Original Paper

Phonetica 2016;73:338–361
DOI: 10.1159/000446001

Received: June 28, 2015
Accepted after revision: April 3, 2016
Published online: February 23, 2017

The Phonetics and Phonology of the Polish Calling Melodies

Amalia Arvaniti[a] Marzena Żygis[b] Marek Jaskuła[c]

[a]University of Kent, Canterbury, UK; [b]Centre for General Linguistics and Humboldt University, Berlin, Germany; [c]West Pomeranian University of Technology, Szczecin, Poland

Abstract

Two calling melodies of Polish were investigated, the *routine call*, used to call someone for an everyday reason, and the *urgent call*, which conveys disapproval of the addressee's actions. A Discourse Completion Task was used to elicit the two melodies from Polish speakers using twelve names from one to four syllables long; there were three names per syllable count, and speakers produced three tokens of each name with each melody. The results, based on eleven speakers, show that the routine calling melody consists of a low F0 stretch followed by a rise-fall-rise; the urgent calling melody, on the other hand, is a simple rise-fall. Systematic differences were found in the scaling and alignment of tonal targets: the routine call showed late alignment of the accentual pitch peak, and in most instances lower scaling of targets. The accented vowel was also affected, being overall louder in the urgent call. Based on the data and comparisons with other Polish melodies, we analyze the routine call as LH* !H-H% and the urgent call as H* L-L%. We discuss the results and our analysis in light of recent findings on calling melodies in other languages, and explore their repercussions for intonational phonology and the modeling of intonation.

1. Introduction

The intonation of calling melodies has often been the topic of research within the autosegmental-metrical framework of intonational phonology (henceforth, AM; see e.g., Ladd, 2008a). In this literature, particular attention has been paid to the *vocative chant*, a type of melody described as *stylized* (Ladd, 2008a). The vocative chant has been investigated in many languages, including English (Liberman, 1975; Pierrehumbert, 1980; Gussenhoven, 2004; Brugos et al., 2006), Dutch (Gussenhoven, 2005), Greek (Arvaniti and Baltazani, 2005), German (Grice et al., 2005), Hungarian (Varga, 2008), Arabic (Chahal and Hellmuth, 2014), Catalan (Borràs-Comes et al., 2015), and Portuguese (Frota et al., 2015). In most (though not all) of these languages,

KARGER

E-Mail karger@karger.com
www.karger.com/pho

Marzena Żygis
Centre for General Linguistics (ZAS)
Schützenstrasse 18
DE–10–117 Berlin (Germany)
E-Mail zygis@zas.gwz-berlin.de

the vocative chant is roughly realized as a rise followed by a fall to sustained level pitch in the middle of the speaker's range. The fact that pitch remains level at the end gives this melody its 'chanting' quality prompting in turn its characterization as stylized.

The fact that the vocative chant ends in sustained level pitch has attracted significant attention within AM. This is so because this type of pitch configuration is not easy to represent by using only L (low) and H (high) tones (Arvaniti, 2016). This is evident in the analyses of the English vocative chant. Liberman (1975) relies on a M (mid) tone to represent its sustained level phrase-final pitch, while Pierrehumbert (1980) analyzes it as a sequence of a H-phrase accent followed by a L% boundary tone: the H- accent is downstepped because it is the second H tone in a HLH tonal sequence (it follows a H*+L pitch accent); the following L% is upstepped because it is preceded by a H-phrase accent. This analysis is retained in Mainstream American English ToBI (MAEToBI) but with explicit marking of the downstep giving !H-L% (Brugos et al., 2006). In contrast, Gussenhoven (2004) analyzes the downstep as part of the pitch accent which he represents as H*H, with the lack of any further tonal specification resulting in the plateau (for detailed discussions, see Gussenhoven, 2004: 313ff., Ladd, 2008a: 110ff., and Arvaniti, 2016).

In addition to the vocative chant, other calling melodies have lately been investigated in some detail. Borràs-Comes et al. (2015) examined three Catalan melodies used with vocatives, L* H%, a rising melody, L+H* HL%, a rise-fall, and L+H* !H%, the vocative chant. The authors examined both the production of these melodies and their acceptability in different pragmatic contexts along four parameters, insistence (first or insistent call), and physical distance, (social) power and social distance between speaker and addressee. They show that although all three melodies can be used in most contexts, some were produced more frequently in certain contexts and were judged by an independent group of participants as more appropriate for those contexts as well. Interestingly, in Catalan, the melody that emerged as the default was L+H* HL%; the rising L* H% melody was considered most appropriate for calling someone at a distance, while the vocative chant was considered insistent and appropriate mainly for addressees of lower power but socially close, such as children. This use of the vocative chant is rather surprising compared to the description of the melody in other languages such as English (Ladd, 2008a) and Hungarian (Varga, 2008), where the vocative chant is not associated with insistence and is considered appropriate for calling an addressee at some distance from the speaker.

The above observations on calling melodies raise several issues that pertain, on the one hand, to the form and representation of such melodies cross-linguistically, and to the relationship between their form and use, on the other. Here these issues are addressed in the study of two Polish calling melodies. The first is used to call someone when the speaker wants to express disappointment or disapproval of the addressee or their actions; we refer to this melody as the *urgent calling melody* or *urgent call* for short. The second melody is used to call a person for an everyday reason; we refer to this melody as the *routine calling melody* or *routine call* for short. Examples of the two melodies are shown in figures 1 and 2 (see also figure 14 for an example of the routine call with a longer utterance). The urgent call is similar in use to the melody referred to as 'impatient' or 'insistent' in analyses of Portuguese (Frota, 2014) and Catalan (Prieto, 2014). The routine call is similar in function to the vocative chant of English (Ladd, 2008), Hungarian (Varga, 2008), Dutch (Gussenhoven, 2004) and

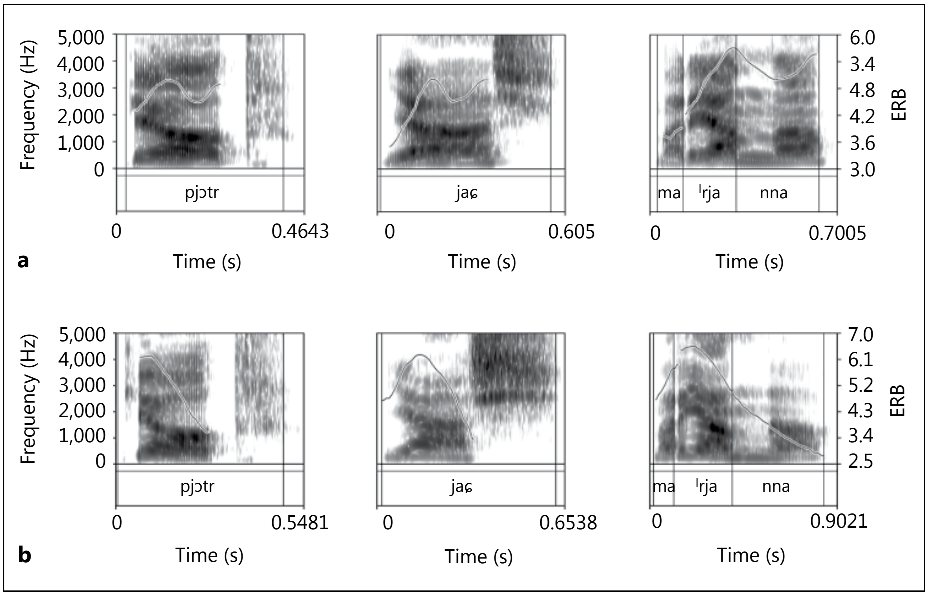

Fig. 1. Illustration of the routine call (**a**) and the urgent call (**b**) in monosyllabic words [pjɔtr] and [jaɕ], and a three-syllable word [maˈrjanna] exhibiting different degrees of tonal crowding.

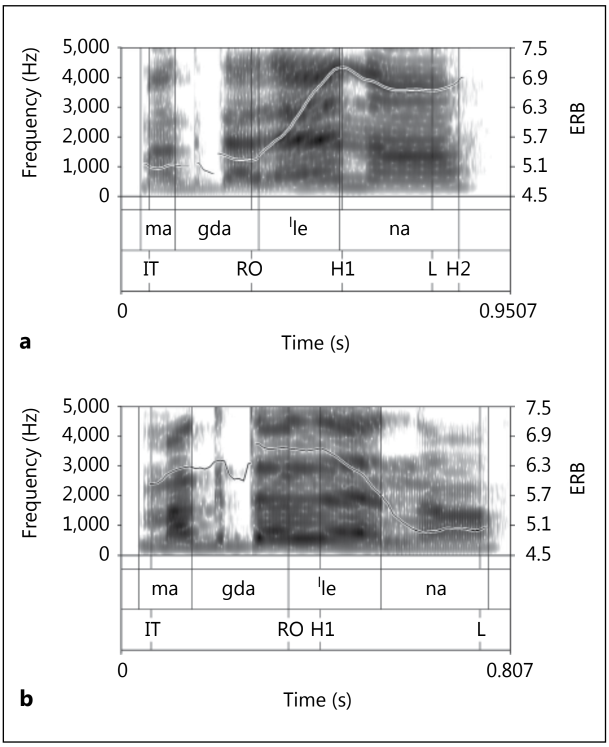

Fig. 2. Illustration of measurements for the routine melody (**a**) and the urgent melody (**b**).

