ISCA Archive
http://www.isca-speech.org/archive

4th International Symposium on
Tonal Aspects of Languages (TAL-2014)
Nijmegen, The Netherlands
May 13-16 2014

F_0 peak timing, height, and shape as independent features

Gilbert Ambrazaitis, Johan Frid

Linguistics and Phonetics, Centre for Languages and Literature, Lund University, Sweden
Gilbert.Ambrazaitis@ling.lu.se, Johan.Frid@ling.lu.se

Abstract

A considerable amount of evidence from several intonation languages (e.g., German, English, Italian) supports the idea that F_0 peak timing, height, and shape variables form a *feature bundle*, which is used to encode two-fold intonational (e.g., sentence-level) pitch accent distinctions such as L+H* vs. L*+H. The three types of features in the bundle can be weighted differently but the outcome seems to be functionally equivalent. In this sense, they are 'substitute phonetic features'. This paper presents data from two distinct prosodic dialect types of Swedish, a pitch-accent language, suggesting that these F_0 variables can also be used independently of each other in order to encode two different contrasts (i.e., a three-fold contrast), each of which phonetically and functionally related to the L+H* vs. L*+H distinction in an intonation language. For Central Swedish, we observe two peak raising strategies which go along with differently shaped rises: 'extending' (= faster rise) and 'shifting' (= slower rise), which tend to be used to signal 'speaker-related' emphasis (e.g., 'surprise') or 'message-related' emphasis (e.g., 'correction'), respectively. For Southern Swedish, we observe an 'extended' peak and an 'extended and delayed' peak.

Index Terms: Intonation, prosody, focal accent, word accent, Swedish, emphasis, paralinguistic

1. Introduction

The timing of pitch events is widely assumed to be crucial for the encoding of both lexical and intonational contrasts in many languages. Well-established examples from intonation languages include the three-fold distinction between the *early* (H+L*), *medial* (H*), and the *late* peak (L*+H) of German, used to signal, roughly, 'matter-of-fact' vs. 'new' vs. 'contrastive information' [1]; as well as the two-fold contrast between L+H* and L*+H in Neapolitan Italian, encoding a narrow focus statement vs. a yes/no question [2].

What is modelled in terms of different timings or tonal associations, may, but need not be strictly realised in terms of the timing of an F_0 event. An illustration of this is provided in Fig. 1c, displaying typical realisations of the *early/ medial/ late* contrast in German. While the early (H+L*) peak is distinguished clearly from the medial (H*) peak in terms of timing, the phonetic distinction between *medial* and *late* is, in this example, a matter of a later, higher, and temporally extended peak in *late* as compared to *medial*.

In the present paper, we focus strictly on intonational (i.e. sentence-level) contrasts as encoded by F_0 peaks with a timing within or after the accented vowel, such as typically in (L+)H* and L*+H in German or Italian. For this specific prosodic condition, it has been suggested that F_0 *height* and *timing* can function as 'substitute phonetic features' [3], in the sense that a *delayed* F_0 peak can enhance the effect of, or even replace a *raised* F_0 peak, a claim very much in line with the observation in Fig. 1c.

The idea of substitute features is generally supported, and even extended, by some recent perception experiments on German H* vs. L*+H [1] and American English L+H* vs. L*+H [4], which both studied the perceptual relevance of *peak shape* features. As a main result, a faster rise [1] or a 'scooped' (as opposed to a 'domed') rise [4] both introduce a bias towards perceiving a later peak timing, i.e. L*+H.

We can hence assume a *feature bundle* of three types of substitute features: F_0 *timing*, *height*, and *shape* variables. Further support for including peak shape variables in this bundle comes from studies on the perception of plateau- (as opposed to peak-) shaped realisations of pitch accents [5, 6], as well as from production studies (e.g., [7]).

This brief review suggests that different types of F_0 manipulations may have equivalent *functional* effects: Given a certain (sentence-level) functional contrast – such as new vs. contrastive in German; statement vs. question in Italian – where one of the categories is signalled by means of a 'relatively early' F_0 peak (such as H*), the other category (L*+H) may be encoded by an F_0 gesture that is *delayed*, *raised*, or *differently shaped* in a critical way.

