lietuvių kalbos tartis: garsai, žodis, intonacija [Pronunciation in Lithuanian Language: Phonemes, Words, Intonation]*. Lithuanian University of Educational Sciences.

Snow, D., Balog, H.L. (2002). Do children produce the melody before the words? A review of developmental intonation research. *Lingua*, 112(12), 1025–1058. https://doi.org/10.1016/S0024-3841(02)00060-8.

Stehwien, S., Schweitzer, A., Vu, N.T. (2020). Acoustic and temporal representations in convolutional neural network models of prosodic events. *Speech Communication*, 125, 128–141.

Sun, G., Zhang, Y., Weiss, R.J., Cao, Y., Zen, H., Wu, Y. (2020). Fully-hierarchical fine-grained prosody modeling for interpretable speech synthesis. In: *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6264–6268. https://doi.org/10.1109/ICASSP40776.2020.9053520.

Suni, A., Aalto, D., Raitio, T., Alku, P., Vainio, M. (2013). Wavelets for intonation modeling in HMM speech synthesis. In: *8th ISCA Workshop on Speech Synthesis*, Vol. 1, pp. 285–290.

Syrdal, A.K., Möhler, G., Dusterhoff, K., Conkie, A., Black, A.W. (1998). Three methods of intonation modeling. In: *Proceedings of the 3rd Esca/Cocosda Workshop on Speech Synthesis*, pp. 305–310.

Talandienė, M. (1970). Alternatyvinių klausimų loginiai ir komunikaciniai santykiai [Logical and Communication Relations in Alternative Questions]. *Kalbos garsai ir intonacija*. Vilnius.

Taylor, P. (1994). The rise/fall/connection model of intonation. *Speech Communication*, 15(1–2), 169–186.

Taylor, P. (1998). The tilt intonation model. In: *Proceedings of the 5th International Conf. on Spoken Language Processing (ICSLP 98)*. https://doi.org/10.21437/ICSLP.1998-153.

Text-Talk (2006). http://www.text-talk.com/lt/apie-kalbos-sinteze.html.

Tokuda, K., Zen, H., Black, A.W. (2002). An HMM-based speech synthesis system applied to English. In: *Proceedings of 2002 IEEE Workshop on Speech Synthesis*. IEEE, pp. 227–230.

Trubetzkoy, N.S. (1969). *Principles of Phonology*. University of California Press.

Tseng, C.Y., Pin, S.H., Lee, Y., Wang, H.M., Chen, Y.C. (2005). Fluent speech prosody: framework and modeling. *Speech Communication*, 46(3–4), 284–309. https://doi.org/10.1016/j.specom.2005.03.015.

Ulvydas, K. (Ed.) (1965). *Lietuvių kalbos gramatika / Lithuanian Grammar 1*. Mintis, Vilnius.

Vaičiūnas, A., Raškinis, G., Kazlauskienė, A. (2016). Corpus-based hidden Markov modelling of the fundamental frequency of Lithuanian. *Informatica*, 27(3), 673–688. https://doi.org/10.15388/Informatica.2016.105.

Venditti, J.J. (2005). The J ToBI model of Japanese intonation. In: Jun, S.A. (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford Unviersity Press, pp. 172–200.

Wang, D., Zheng, T.F. (2015). Transfer learning for speech and language processing. In: *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*. IEEE, pp. 1225–1237. https://doi.org/10.1109/APSIPA.2015.7415532.

Xu, Y. (2004). Transmitting tone and intonation simultaneously-the parallel encoding and target approximation (PENTA) model. In: *Proceedings of the First International Symposium on Tonal Aspects of Languages (TAL 2004)*, pp. 215–220.

Xu, Y. (2015). Speech prosody: theories, models, and analysis. In: Meireles, A.R. (Ed.), *Courses in Speech Prosody*, pp. 146–177.

Yngve, V.H. (1954). Language as an error correcting code. In: *Quarterly Progress Report of the Research Laboratory of Electronics*. MIT, Cambridge, MA, pp. 35–36.