Portuguese (Frota, 2014), among others, and impressionistically sounds very similar to them. However, inspection of pilot recordings of the Polish routine call suggested that unlike those melodies it ends in a small final rise rather than a plateau.[1]

Our aim in investigating these two melodies was threefold. First, we wished to systematically examine their characteristics in order to account for their phonetics and determine their tonal composition, thereby situating them within the intonation system of Polish. Despite several studies on Polish intonation (Demenko, 1999, 2015; Dukiewicz, 1978; Dłuska, 1947; Francuzik et al., 2004; Jassem, 1987; Jassem and Demenko, 1986; Karpiński, 2006; Wagner, 2008), there are no detailed accounts of its calling melodies. To the best of our knowledge, calling melodies are only briefly discussed by Steffen-Batogowa (1996: 164ff.) who points to differences depending on pragmatic context. One context investigated by Steffen-Batogowa is comparable to our 'dinner' context (see section 2.3): a person is calling someone who is in another room to answer a phone call. Steffen-Batogowa describes this melody as falling-rising, so it could be the routine call examined here, but since no detailed descriptions are provided it is difficult to interpret this pattern with any certainty.

To determine the phonetics and phonological representation of the melodies, we used tonal crowding as a diagnostic (cf. Arvaniti and Ladd, 2009). Tonal crowding refers to the situation in which the composition of a melody is such that there are more tonal elements than tone-bearing units (typically vowels) on which the tonal elements may be realized (see Ladd, 2008a for a review). As Arvaniti and Ladd (2009) have shown, tonal crowding can lead to adjustments in the realization of tonal elements, such as the earlier alignment of peaks and the curtailment or even elimination of particular parts of a melody; e.g., Arvaniti and Ladd (2009) show that the F0 of Greek wh-questions beginning with unstressed syllables (e.g. [apoˈpu] 'from where?') starts low and rises to a peak, whereas in those beginning with a stressed syllable (e.g. [ˈpu] 'where?') F0 starts high in the speaker's range and shows a substantially curtailed rise. Because of its effects on F0, tonal crowding can be used as a diagnostic of the status of specific parts of a melody: it can be used to determine whether a particular pitch excursion, such as a rise or a dip, is required or can be dispensed with during production, the assumption being that retained elements are more likely to be phonologically specified. Tonal crowding allowed us to investigate in detail the following elements of the Polish melodies observed in pilot recordings. The status of the initial rise: we wished to examine whether the initial rise is an element of both melodies and if so whether it is associated to the stressed syllable in a similar manner in both (or whether indeed it bears no connection with the stressed syllable). The status of the final rise in the routine call: as noted, this rise is somewhat unusual and we wished to ascertain whether it is simply a return to a default mid-level, as some analyses might indicate (cf. Gussenhoven, 2004), or a deliberate pitch movement. If the former hypothesis is correct, one would expect the final rise to be dramatically reduced under tonal crowding leading to plateaux.

This treatment of the data and the use of tonal crowding as a diagnostic presuppose acceptance of certain assumptions, such as that the segmental string of an

[1] In his work, Gussenhoven also mentions a vocative fall-rise used in Dutch (Gussenhoven, 2005) and English (Gussenhoven, 2004); based on Gussenhoven's illustrations this melody is probably similar in form to the Polish routine call investigated here; however, we have not been able to find a definite description of the meaning of this melody in either English or Dutch and thus we do not discuss it further.

utterance is independent of the tonal string and that the latter (like the former) is composed of discrete elements (for a discussion, see Arvaniti, 2011). These assumptions stem directly from AM, the theoretical framework adopted here. AM was adopted as it is the only fully developed *phonological* framework for intonation and one of our aims was, as mentioned, to provide both a phonetic account and a *phonological* analysis of the melodies involved (on the difference between the two approaches to intonation, see among others Arvaniti, 2011). Our results, however, are of relevance to other models as well; this applies particularly to the comparison of the melodies as realized under tonal crowding and in the absence of it, as not all models can handle tonal crowding effects equally well (Arvaniti and Ladd, 2009). We will return to this point in section 4.3.

Our results are also of relevance to AM. As mentioned, melodies like the Polish urgent call have not been examined in much detail before, while the vocative chant poses challenges for AM. We, therefore, wished to use the present results to contribute to this discussion on the typology of intonation and the nature of AM representations. Researchers have argued that the vocative chant is essentially the same phenomenon across a variety of languages and should be autosegmentally represented in a similar manner in all of them using phonetically transparent representations (e.g., Ladd, 2008a: 110 ff.). Others have argued that cross-linguistically, similarity of representations and phonetic transparency are largely incompatible (e.g. Arvaniti, 2016), and that language-specific autosegmental representations are necessary if one is to capture contrasts specific to each linguistic system and adhere to analytic coherence (Gussenhoven, 2005; Arvaniti, 2016). We will return to this point in section 4.2.

Finally, our aim was to compare the routine and urgent calls both with respect to their tonal composition and with respect to additional acoustic properties that may differentiate them. Such differences could be manifested in the duration, amplitude and spectral properties of the tone-bearing units, particularly the accented vowels, but could extend beyond this domain. There is now an increasing body of research showing that such effects of intonation on segmentals can be systematic. For instance, Arvaniti et al. (2006) show that the L+H* pitch accent of Greek is accompanied by greater duration of the accented syllable when used to encode contrastive focus. Arvaniti and Garding (2007) also show systematic differences in duration between syllables accented with L*+H vs. L+H* in English with the latter being consistently longer. Niebuhr reports that both [t] aspiration (Niebuhr, 2008) and voiceless fricatives (Niebuhr, 2012) have a higher centre of gravity (i.e. sound more high pitched) in rising than falling melodies in German. Similar results are reported by Żygis et al. (2014) about Polish voiceless sibilants, which were pronounced with higher spectral peaks, higher intensity, centre of gravity and standard deviations in polar questions (rising intonation) as compared to statements (falling intonation). Despite these results, the link between intonation and segmentals has not been sufficiently investigated; thus one aim of the present study was to develop this line of research further.

2. Methods

2.1. Speakers

Data were elicited from sixteen native speakers of Standard Polish, eight male and eight female. The speakers' age ranged from 18 to 23 years. They were all monolingual speakers of Standard Polish from Szczecin (North Poland) or its vicinity. They were all volunteers and naïve as to the purposes of the experiment; none reported a history of speech or hearing disorders. For reasons explained in more detail in section 2.3, the results reported here are based on data from 11 of these speakers.

Table 1. The names used in the study

Number of syllables	Names		
1	*Jan* [jan]	*Jaś* [jaɕ]	*Piotr* [pjɔtr]
2	*Daniel* [ˈdaɲɛl]	*Bartek* [ˈbartɛk]	*Patryk* [ˈpatrɨk]
3	*Natalia* [naˈtalja]	*Marianna* [maˈrjanna]	*Małgośka* [mawˈgɔɕka]
4	*Magdalena* [magdaˈlɛna]	*Alexandra* [alɛˈksandra]	*Małgorzata* [mawgɔˈʐatka]

2.2. Materials

The materials were Polish names one to four syllables long; see table 1. All names longer than one syllable were stressed on the penult. Three names per syllable count were selected; some included mostly sonorants, e.g. *Natalia*, *Daniel*, while others included either voiceless fricatives, e.g. *Jaś*, or voiceless stops, e.g. *Piotr*, *Bartek*.

The purpose of incorporating such differences was to observe the above-mentioned effect of tonal crowding on the two melodies. First, we wanted to examine the effect of tonal crowding on the scaling and alignment of the tonal targets in the two melodies, particularly when they have to be realized on monosyllables; adjustments of this sort are problematic for intonation models which rely on syllable-by-syllable tonal specifications, such as PENTA (e.g., Xu et al., 2015). Second, with respect to the routine call, we wanted to see if the rise at its end would be dramatically reduced in contexts leading to tonal crowding and whether reduction would lead to the elimination of the rise or to plateaux (as reported in other languages); elimination would suggest that the rise is not a required part of the melody, while the replacement of the rise by a plateau would indicate that the rise is not phonologically specified but is simply a return to some default.

2.3. Procedures

Recordings were conducted in a sound-proof room at the Electrical Engineering Department of the West Pomeranian University of Technology in Szczecin, Poland. The recordings were conducted with a TLM103 Neumann microphone placed at a distance of ca. 30 cm from the speaker's mouth and connected to a ProTools system with a Digi 003 interface (sample rate 44,100 Hz, 16 bit).