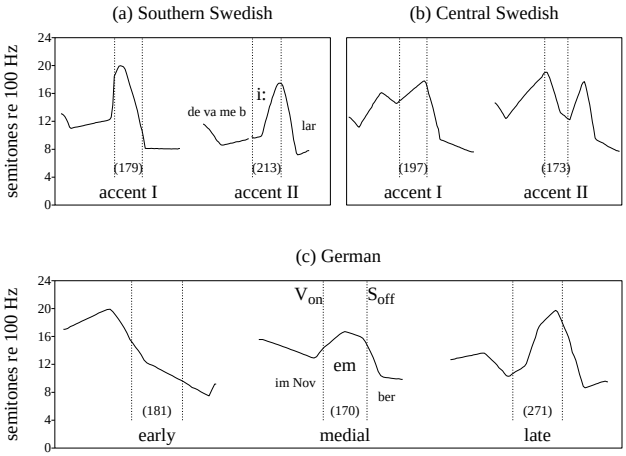

Figure 1: *Typical realisations of the Swedish word accents I and II in (a) Southern and (b) Central Swedish, compared to (c) the three-fold contrast of sentence-level pitch accents in German. Stylised F_0-curves based on authentic productions of a short phrase: Swedish accent I (a&b):* det var med bilen *'by car'; Swedish accent II (a&b):* det var med bilar *'with cars'; German (c):* im November *'in November'. Dotted lines delimit the rhyme of the stressed syllable (in parentheses: duration in ms); V_{on} = stressed vowel onset; S_{off} = stressed syllable offset.*

This conclusion does not, however, exclude the possibility that each of the F_0 variables under discussion can be used distinctively by itself. For instance, pitch timing [8, 9] as well as pitch height or shape variables [10, 11] can be distinctive in the encoding of lexical tonal contrasts.

In this paper, we present data from Central and Southern Swedish, which suggest that – in conditions equivalent to the *medial/late* (H* vs. L*+H) contrast of German – F_0 peak timing, height, and shape features not only function as substitute features, but also independently of each other in order to encode finer (sentence-level) functional contrasts.

1.1. Tonal prosody in Central and Southern Swedish

Swedish has a tonal contrast at the word level (Accent I vs. Accent II). As for sentence intonation, two basic types of Swedish dialects can be distinguished: those that mark focus by means of an additional tonal peak and those that do not [12]. The two dialects dealt with in this paper represent these two types.

In Southern Swedish, the F_0 contour of an utterance is mostly a result of the tonal patterns that encode the word accents. Each word accent is typically realised as a rising-falling F_0 movement, timed later for Accent II than for Accent I (see Fig. 1). Focus is signalled through an increased F_0 range in the focused word [13]. This is different in Central Swedish, where focus is signalled by an additional tonal movement: the *focal accent* (*sentence accent* in [8]), which is realised after the word accent gesture (see the two-peaked patterns in Fig. 1b).

The current paper studies Accent I materials only. The important aspect to be learned from this brief introduction is that the F_0 peak occurring within the stressed syllable in Accent I represents the *word accent* gesture in *Southern* Swedish, while it represents the *sentence* accent in *Central* Swedish.

2. Method

The data presented here has been collected in [14] and [15], where a full account of the methods and materials can be found. Both data sets (Southern and Central Swedish) were recorded based on a common set of materials: The test phrase *i november* 'in November' was elicited in a number of conditions, representing a variety of 'discourse contexts'. A condition was made up, first, of a written context, presented to subjects on a computer screen; in some conditions, the written context was followed by an audio prompt: a context question, which was prerecorded by a dialect-matched speaker and presented via headphones. Following this context (either text only or both text and audio prompt), the target sentence was presented, which was to be read aloud by the subject and audio-recorded. The target sentence consisted of the test phrase, with some minor additions depending on the condition, such as 'Wow! In November! Not bad!'. An example of a complete condition is presented in (1).

(1) Corrective response (text and audio prompt)
Written context: *Du är polis och träffar en gammal kollega. Ni småpratar lite om jobbet.*
You are a police officer and meet an old colleague. You talk about your job.
Audio context question: *Och du tar din semester alltså i oktober igen, då?*
You're going on holiday in October again, right?
Target sentence: *Nej! I november.*
No! In November.