Zen, H., Sak, H. (2015). Unidirectional long short-term memory recurrent neural network with recurrent output layer for low-latency speech synthesis. In: *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4470–4474.

Zerbian, S. (2010). Developments in the study of intonational typology. *Linguistics and Language Compass*, 4(9), 874–889. https://doi.org/10.1111/j.1749-818X.2010.00233.x.

G. A. Melnik-Leroy is a researcher at the Cognitive Computing Group at the Institute of Data Science and Digital Technologies (Vilnius University). She holds a doctoral degree in cognitive science from Ecole Normale Superieure, CNRS, EHESS (Paris), one of the world's leading institutions in the field. Her research is mainly focused on topics in cognitive psychology and psycholinguistics, including speech processing, the mental lexicon and language acquisition. The researcher also works on practical applications of findings from cognitive psychology to other fields, such as speech technologies, operational research and educational technologies. She has won several international research grants and successfully led research projects.

J. Bernatavičienė graduated from the Vilnius Pedagogical University, in 2004 and received MS degree in informatics. In 2008, she received the doctoral degree in computer science from Institute of Mathematics and Informatics jointly with Vilnius Gediminas Technical University. She is a senior researcher at Institute of Data Science and Digital Technologies of Vilnius University. Her research interests include data bases, data mining, neural networks, image analysis, visualization, decision support systems and Internet technologies. She supervises 2 PhD students and has written more than 60 articles, 15 of which are in CA WoS database.

G. Korvel received the BS degree in mathematics and the MS degree in informatics from Vilnius Pedagogical University (recently Vytautas Magnus University Education Academy), Lithuania, in 2007 and 2009, respectively, and the PhD degree from the Institute of Data Science and Digital Technologies, Vilnius University, in 2013. She is currently a senior researcher with the Institute of Data Science and Digital Technologies. Her research interests include speech signal processing, natural language processing, development of mathematical models, applications of soft computing, and computational intelligence. The main scientific results have been published in more than 30 papers and discussed at more than 40 national and international conferences. Some of her works received the Diploma for the Best Presentation. She is a three-time winner of the Lithuanian Academy of Sciences Young Scientist Award. She received acknowledgment from the Prime Minister of Lithuania for her obtained scientific results, in 2013 and 2019. G. Korvel took part in 4 research projects and 3 COST actions. She is a reviewer of many scientific journals, a member of the editorial board of *The Journal of Intelligent Information Systems*, and has been the session organizer at international conferences.

G. Navickas graduated from Vilnius Gediminas Technical University: in 2000 received BS degree in engineering informatics and in 2002 MS degree in statistics. He works at the Institute of Data Science and Digital Technologies of Vilnius University. His research interests include automatic Lithuanian speech recognition, Lithuanian speech synthesis, deep neural networks, speech signal processing, speech recognition methods and algorithms, speech interface applications in different fields, robotics.

G. Tamulevičius is a senior researcher at the Institute of Data Science and Digital Technologies (Vilnius University). His research interests include the analysis and modelling of speech signals, the digital processing of speech signals, and the applications of acoustic analysis of speech signals. Current activities include academic research, administration of the study process, and teaching students.

P. Treigys is a professor at the Faculty of Mathematics and Informatics at Vilnius university. He is a principal researcher and the head of the Signal and Image Analysis group at Vilnius University Institute of Data Science and Digital Technologies. His interests include image analysis, detection and object feature extraction in image processing, automated image objects segmentation, optimization methods, artificial neural networks, and software engineering. Povilas Treigys is a reviewer of the journals *Informatica*, *Sensors*, *Nonlinear Analysis*, and *The Baltic Journal of Modern Computing*, recently was invited to the editorial board of the conference DAMSS (Data Analysis Methods for Software Systems). He has supervised 1 postdoctoral and 6 PhD students and written more than 70 articles, 27 of which are in CA WoS database. He was the leader of the Lithuanian work group of 2 international projects.