Three repetitions were elicited using a Discourse Completion Task (DCT). The speakers were asked to imagine calling a child who is playing in the garden (i) to dinner, (ii) to get a present, or (iii) because they discovered she had broken a vase. Generally, contexts (i) and (ii) elicited the routine call, while context (iii) elicited the urgent call. We chose to represent the addressee as a child, as we judged this to make for natural DCTs. Nevertheless, the intuition of the two authors who are native Polish speakers is that neither melody is confined to child-directed speech (though this is clearly a point that requires further investigation). Speakers found the DCT easy, but a small number of them were not consistent in their use of the two intended melodies which they replaced with others. The main reason appears to be that these speakers misinterpreted the instructions, so they did not actually call a person but, rather, talked to them as if they were in the same room using a variety of melodies in all contexts. Since only three repetitions of each melody were elicited per test word and speakers were not asked to repeat material if the sought melody was not used, the variability meant that for these speaker's datasets were incomplete, with some test words missing altogether; for this reason their data were discarded. In our view, this points to the difficulties with consistently eliciting specific melodies in a laboratory setting, but does not affect the legitimacy of the melodies examined here and their place in the Polish intonational system: the majority of the speakers consistently used the melodies we sought to elicit. The reported results are based on eleven speakers, six female and five male. For statistical calculations, 633 observations were obtained, 339 for the routine call and 294 for the urgent call.

2.4. Measurements and Statistical Analysis

F0 was measured in ERB, i.e., the perceptual equivalent bandwidth calculated from acoustic frequency in Hz, using the formula provided by PRAAT (11.17 ln((*x*+312)/(*x*+14680))+43). ERB was used to reduce differences between male and female speakers (but see below). Measurements were taken at specific points in the contour. Measurements were obtained semi-automatically at specific points in the contour: areas of interest in terms of F0 were selected and maxima and minima were located and annotated using PRAAT's functions 'Get maximum pitch' and 'Get minimum pitch', respectively. In 17 instances in which F0 was halved due to creaky voice, the value obtained by the script was doubled. Stressed vowels, syllables and word boundaries were also marked.

A variety of measurements were obtained from the annotated data by means of PRAAT scripts. The tonal targets measured were as follows (fig. 2):

- *Initial Tone* (henceforth, *IT*): the F0 at the onset of the contour;
- *Rise Onset* (henceforth, *RO*): in the routine melody, RO was defined as the F0 at the onset of the rise to the first peak in the contour; in the urgent melody, RO could not be defined in the same way because there was no low-level stretch (fig. 1b), so RO was operationally defined as the F0 at the onset of the stressed syllable; in both melodies, RO coincided with IT in 1- and 2-syllable words;
- *H1*: in the routine melody, H1 was the F0 of the first peak in the contour; in the urgent melody H1 was the only peak;
- *L*: in the routine melody, L was the F0 minimum in the dip in the contour (between the two rises); in the urgent melody it was the lowest point reached at the end of the contour;
- *H2*: the F0 maximum at the contour's end; this measurement applied to the routine melody only.

In addition, the distance of RO, H1 and L from the following landmarks was measured:

- The distance of RO from the onset of the stressed vowel (SVO), i.e. tSVO-tRO (for the routine melody only);
- The distance of H1 from the onset of the stressed vowel expressed as a proportion of the vowel's duration, i.e. (tH1-tSVO)/durSV;
- The distance of L from H1 and H2, i.e. tH1-tL and tH2-tL, respectively (for the routine melody only);
- The distance of L from the onset of the last vowel (which coincided with the stressed vowel in monosyllables), as a proportion of the vowel's duration, i.e. (tL-tLVO)/durLV (for the routine melody only).

Finally, properties of stressed vowels, stressed syllables and words were investigated by measuring the following parameters: duration, RMS amplitude, amplitude integral of the stressed vowel (a summation of duration and RMS amplitude; see e.g., Beckman, 1986; and references therein, Gordon, 2002, 2005), spectral balance of the stressed vowel, and centre of gravity of sibilants in word-final position.

The duration and RMS of the stressed vowel were normalized by expressing them as ratios of the word duration and RMS amplitude, respectively. The duration and RMS amplitude of stressed syllables were not normalized as the corpus contained several monosyllabic words.

To measure spectral balance, which cannot be calculated for the same band ranges for all vowels, we selected the vowel /a/ which was present in stressed syllables in 9 out of 12 words and in words of all syllable counts (see table 1). The signal of the vowel /a/ was band-pass filtered at the following band ranges: B1 = 0–500 Hz, B2 = 500–1000 Hz, B3 = 1000–2000 Hz, B4 = 2000–4000 Hz (see Sluijter and van Heuven, 1996a). The filtered signals were transformed into intensity objects in PRAAT from which the intensity means were extracted (b1, b2, b3, b4), and differences in intensity between frequency bands (b2-b1, b3-b2 and b4-b3) were calculated.

All statistical calculations were made in R (R Development Core Team 2008). Linear mixed-effect models were used to investigate the influence of SYLLABLE COUNT and GENDER on the variables listed above. When analyzing F0-based differences between the two melodies, we also included TONE TYPE as a fixed effect with IN, RO, H1, L and H2 as sublevels. To account for variability among speakers and items and to minimize Type I error (Barr et al., 2013), random intercepts for participants and items as well as their slopes for SYLLABLE COUNT and TONE TYPE were included. In addition, very high correlations between random-effects terms were excluded from the model. No high correlations between fixed effects were observed. The maximized models were tested against less complex models by means of ANOVAs. The best fit model was taken as the final model.

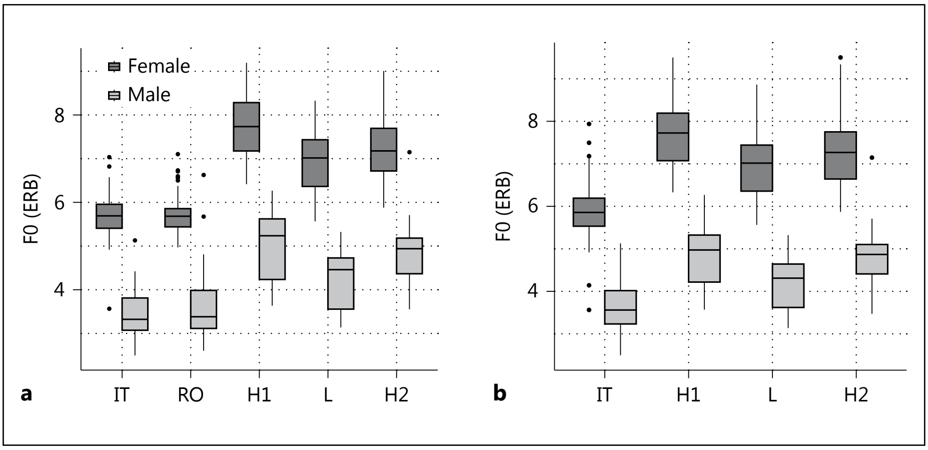

Fig. 3. Scaling of Initial Tone (IT), Rise Onset (RO), H1, L and H2 in the routine call in (**a**) 3- and 4-syllable words only (so as to include RO) and (**b**) all test words.

Residuals of the model were normally or nearly normally distributed. All *p*-values reported in this paper are based on Satterwhite approximation available in the package 'lmerTest' (Kuznetsova et al., 2015) which provides different kinds of tests for linear mixed-effect models as implemented in the 'lme4' package (Bates et al., 2015).

3. Results

3.1. F0 Scaling and Alignment: Routine Calling Melody

Preliminary inspection of the data indicated that in polysyllabic words the routine call is realized on the last foot (i.e. the last two syllables), with low F0 before the stressed syllable, rising F0 on the stressed penult and a reduced rise on the ultima (fig. 1a, 2a). Comparisons of F0 scaling between IT and RO in 3- and 4-syllable words (where IT and RO could be measured as distinct points in the contour; see section 2.4) confirmed this observation. As shown in figure 3a, IT and RO were similar in scaling: there was no statistically significant difference between them [t = 0.628]. Syllable count did not affect the scaling of IT and RO. The only significant difference was one of gender: as expected, male speakers produced lower IT and RO in comparison to female speakers [mean IT: males 3.38 ERB (s.d. 0.45) vs. females 5.70 ERB (s.d. 0.37); mean RO: males 3.49 ERB (s.d. 0.52) vs. females 5.69 ERB (s.d. 0.36), $t = -12.183$, $p < 0.001$]. The differences are illustrated in figure 3a. [In figure 3 and all remaining box plots, boxes correspond to the 25th to 75th percentile range, black lines represent medians and whiskers correspond to ±1.5 interquartile range; data above or below this range are outliers and are represented as points in the graph.]