Five of the test conditions are discussed here for Central Swedish: (a) ***new-information response***: henceforth, NEW, similar structure as (1), lacking the contrastive component ('In November.'); (b) ***corrective response***: COR, see (1); (c) ***exclamation***: EXC ('Ok! In November! Now I understand.'); (d) ***surprised feedback***: SUR ('Wow! In November! Not bad!'); finally, (e) ***question***: QUE ('And when is your exam? In November?'). Four out of these conditions – all but QUE – are also discussed for Southern Swedish.

A comment on the condition-dependent additional words such as 'No!' or 'Wow!' might be in order. A risk with using such words might be that they already convey the discourse function or expressive meaning to be elicited (e.g., 'correction', 'surprise') in a sufficient manner, making it less necessary to also express, say, correction or surprise intonationally on the target phrase. However, these additions were regarded useful for reinforcing the intended interpretation of the test condition, i.e. as a support for the subjects. As the results will show, the test phrase was indeed intoned differently in the different conditions, despite the additional words.

Five repetitions of each condition were recorded by each speaker. The data presented here are based on nine adult speakers for each dialect (five females in both groups).

F_0 data were normalised for both time and speaker, in order to support visual presentation and comparison of the data such as to make it possible to calculate mean F_0 contours across several repetitions of the same intonation patterns, either for a single speaker or across speakers. Time normalisation was achieved by taking ten temporally equidistant F_0 measurements for each segment; the utterances were segmented into six phonetic sections, according to the following broad phonetic transcription, representing a possible Central Swedish realisation: [i n] [ʊ] [ʊ] [ɛ] [m] [bəɪ].

Speaker normalisation was achieved by relating F_0 values in semitones to a subjects' base F_0 value F_b [16]. For estimation of F_b and further details on data processing, see [15].

All tokens were labeled according to a simple scheme, classifying it as one of a few, dialect-dependent, pattern types. For instance, all tokens of Central Swedish discussed in this paper were of the type 'non-early' [14], referring to a focal F_0 peak realised after the onset on the stressed vowel. All tokens averaged in a mean curve are of the same basic pattern type.

3. Results

3.1. Central Swedish

Results for Central Swedish, across all nine speakers as well as individually for five selected speakers, are plotted in Figure 2. In the average across all speakers (Fig. 2a), the following pattern seems to emerge: In condition NEW, speakers produce a rising-falling F_0 pattern across the stressed and the post-stress syllable (*-vember*). This pattern reflects the focal accent of Central Swedish (see 1.1). When defining the accent pattern observed in NEW as a baseline, the patterns observed in the remaining conditions could be conceived of as variants of the focal accent pattern found in NEW, which all share the common feature of a 'raised' F_0 peak, as compared to NEW.

A closer look at these average curves reveals a more nuanced analysis, as there seem to be different types of 'raising' involved: In SUR, for instance, the accentual rise starts off at approximately the same F_0 level as in NEW, while the level reached at the end of rise is much higher than in NEW. In other words, the *range* of the rise is clearly extended. This seems to

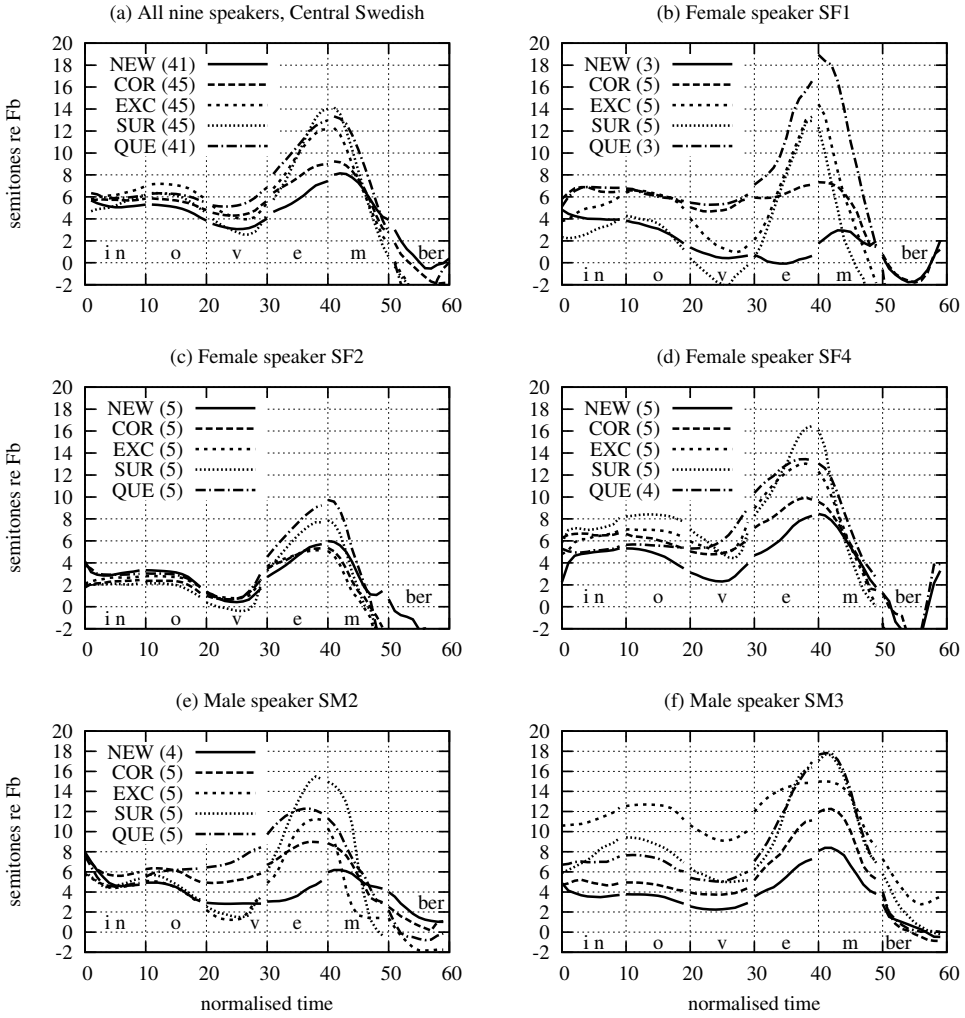

Figure 2: *Mean F_0 contours (N in parentheses) for Central Swedish speakers: (a) average across all nine speakers; N=45 where no tokens are missing; (b-f) individual plots for five selected speakers (N=5 for SM3, all conditions). The normalised time scale indicates the number of measurements; vertical lines are segment boundaries. Segments are labelled orthographically.*

be different in COR, where the entire rising movement, including the onset level, seems to be shifted upwards, while the range seems to be kept constant.

These two basic strategies of F_0 peak raising – i.e., *shifting* vs. *extending* – seem to be present in the intonational repertoire of some speakers, but absent in others. Most clearly, speaker SM2 (Fig. 2e) differentiates between all five conditions by employing *two gradual steps of each of the two strategies*, counted from the baseline (NEW). A distinction between the two strategies, or at least traces of it, is observed for six speakers in total (four of which are included in Fig. 2: SF1, SF4, SM2, and SM3), with some restrictions or alternations: For instance, SF1 combines the *shifting* and *extending* strategies in condition QUE (see Fig.2b). Despite this variability, there is a tendency for a differential usage of the two strategies: For the six relevant speakers, *extending* is clearly preferred in condition SUR, while *shifting* is preferred in QUE and COR; for EXT, *extending* and *shifting* was observed in three speakers each.

Finally, *shifting* seems to be realised differently by different speakers: some seem to perform a *register* shift already from the onset of the utterance (see, e.g., COR vs. NEW in speaker SF1, or EXC vs. NEW in SM3; Fig. 2); for others, the shifting seems to increase gradually from the utterance onset (e.g., SM2). In both cases, however, the result is a rather slow or shallow rise in comparison to the rise resulting from *extending* the peak.

3.2. Southern Swedish

For Southern Swedish, the most crucial result in the present context comes out sufficiently well in an average plot across all nine speakers (Fig. 3). In contrast to the case of Central Swedish, we only take into account four of the conditions. As a first observation, the F_0 pattern produced in condition COR does not seem to differ crucially from the pattern in NEW. That said, there still might be prosodic differences, e.g. in terms of durations, which we neglect in this paper.