The data also confirmed that the Polish routine call involves a final rise. As illustrated in figure 3, H2 was scaled significantly lower than H1 but still significantly higher than L. This applied not only when data from 3-and 4-syllable words were considered (fig. 3a), but also for all the data together (fig. 3b) [for all data: mean H2 =

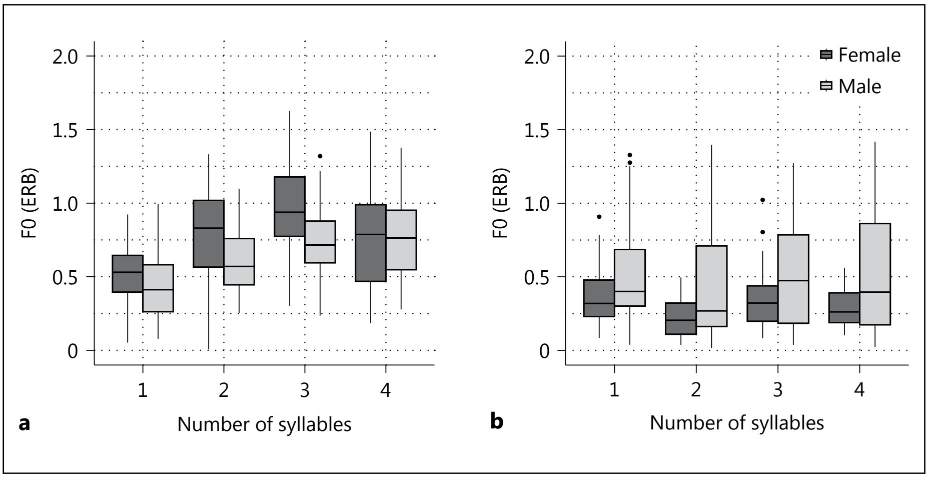

Fig. 4. (**a**) Scaling difference between H1 and L and (**b**) H2 and L by syllable count in the routine call.

6.05 ERB (s.d. 1.41) vs. H1 = 6.32 ERB (s.d. 1.57), $t = -3.186$, $p < 0.01$; mean H2 vs. L 5.62 ERB (s.d. 1.5), $t = 5.133$, $p < 0.001$]. These differences between H1 and L, on the one hand, and L and H2, on the other, were largely unaffected by tonal crowding as they held at all syllable counts (the influence of the number of syllables was nonsignificant), indicating that the melody is not realized as a plateau even under tonal crowding. Rather, undershoot under tonal crowding affected H1 instead, i.e. the first peak in this melody: H1 showed lower scaling in monosyllables compared to longer words [mean H1: 1 syllable 6.01 ERB (s.d. 1.59) vs. 2 syllables 6.45 ERB (s.d. 1.5), $t = -4.180$, $p < 0.01$; for 1 syllable vs. 3 syllables 6.53 ERB (s.d. 1.63), $t = -6.38$, $p < 0.001$; for 1 syllable vs. 4 syllables 6.32 ERB, $t = -5.268$, $p < 0.001$]. Finally, the influence of gender was significant: as expected, males produced all tonal targets with lower values than females [mean H1: males 4.87 ERB (s.d. 0.67) vs. females 7.65 ERB (s.d. 0.64); mean H2: males 4.14 ERB (s.d. 0.57) vs. females 6.89 ERB (s.d. 0.65), mean L1: males 4.68 ERB (s.d. 0.6) vs. females 7.20 ERB (s.d. 069), t = –11.01, p < 0.001].

The same pattern emerges if one compares the differences in scaling between H1 and L, on the one hand, and H2 and L, on the other; see figure 4. The difference between H1 and L was significantly smaller in monosyllables compared to longer words [means: 1 syllable 0.46 ERB (s.d. 0.24) vs. 2 syllables 0.69 ERB (s.d. 0.28), $t = -2.31$, $p < 0.05$; for 1 syllable vs. 3 syllables 0.85 ERB (s.d. 0.28), $t = -4.68$, $p < 0.001$; for 1 syllable vs. 4 syllables 0.76 ERB (s.d. 0.31), $t = -3.80$, $p < 0.001$]. In addition, 2-syllable words showed a smaller difference between H1 and L in comparison to 3-syllable words [$t = -2.79$, $p < 0.05$]; see figure 4a. Again, the effect of speaker's gender was significant [mean: females 0.74 ERB (s.d. 0.33) vs. males 0.64 ERB (s.d. 0.28), $t = 3.002$, $p < 0.05$]. On the other hand, the scaling difference between H2 and L was largely unaffected by syllable count: the only difference found was between monosyllables and 2-syllable words with the former showing greater H2-L than the latter [mean: for 1 syllable 0.48 ERB (s.d. 0.36) vs. 2 syllables 0.32 (s.d. 0.3), $t = 3.13$, $p < 0.05$]; see figure 4b. Male and female speakers did not differ significantly with respect to the scaling difference H2-L [$t = 0.550$].

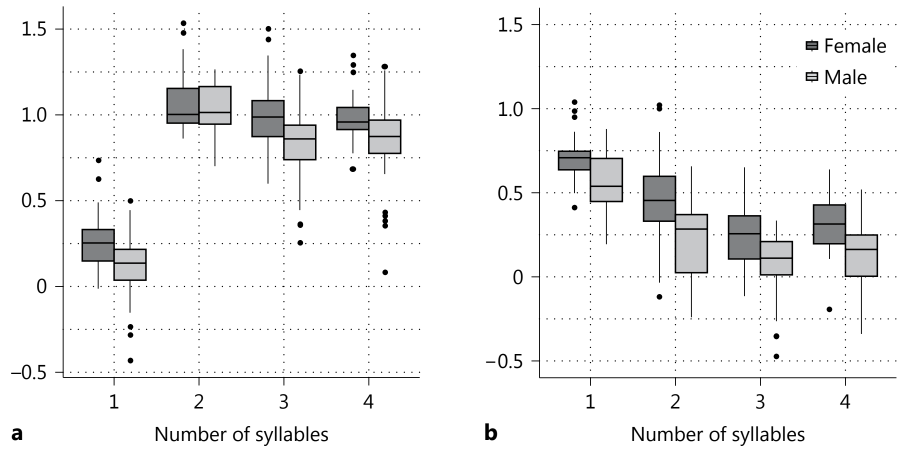

Fig. 5. (**a**) Proportional alignment of H1 with respect to the stressed vowel duration; (**b**) proportional alignment of L with respect to the duration of the last vowel (this vowel coincides with the stressed vowel in monosyllables) in the routine call.

In addition to scaling, the alignment of RO, H1 and L was examined. The alignment of RO with respect to the onset of the stressed vowel was somewhat variable but unaffected by syllable count [$t = 0.227$]. RO occurred on average 87 ms (s.d. = 37) before the stressed vowel onset. The effect of gender was nonsignificant [$t = 0.087$]. The alignment of H1, on the other hand, was affected by syllable count. H1 was aligned roughly with the end of the stressed vowel, except in monosyllables where it occurred close to the vowel's onset. All comparisons between monosyllables and all other words were significant [mean based on ratio: 1 syllable 0.18 (s.d. 0.19) vs. 2 syllables 1.05 (s.d. 0.15), $t = –13.8$, $p < 0.001$; 1 syllable vs. 3 syllables 0.92 (s.d. 0.24), $t = –11.68$, $p < 0.001$; 1 syllable vs. 4 syllables 0.91 (s.d. 0.28), $t = 11.73$, $p < 0.001$]; no significant differences between male and female speakers were found [$t = –1.892$]; see figure 5a. Finally, L generally occurred early within the last vowel, and its alignment was also affected by syllable count. Specifically, L aligned later in monosyllabic words, where it occurred roughly in the middle of the only vowel; in 2-, 3- and 4-syllable words, on the other hand, it occurred closer to the last vowel's onset. Significant differences were found with respect to monosyllables and other words [mean based on ratio: 1 syllable 0.65 (s.d. 0.2) vs. 2 syllables 0.33 (s.d. 0.33), $t = 2.65$, $p < 0.05$; 1 syllable vs. 3 syllables 0.07 (s.d. 0.41), $t = 4.13$, $p < 0.01$; 1 syllable vs. 4 syllables 0.23 (s.d. 0.25), $t = 3.29$, $p < 0.05$]; on the other hand, no statistically significant differences were found among 2- vs. 3- and 4- syllable words; The alignment of L was significantly affected by gender [mean based on ratio: females 0.41 (s.d. 0.32) vs. males 0.22 (s.d. 0.4), $t = 4.708$, $p < 0.001$]; see figure 5b.

Results also show that L was realized closer in time to H2 than H1 [mean: tH2-tL: 0.121 s (s.d. 0.52) vs. tL-tH1 0.168 s (s.d. 0.73), $t = –57.8$, $p < 0.001$]. The temporal

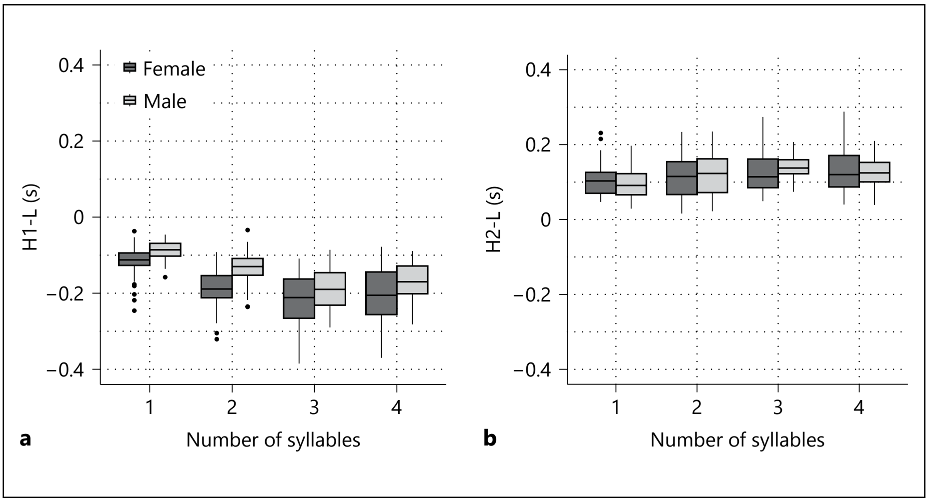

Fig. 6. Distance of L from (**a**) H1 (H1-L) and (**b**) H2 (H2-L) in the routine call.