Turning to EXC and SUR, we observe a sharper rise with an extended range in both conditions, when compared to NEW, which is mainly achieved by a lowering of the rise onset. However, EXC and SUR also differ clearly from each other, in that the peak is *delayed* in EXC, while, instead, it seems to be slightly

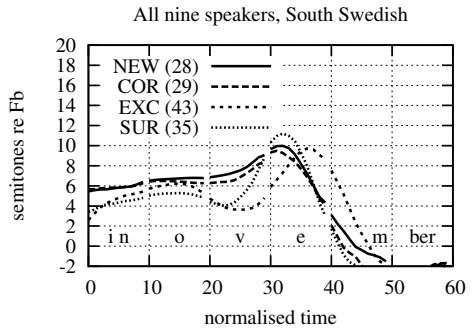

Figure 3: *Mean F_0 contours (N in parentheses) for Southern Swedish speakers; N=45 where no tokens are missing. For further explanations, see Fig. 2.*

raised in SUR (in comparison to both NEW and EXC). This results in a clear acoustic (and perceptual, which is, however, not confirmed by listening tests yet) distinction between the three conditions NEW, EXC, and SUR.

4. Discussion

The test conditions included in this study – *correction, exclamation, surprised feedback, and question* – may all be said to involve the addition of emphasis of some sort, in relation to the baseline condition *new-information response*. It is well known that emphasis is often, cross-linguistically, realised by means of prosodic prominence, and increased prosodic prominence in turn is often achieved by 'increasing' the accentual F_0 gesture. It was therefore quite expected to observe a 'raising' or 'extension' of the F_0 peak in our conditions. However, unexpectedly, for each of the two dialects, we observed *two different strategies* of encoding additional emphasis by means of different combinations of timing, height, and/or shape variables, used to distinguish between different types of emphasis, i.e. different conditions in the present study.

Speakers of Southern Swedish varied peak *shape and height*, on the one hand, to distinguish SUR from NEW, while they manipulate *shape and timing*, on the other hand, to distinguish EXC from NEW, suggesting that timing alone might suffice to distinguish between the expression of *surprise* and *exclamation* (for further discussion, see [15]).

Turning to Central Swedish, the results suggest two different basic strategies of F_0 peak 'raising': *shifting* vs. *extending* the focal F_0 peak, where, roughly, *extending* implies a manipulation of peak *height* and *shape* (= faster rise), while *shifting* mainly implies a manipulation of *height* (while preserving a rather slow rise). They seem to be applied by a majority of speakers, with certain degrees of freedom. For those speakers who made a distinction, SUR was most often realised using the *extending* strategy, while *shifting* was preferred for both QUE and COR. Thus, there seems to be at least some correlation between the choice of strategy and the 'linguistic quality' of the test condition: Both conditions EXC and SUR seek to elicit a certain flavour of expressiveness, and are thus 'speaker-related', whereas QUE and COR are 'message-related'.

To sum up, two quite different phenomena were observed for the two dialects of Swedish included in this study, which, however, are similar in a crucial respect: Speakers from both dialects seemed to group the conditions by means of different phonetic strategies, but these groups were different for Southern (NEW/COR vs. EXC vs. SUR) and Central Swedish (NEW vs. COR/QUE vs. SUR, where results for EXC were unclear).

What the results for Central and Southern Swedish have in common is that speakers of both dialects used F_0 peak timing, height, or shape variables in order to define, although not in the same way, two *different* phonetic strategies in order to distinguish between *different* emphasis-eliciting conditions.

Recall that this result is different from what is typically found for intonation languages: Even for German, it is reasonable to assume that timing, height, or shape features alone might be able to signal additional emphasis (typically modelled as an L*+H accent) compared to an H* baseline [1, 17]. However, we would not expect *different* strategies such as those observed in the present study to be used to distinguish say, a correction from a surprised feedback. Rather, our conditions COR, EXC, and SUR would be encoded by the same strategy (possibly distinguished by means of gradual variants of that strategy) – basically, a somewhat delayed, raised, or sharpened F_0 peak as compared to NEW, i.e. using peak timing, height, and shape variables as a bundle of substitute, rather than independent features – which is indeed what was found in [14] using equivalent German materials.

The original research question in [14] was to test whether speakers of Central Swedish – a language with a lexical pitch distinction – would make similar intonational distinctions as German. Given the simplicity of the lexical pitch-accent system in Swedish, it is not surprising that Swedish may exhibit a similar intonational repertoire as German. It is noteworthy, however, that both Central and Southern Swedish seem to make even finer intonational distinctions than German, at least as far as the conditions of this study are concerned.

In this connection, we should point that a detailed discussion of possible implications for the intonational *phonology* of the Swedish dialects investigated is outside the scope of this paper. Some of the differences observed between conditions should certainly be regarded as within-category, or 'paralinguistic' [18] variations of the word accent (Southern Swedish) or the focal accent gesture (Central Swedish), respectively. However, clarifying whether this applies to all of the distinctions observed in the present data, and whether this also might apply to the German *medial/late* distinction (which is typically notated H* vs. L*+H suggesting a phonological distinction), is less relevant in the present context.

To conclude, in addition to the frequent and well-attested use of F_0 peak timing, height, and shape variables as 'substitute phonetic features' [3] in the signalling of enhanced intonational emphasis, our data suggest that these F_0 variables can be used *independently* of each other, in order to encode *different nuances* of emphasis, such as to distinguish a 'correction' from a 'surprised' feedback, or the latter from an 'exclamation'.

5. Future directions

Future work will attempt to corroborate the present findings with data from further speakers, including Swedish Accent II materials, as well as investigate further acoustic measures (such as durations) and their perceptual relevance.

6. Acknowledgements

This study was supported by the Swedish Research Council (grant 2009–1566).

7. References

[1] Niebuhr, O., “The signalling of German rising-falling intonation categories – The interplay of synchronization, shape, and height”, Phonetica, 64: 174–193, 2007.

[2] D’Imperio, M. and House, D., “Perception of questions and statements in Neapolitan Italian”, Proc. Eurospeech’97, Rhodes, Greece, 251–254, 1997.

[3] Gussenhoven, C., “The Phonology of Tone and Intonation”, Cambridge University Press, 2004.

[4] Barnes, J., Veilleux, N., Brugos, A. and Shattuck-Hufnagel, S., “The effect of global F_0 contour shape on the perception of tonal timing contrasts in American English intonation”, Proc. 5th Speech Prosody, Chicago, USA, 2010.

[5] Knight, R. and Nolan, F., “The effect of pitch span of intonational plateaux”, JIPA, 36(1): 1–28, 2006.

[6] D’Imperio, M., Gili Fivela, B. and Niebuhr, O., “Alignment perception of high intonational plateaux in Italian and German”, Proc. 5th Speech Prosody, Chicago, USA, 2010.

[7] Niebuhr, O., D’Imperio, M., Gili Fivela, B. and Cangemi, F., “Are there ‘shapers’ and ‘aligners’? Individual differences in signalling pitch accent category”, Proc. 17th ICPhS, Hong Kong, China, 120–123, 2011.

[8] Bruce, G., “Swedish Word Accents in Sentence Perspective”, Travaux de l’institut de linguistique de Lund, 12, 1977.

[9] Remijsen, B., “Tonal alignment is contrastive in falling contours in Dinka”, Language, 89(2): 297–327, 2013.

[10] Kuang, J., “The tonal space of contrastive five level tones”, Phonetica, 70: 1–23, 2013.

[11] Morén, B. and Zsiga, E., “The lexical and post-lexical phonology of Thai tones”, Natural Language & Linguistic Theory, 24: 113–178, 2006.

[12] Bruce, G., “Components of a prosodic typology of Swedish intonation”, in T. Riad and C. Gussenhoven [Eds], Tones and Tunes – Volume 1: Typological Studies in Word and Sentence Prosody, 113–146, Mouton de Gruyter, 2007.

[13] Bruce, G. and Gårding, E., “A prosodic typology for Swedish dialects”, in E. Gårding, G. Bruce and R. Bannert [Eds], Nordic Prosody – Papers from a Symposium, 219–228, Lund University, 1978.

[14] Ambrazaitis, G., “Nuclear Intonation in Swedish – Evidence from Experimental-Phonetic Studies and a Comparison with German”, Travaux de l’institut de linguistique de Lund, 49, 2009.

[15] Ambrazaitis, G., Frid, J. and Bruce, G., “Revisiting South and Central Swedish intonation from a comparative and functional perspective”, in O. Niebuhr [Ed], Understanding Prosody – The role of context, function, and communication, 138–158, De-Gruyter, 2012.

[16] Traunmüller, H., “Conventional, biological, and environmental factors in speech communication: A modulation theory”, Phonetica, 51: 170–183, 1994.

[17] Kohler, K., “Categorical pitch perception. Proceedings of the XIth ICPhS, Tallin, Estonia, 331–333, 1987.

[18] Ladd, D.R., “Intonational Phonology (2nd ed.)”, Cambridge University Press, 2008.