distance tH1-tL was generally larger for monosyllables as compared to 3- and 4-syllable words [mean: 1 syllable –0.104 s (s.d. 0.03) vs. 3 syllables –0.206 s (s.d. 0.072), $t = 3.13$, $p < 0.05$; 1 syllable vs. 4 syllables –0.191 s (s.d. 0.068), $t = 2.71$, $p < 0.05$]; see figure 6a. The distance was greater for male than female speakers [mean tH1-tL: males –0.146 s (s.d. 0.058) vs. females –0.187 s (s.d. 0.08), $t = 3.473$, $p < 0.001$]. Furthermore, tH2-tL was significantly shorter in monosyllables as opposed to 3- and 4-syllable words [mean: 1 syllable 0.101 s (s.d. 0.044) vs. 3 syllables 0.132 s (s.d. 0.04), $t = –2.646$, $p < 0.05$; 1 syllable vs. 4 syllables 0.134 s (s.d. 0.057), $t = –2.821$, $p < 0.05$]; see figure 6b. No significant differences in tH2-tL were found between male and female speakers [$t = –0.629$]. Finally, for both measurements, 1- and 2-syllable words behaved in a similar manner [for tH1-tL $t = –1.858$; for tH2-tL $t = 1.358$].

3.2. F0 Scaling and Alignment: Urgent Calling Melody

In the urgent call, F0 starts low and rises directly to a peak, without creating a low plateau first; after the peak it drops to the bottom of the speaker's range; cf. figures 1b and 2b. This description is supported by the results. First, the lack of a low plateau is indicated by the comparison of IT and RO scaling in 3- and 4-syllable words (the two types of words in which both IT and RO were distinct measures): in the urgent call, RO was significantly higher than IT, as shown in figure 7a [mean: RO 6.5 ERB (s.d. 1.32) vs. IT 5.65 ERB (s.d. 1.32), $t = 5.488$, $p < 0.001$]. Values were lower for male than female speakers [mean IT: males 4.23 ERB (0.77) vs. females 6.35 ERB (s.d. 0.59); mean RO: males 5.17 ERB (s.d. 0.94) vs. females 7.14 ERB (s.d. 0.95), $t = –7.63$, $p < 0.001$]. When the entire corpus was considered, H1 was shown to be scaled significantly higher than both IT and L [mean: H1 = 6.98 ERB (s.d. 1.33) vs. IT = 5.8 ERB (1.32), $t = 8.499$, $p < 0.001$; H1 vs. L = 4.25 ERB (s.d. 0.99), $t = 10.606$, $p < 0.001$]; see figure 7b. The differences held for all syllable counts (the effect of syllable count was not significant). Again, all tones were produced significantly lower by male than female speakers [mean H1: males 5.62 ERB (s.d. 0.86) vs. females 7.75 ERB

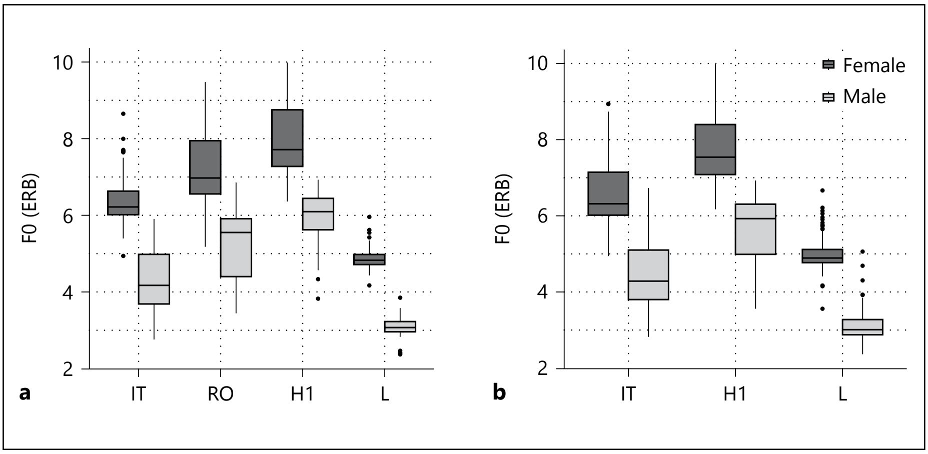

Fig. 7. Scaling of IT, RO, H1 and L in the urgent call in (**a**) 3- and 4-syllable words and (**b**) all words.

(s.d. 0.86); mean L: males 3.05 ERB (s.d. 0.41) vs. females 4.94 ERB (s.d. 0.38), t = –18.33, p < 0.001].

Proportional alignment data showed that H1 in the urgent calling melody was by and large aligned with the onset of the stressed vowel, independently of syllable count. The only exception was disyllabic words, where H1 occurred approximately two thirds into that vowel; the difference between 2-syllable words and all other words was statistically significant [means based on ratio: 2 syllables 0.63 (s.d. 0.32) vs. 1 syllable –0.02 (s.d. 0.25), t = 6.267, p < 0.001, 2 syllables vs. 3 syllables 0.16 (0.38), t = 4.482, p < 0.01, 2 syllables vs. 4 syllables –0.13 (s.d. 0.51), t = 5.345, p < 0.001]. No significant differences were found between male and female speakers [t = –0.032]; see figure 8.

3.3. F0 Scaling and Alignment: Comparing the Routine and Urgent Calling Melodies

A comparison of the initial tone (IT) in the routine versus urgent calling melody showed that IT was significantly higher in the latter [mean: urgent call 5.79 ERB (s.d. 1.3) vs. routine call 4.79 ERB (s.d. 1.24), t = 3.641, p < 0.01]. RO (measured on 3- and 4-syllable words only) was also significantly higher in the urgent calling melody [mean: 6.58 ERB (s.d. 1.27)] than in the routine calling melody [mean: 4.69 ERB (s.d. 1.19), t = 4.56, p < 0.001]. H1 was produced with similar scaling in both contours [mean: routine call 6.35 ERB (s.d. 1.57) vs. urgent call 6.93 ERB (s.d. 1.38), t = 1.651]. On the other hand, L was significantly lower in the urgent melody compared to the routine melody [mean: urgent call 4.23 ERB (s.d. 1.01) vs. routine call 5.68 ERB (s.d. 1.41), t = –26.61, p < 0.001]; compare figures 3 and 7.

In addition, differences between the two melodies were found in the alignment of H1 with respect to the stressed vowel: as noted earlier, H1 was aligned generally close to this vowel offset in the routine melody, while in the urgent melody it aligned close to the onset of the vowel. The later peak alignment in the routine melody relative to the urgent melody applied to words of all syllable counts [mean based on ratio: 1 syllable, routine melody 0.18 (s.d. 0.19) vs. urgent melody –0.02 (s.d. 0.35), t = 4.283, p <

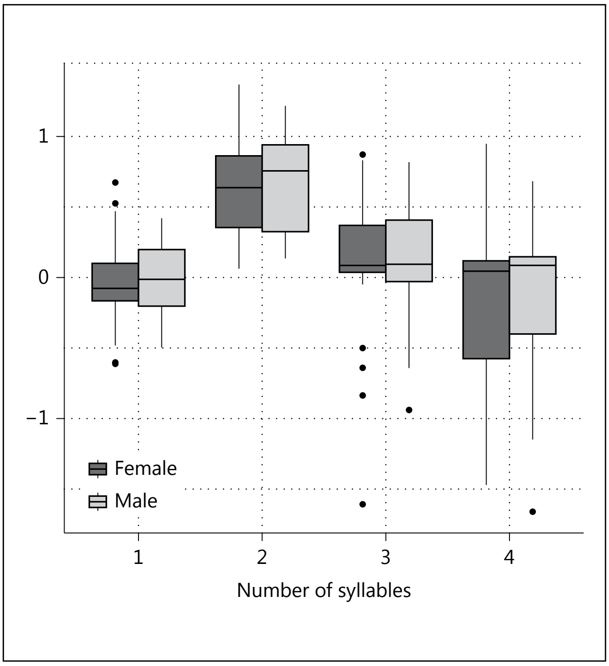

Fig. 8. Proportional alignment of H1 with respect to the stressed vowel duration in the urgent calling melody.

0.01; 2 syllables, routine melody 1.08 (s.d. 0.19) vs. urgent melody 0.83 (s.d. 0.3), t = 12.65, $p < 0.001$; 3 syllables, routine melody 0.92 (s.d. 0.21) vs. urgent melody 0.16 (s.d. 0.197), $t = 22.33$, $p < 0.001$; 4 syllables, routine melody 0.96 (s.d. 0.24) vs. urgent melody –0.04 (s.d. 0.31), $t = 23.67$, $p < 0.001$]; compare figures 5a and 8.

3.4. Stressed Vowel Differences in Spectral Balance

Our results showed that the spectral balance of higher frequency bands was significantly different for stressed vowels in the two melodies; In particular, spectral balance in the b3-b2 band was significantly higher in the urgent call as compared to the routine call [mean: urgent call 1.25 dB (s.d. 5.06) vs. routine call –1.41 dB (s.d. 5.18), $t = 3.07$, $p < 0.05$]. The spectral balance of lower frequencies (b2-b1) and higher frequencies (b4-b3) did not show any significant differences between melodies [mean b2-b1: routine call 8.75 dB (s.d. 4.43) vs. urgent call 8.98 dB (s.d. 4.6), $t = -0.66$; mean b4-b3: routine call –12.2 dB (s.d. 4.32) vs. urgent call –13.14 dB (s.d. 4.3), $t = -1.38$].

3.5. Duration Differences

The relative duration of the stressed vowel was significantly longer in the routine call than in the urgent call [mean: 0.24 (s.d. 0.12) vs. 0.23 (s.d. 0.10), $t = 2.62$, $p <$ 0.05]. Stressed vowel duration was also affected by the number of syllables in the test word, in that longer words had shorter stressed vowels than shorter words (fig. 9). In particular, monosyllables had a significantly longer stressed vowel duration than 2-, 3-, and 4-syllable words [1 syllable vs. 2 syllables, $t = -8.32$, $p < 0.001$; 1 syllable vs. 3 syllables, $t = -9.93$, $p < 0.001$; 1 syllable vs. 4 syllables, $t = -11.35$, $p < 0.001$]. No significant effects regarding the duration of the stressed vowels were found when

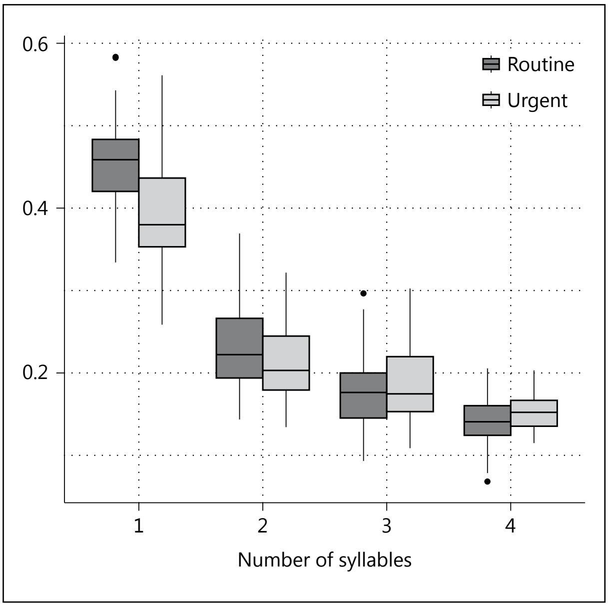

Fig. 9. Relative duration of the stressed vowel by syllable count and melody; data pooled over participant gender.

comparing 2- vs. 3-syllable words [$t = -1.61$], and 3- vs. 4-syllable words [$t = 1.42$], but stressed vowels were longer in 2- than 4-syllable words [$t = 3.04$, $p < 0.05$]. Male speakers produced shorter stressed vowels than female speakers [mean: male 0.22 (s.d. 0.10) vs. female 0.24 (s.d. 0.11), $t = -2.736$, $p < 0.05$].

On the other hand, the duration of the stressed syllables[2] was not affected by melody [mean: routine call 0. 309 s (s.d. 0.143) vs. urgent call 0. 303 s (s.d. 0.144), $t = 1.12$]. However, as with the stressed vowel duration, syllable duration was longer in monosyllables as compared to multisyllabic words (1 syllable vs. 2 syllables, $t = 7.38$, $p < 0.001$, 1 syllable vs. 3 syllables, $t = 6.23$, $p < 0.001$, 1 syllable vs. 4 syllables, $t = 6.54$, $p < 0.001$), while male speakers produced shorter stressed syllables than female speakers [mean: male 0.275 s (s.d. 0.121) vs. female 0.327 s (s.d. 0.153), $t = -3.07$, $p < 0.05$]. Similarly, duration of the entire words was not significantly different in the two contexts [mean: routine call 0.655 s (s.d. 0.175) vs. urgent call 0.659 s (s.d. 0.199), $t = 0.05$]. For details, see online supplementary Appendix B (see www.karger.com/doi/10.1159/000446001).

3.6. RMS, and Amplitude Integral Differences

The relative RMS amplitude was significantly higher in stressed vowels produced in urgent calls than in routine calls [mean based on ratio: urgent call: 1.641 (s.d. 0.269) vs. routine call 1.449 (s.d. 0.234), $t = 4.70$, $p < 0.001$]. The same conclusion applies to stressed syllables [mean: urgent call 0.094 Pa (s.d. 0.044) vs. routine call 0.088 Pa (s.d. 0.041), $t = 4.06$, $p < 0.001$] and words [mean: urgent call 0.086 Pa (s.d. 0.038) vs. routine call 0.071 Pa (s.d. 0.030), $t = 3.38$, $p < 0.01$]. The integral of the stressed

[2] Please recall that we report only raw values for syllables and words.

vowel was also significantly higher in the urgent call than in the routine call [mean: urgent call 0.020 (s.d. 0.012) vs. routine call 0.016 (s.d. 0.011), $t = 2.23$, $p < 0.05$]; the same applied at the word level [mean: urgent call 0.056 (s.d. 0.027) vs. routine call 0.048 (s.d. 0.025), $t = 4.74$, $p < 0.001$], though not at the syllable level: there was no statistically significant differences between the stressed syllable in the routine and urgent calls.

4. Discussion

4.1. Phonetic Realization of the Calling Melodies

Overall, the results showed systematic differences between the two melodies. These differences extended beyond the tonal events examined here and encompassed aspects of segmental realization as well.

With respect to the routine call, the results confirmed preliminary observations based on pitch tracks from pilot recordings. Specifically, the results from IT and RO confirmed that in the routine call, F0 remains low until the onset of the stressed syllable, if there is a sufficient number of syllables to do so, i.e. if the utterance includes three or more syllables; if not, the contour still starts low but rises quickly. The ensuing pitch peak is aligned with the stressed vowel and is followed by a dip and a final rise of reduced excursion (male speakers 0.5 ERB; female speakers: 0.33 ERB on average), high in the speaker's frequency range. Tonal crowding had a significant effect manifested primarily on monosyllables which showed earlier alignment and lower scaling of H1, and later alignment of L. However, even under extreme tonal crowding, as in monosyllables like *Piotr* where only the glide and the vowel can carry the tune (since the final /r/ is devoiced in this context), the difference between L and H2 was present. This indicates that the final rise is deliberate and not a return to some default; if that were the case, one would expect more significant undershoot of the final tone (H2) when tonal crowding is extreme, resulting consistently in plateaux or at least in substantial curtailment of this rise, but clearly this was not the case as illustrated in figure 1a.

The urgent calling melody, on the other hand, shows an abrupt rise to a peak reached early on the stressed vowel (much earlier than the equivalent peak of the routine call). In the urgent call, the rise is not preceded by a low F0 stretch even when there are unstressed syllables on which low F0 could be realized, as when the names are three or four syllables long. This is illustrated in figures 1b and 2b, which also indicate that this melody can start with high F0 under extreme tonal crowding; this happens with monosyllabic names, particularly when they start with a voiceless consonant (cf. [pjɔtr] and [jaɕ] in figure 1b). These patterns of realization indicate that the lack of the low F0 stretch in the urgent call is deliberate, as it is absent in all words independently of tonal crowding. Further, the tonal crowding data indicate that the rise in the urgent call is dispensable, while this is not the case in the routine chant (cf. the melody of [pjɔtr] in panels a and b of figure 1). Finally, the urgent call ends in a fall that reaches the bottom of the speaker's range, a pattern that differs substantially from the realization of the F0 dip in the routine call which remains high in scaling even when tonal crowding is absent.

Generally, tonal crowding had only a small effect on the urgent call melody (beyond the elimination of the initial rise): the alignment of the tonal targets was consistent independently of syllable count, with the notable exception of disyllables which

DOI: 10.1159/000446001

show later peak alignment than all other words. This result is somewhat unusual and cannot be explained by the composition of the relevant stressed syllables or other characteristics of the disyllables used here. It is possible, however, that it reflects priorities in the realization of the tonal targets involved. In monosyllables, reaching the bottom of the speaker's range at the end takes priority and is achieved by curtailing or eliminating the rise to the initial peak; in polysyllabic words there is sufficient segmental material for both the initial rise and subsequent fall. Disyllables then may be a special case: as they are longer than monosyllables, the rise to the peak need not be eliminated, but its realization is only possible with significant peak delay; since this peak delay does not compromise the final fall, it is the preferred solution.

The comparison of the urgent and routine calling melodies showed systematic differences between the two. First, as noted, the routine call starts with a low F0 stretch that is absent in the urgent call. Second, the accentual peaks of the two melodies are aligned differently: by and large, the routine call shows late peak alignment, while the urgent call shows early peak alignment, practically at the onset of the accented vowel. Finally, the two melodies end very differently: the routine call remains high in the speaker's range, while the urgent call ends at the bottom; as a result, the dip in F0 between the two peaks of the routine call is consistently higher than the onset of both melodies and the L reached at the end of the urgent call; in other words, it is higher than all targets that can be reasonably seen as reflexes of L tones.

In addition to the differences in the scaling and alignment of tonal targets, the two melodies showed small but systematic differences in the segments involved, particularly the stressed vowel. Specifically, the stressed vowel in the urgent call is somewhat shorter but also louder that the vowel in the routine call. The difference in duration is due primarily to monosyllables (see fig. 9) in which the vowel is extended to accommodate the more complex melody of the routine call; this interpretation is supported by the minimal differences in duration between the two melodies in longer words. In turn, this suggests that the loudness differences are due to RMS amplitude, not to a combination of amplitude and duration. This conclusion is further supported by the finding that the spectral balance is higher in the urgent than in the routine call in the b3-b2 frequency band which contributes significantly to perceived intensity (Sluijter et al., 1996b). In short, when the duration, RMS, amplitude integral and spectral balance are considered together, they strongly suggest that the stressed vowel is louder in the urgent than the routine call of Polish.

These results indicate that intonation differences are not conveyed exclusively by F0; they also affect the segments involved (cf. Niebuhr, 2008, 2012; Żygis et al., 2014; Borràs-Comes et al., 2015). It remains to be seen whether the small but systematic differences observed here have an effect on perception. Clearly, in this particular instance, differences in amplitude, duration or vowel spectral balance would be extremely unlikely to change a speaker's understanding of the melody used. However, they could affect the naturalness of the rendition, while the possibility that segmental effects may shift listeners' perceptions in melodies that are more similar to each other cannot be ruled out.

4.2. Autosegmental Analysis of the Melodies

Based on the patterns of scaling and alignment of the tones in the two melodies, the proposed autosegmental analysis of the Polish routine call is LH* !H-H% and that of the urgent call H* L-L%. We discuss these analytical decisions in more detail below in light of additional observations regarding the intonation system of Polish.

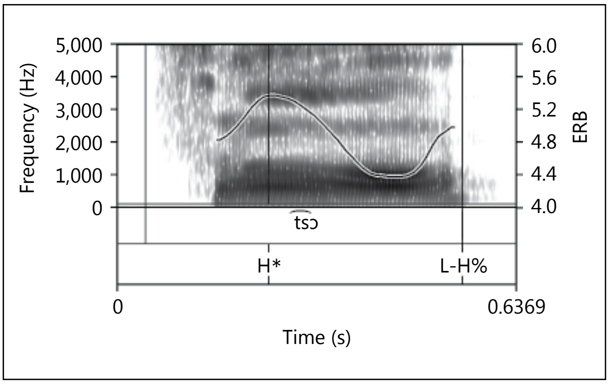

Fig. 10. A melody used with Polish wh-questions to indicate incredulity as realized on the monosyllabic wh-word *co* 'what?'

In the routine call, since the initial rise is aligned with the stressed syllable, it is analyzed as a LH* pitch accent. The use of a bitonal representation is necessary here as the melody remains consistently low at the beginning. This contrasts with the urgent calling melody which clearly lacks the initial low F0; cf. figures 1a, b, 2a, b. The fact that the melody of the urgent call lacks this initial low F0 stretch is reflected in the analysis of its accent as H*. At first glance, these representations appear to leave the low F0 beginning of the two melodies unaccounted for. Inspection of other melodies in our extended Polish corpus, however, indicates that low F0 at the beginning of utterances may be a default for Polish. If so, it does not need to be annotated, but can be considered part of the phonetic realization of Polish intonation: in the absence of any specification, start low (tonal crowding permitting). It is, however, possible that a %L boundary tone aligned with the left boundary of the utterance would have to be posited for the onset of the melodies discussed here, if other melodies are found that start high (i.e. if a principled distinction has to be made between %L and %H because no other configuration can account for the low or high onset of different melodies, or if a change from %L to %H is shown to change the meaning of a melody).

The dip and rise after the accentual peak in the routine call are analyzed as a sequence of a downstepped !H-phrase accent and a H% boundary tone. The phrase accent is analyzed as !H-, rather than L-, as it is scaled much higher than all the other targets that can be reasonably considered to be reflexes of L tones; this group includes IT and RO both in the routine call itself and in the urgent melody, and the uncontroversially L tone at the end of the urgent call (cf. fig. 3, 7). In addition, evidence from other Polish melodies suggests that L tones are not upstepped between H tones even in extreme tonal crowding. This is illustrated in figure 10 which shows a melody used with Polish wh-questions to indicate incredulity; as can be seen, even with the monosyllabic wh-word *co* ('what') the dip between the two peaks reaches the bottom of the speaker's range. Given contours like that in figure 10, the dip in the routine call cannot be plausibly attributed to an upstepped L tone as this would require a similar realization in wh-questions. Finally, the use of H% reflects the small but consistent final rise of the melody.

The analysis of the routine call's final part as consisting of two distinct tonal events, a phrase accent and a boundary tone, is based on the following observations. First, it is clear from the results on the temporal and F0 distance between the !H- (i.e. F0 dip or L target in the measurements) and the two H tones that flank it (i.e. the accentual H*

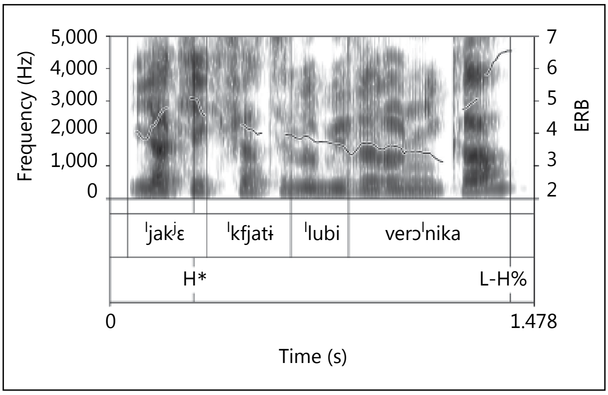

Fig. 11. The melody used with Polish wh-questions to indicate incredulity as realized on a long utterance, *Jakie kwiaty lubi Weronika?* 'What kind of flowers does Veronika like?'

and the H% boundary tone) that this target is not strongly connected to the accentual H* (H1 in the measurements): it is further away from it than from H% and does not have a consistent relationship with it either in terms of scaling or in terms of alignment. Thus, it is unlikely that the tone represented here as an independent !H- phrase accent is part of the melody's pitch accent instead. Second, inspection of other Polish melodies indicates that tones spread between the last pitch accent and the boundary tone: this is illustrated in figure 11 which shows the contour of a long wh-question uttered with the same melody as *co* 'what' in figure 10. This spreading is consistent with the presence of a phrase accent (Grice et al., 2000; cf. Barnes et al., 2006). If phrase accents are a necessary part of the Polish intonational system, parsimony and analytical coherence (Gussenhoven, 2007) dictate a uniform analysis involving phrase accents for all melodies. Thus, it appears best to analyze the dip and subsequent rise of the Polish routine call as the reflexes of two distinct tonal events, a !H- phrase accent, followed by a H% boundary tone. In turn, this analysis indicates that the explicit marking of downstep is necessary for the Polish intonation system. This is also indicated, as noted above, by the fact that it is possible to find LH and HL sequences that do not involve upstepping of L tones or downstepping of H tones: the former is exemplified by the melody illustrated in figures 10 and 11, the latter by the urgent calling melody itself (fig. 1b, 2b). Finally, the end part of the urgent calling melody, a plain fall, is analyzed as a L-L% sequence for reasons of system-internal consistency (cf. Gussenhoven, 2007; Arvaniti, 2016).

4.3. Repercussions for the Phonological Representation of Intonation

The phonology of the Polish calling melodies has repercussions for the representation of intonation in general and the notion of similarity that is often relied upon in cross-linguistic comparisons of intonational patterns. As shown in table 2, the representations of the Polish calls differ in some respects from representations of similar calling melodies in other languages. As noted in the introduction, this kind of variability has been the focus of discussion in AM, as it is thought to be a hindrance to cross-linguistic comparison (e.g., Ladd, 2008a, b), though others argue it is inevitable and possibly desirable (Arvaniti, 2016). It is thus of interest to this debate to consider the melodies studied here in relation to similar melodies in other languages.

Inspection of table 2 quickly shows that the differences in representation are not all of one type. Rather, disagreements reflect (i) distinct theoretical positions, (ii)

Table 2. AM representations of melodies associated with routine calls (top) and urgent calls (bottom)

Language	Pitch accent	Phrasal tones
Routine calls		
Arabic (Chahal and Hellmuth, 2014)	L+H*	H–L%
Catalan (Borràs-Comes et al., 2015)	L+H*	!H%
Dutch (Gussenhoven, 2005)	H*!H*	%
English (Brugos et al., 2006)	H* or L+H*	!H–L%
German (Grice et al., 2005)	L+H*	H–%
Greek (Arvaniti and Baltazani, 2005)	L*H	!H–!H%
Hungarian (Varga, 2008)	H*	!H–0%
Polish (this study)	LH*	!H–H%
Portuguese (Frota, 2014; Frota et al., 2015)	(L+)H*	!H%
Urgent (or insistent) calls		
Catalan (Prieto, 2014)	L+H*	HL%
Portuguese (Frota, 2014; Frota et al., 2015)	H*	L%
Polish (this study)	H*	L–L%

matters of convention, (iii) genuine differences between systems. The use of two tonal events to represent the final rise in the routine call of Polish, !H- and H%, reflects the first type of disagreement: analyses differ as to whether phrase accents are considered to be distinct tonal entities or not (cf. Brugos et al., 2006; Gussenhoven et al., 2003; Gussenhoven, 2005; for a discussion see also Ladd, 2008b). Such analytical differences reflect the difficulty of dividing an F0 curve into discrete elements and highlight the need for specific criteria that may allow researchers to do so with some consistency (for a discussion, see Arvaniti, 2016). In contrast, the use of LH* in Polish and L+H* in German is largely a notational difference, i.e. a matter of convention and thus easily amenable to standardization.

In contrast to examples like those discussed above, other disagreements in table 2 reflect differences among the systems themselves. One such instance is the analysis of the Dutch melody as H*!H* % (i.e. with the rise and subsequent downstep as one unit), which reflects the fact that the Dutch melody applies in steps to successive feet when possible (Grice et al., 2000; Gussenhoven et al., 2003; Gussenhoven, 2005). An analysis of this sort would be clearly unsuitable for Polish in which the entire melody is confined to the last foot even in long calls. This difference in the melody's association with Polish metrical structure is illustrated in figure 2a, which shows a word that phonologically consists of two feet (Rubach and Booji, 1985; Kraska-Szlenk, 2003); it is even clearer in figure 12 which shows a much longer version of the Polish routine call. As can be seen in figure 12, F0 remains low and flat until the last foot of the entire utterance, whereas in a linguistic system like Dutch it would be manifested as a series of three steps, one on each foot (Gussenhoven, 2005).

As noted in the introduction, an intriguing difference between Polish and other languages relates to the edge tones of the Polish routine call, !H-H%, which indicate a step down in pitch followed by a rise. This differs from the other routine calls in table 2, which are best characterized as instantiations of the vocative chant since they end in sustained level pitch. Though this difference is relevant for and reflected in the AM representation of the Polish routine call, it is important to recognize that it pertains

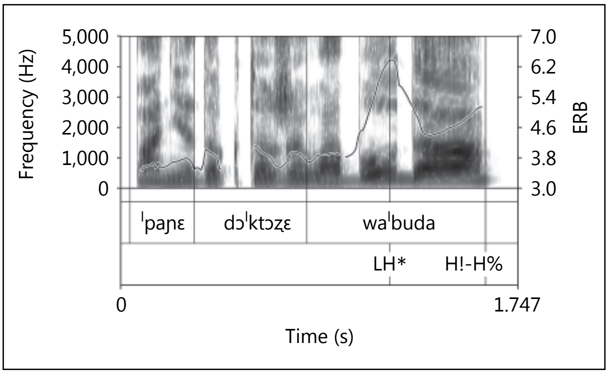

Fig. 12. The Polish routine call as realized on a longer utterance *Panie Doktorze Łabuda* 'Mr Dr Wabuda'.

to form alone and not to the function of the melody. In some of the languages represented in table 2, such as English (Gussenhoven, 2004; Ladd, 2008a), and Hungarian (Varga, 2008), the vocative chant is used to convey 'routineness': it is used to call for an everyday reason someone who is not in the immediate vicinity. This is the context in which the Polish routine call was also elicited: speakers were asked to imagine calling a child who is in the garden. This meaning, however, does not apply to all the melodies in table 2. Frota (2014) notes that the vocative chant is used in Portuguese for a first call or as a greeting. Borràs-Comes et al. (2015) found that for Catalan speakers 'the vocative chant L+H* !H% [was used] for insistent calls' (p. 78). Gibbon (1998), whose description of the pragmatics of the German vocative chant is followed by Grice et al. (2005), notes that in German the melody is not used only for calls, but also for a range of other functions, including greeting, leave-taking and thanking, and for discourse repairs caused by mishearing. Some of the examples provided by Gussenhoven (2005) suggest that the vocative chant has similar functions in Dutch as well.

In short, the calling contours that can best be described as vocative chants in terms of form are not used for the same purposes across all languages that have them. Conversely, the function often associated with the vocative chant in the literature, namely calling someone not in the vicinity for a routine reason, may be accomplished using melodies comparable but not identical to the vocative chant, like the one used in Polish.[3]

[3] We note that among the discarded data from the present study were instances of a melody that ended in sustained pitch; this melody we would provisionally analyze as LH* !H-%, to indicate the drop in pitch and its subsequent sustainment. However, we stress that at present it is not possible to ascertain that this melody is indeed phonologically distinct from LH* !H-H%, i.e. used for different purposes and with a different meaning (possibly determined by dimensions like those examined by Borràs-Comes et al., 2015). Additional research on the production and perception of the Polish calling melodies would be necessary to determine this point. Nevertheless, the presence of this melody does not distract from the fact that LH* !H-H% was used here consistently and unprompted by the majority of the study participants for the purpose of calling for a routine reason a person who is not close-by, i.e. in the prototypical context for the vocative chant. Finally, given that, as mentioned earlier, melodies similar to the Polish routine call have been reported also for English and Dutch (Gussenhoven, 2004, 2005), though without discussion of their function, it is possible that the vocative chant and the routine call are variants of the same underlying melody, with one variant being more prevalent in some languages than in others. This is clearly a matter that deserves further attention, but is beyond the scope of this paper.

Phonetica 2016;73:338–361
DOI: 10.1159/000446001

Such cross-linguistic discrepancies abound and are unsurprising; for instance, the melodies used in Portuguese for requests and commands (Frota, 2014) are similar in form to melodies attested in Greek but with precisely the reverse meaning (Arvaniti and Baltazani, 2005; Baltazani, 2006); similarly, the Catalan L+H* HL% is considered appropriate for most instances of calling someone (Borràs-Comes et al., 2015), but the Polish contour that most resembles it is used to show disapproval instead. What this discussion serves to show then is that clear-cut cross-linguistically valid categories – such as 'vocative chant' – may be fictions as they can only be determined if one relies exclusively on *phonetic* form and ignores meaning altogether. However, if we accept a basic AM tenet, namely that tonal elements are morphemes with pragmatic meaning, then classification should be based on both meaning and form. The Polish routine and urgent calls, in comparison to their counterparts in other languages, bring this point to the fore and reveal issues that need to be addressed in AM regarding the nature of intonational meaning and the relationship between phonetics and phonology (Arvaniti, 2016).

4.4. Repercussions for the Modeling of Intonation

As noted in the introduction, the study relied on tonal crowding to examine the phonetics of the melodies and used it as a diagnostic for determining their phonology (see section 4.2). In addition, however, effects of tonal crowding can serve as a testing ground for different models of intonation (see Arvaniti and Ladd, 2009, 2015). This is because particular models rely on different conceptions of how F0 relates to elements of the segmental string.

In AM it is assumed that F0 is the reflex of phonological structure, specifically of H and L tones, pitch accents that associate with the heads of prosodic constituents, and phrasal tones (phrase accents and boundary tones) that associate with prosodic boundaries. The independence of the tonal elements from the segmental string entails that the two are not in a 1-to-1 relationship: a given utterance may involve a long string of segments but few tones, or a small number of segments but several tones. The former is exemplified by utterances like that in figure 12, which includes eight syllables but just four tones viz. LH* !H–H%; in this instance, the melody is seen as both phonologically and phonetically *underspecified*, i.e. parts of the utterance have no specification for pitch which is instead derived by interpolation between the specified points. The opposite situation, having more tones than tone-bearing units, is tonal crowding and is evidenced here particularly in the routine call when it is realized on monosyllables: in words like *Piotr*, there is only one tone-bearing unit, the vowel [ɔ], which is associated with the same four tones as *Panie Doktorze Łabuda*. As shown above, tonal crowding leads to adjustments in the realization of tones: e.g. the peak which is the reflex of H* is realized earlier, closer to the onset of the vowel and is also scaled lower. Crucially, however, all four tones are realized and this becomes possible both because the vowel lengthens to accommodate the melody and because the location of peaks and dips is adjusted. Such adjustments follow naturally from the way AM conceptualizes the structure of intonation and its relation to F0; this is because both underspecification and tonal crowding are predicted, and because from a phonetic perspective both anticipatory and carry-over coarticulation of tones is possible (Arvaniti and Ladd, 2009).

In other models, however, this relationship is conceptualized in a different way. In models like PENTA (among others, Xu et al., 2015), the assumption is that each syllable has a *pitch target* which is related to a specific communicative function; calls for

DOI: 10.1159/000446001

a routine or an urgent purpose would be treated as two such functions in this model. This conceptualization of intonation is problematic for reasons related to both phonology and phonetic implementation. First, PENTA's primitives are rises, falls and levels and only one of these can be the pitch target of a given syllable; however, in the Polish routine call a rise-fall-rise must be realized on one syllable if the associated word is monosyllabic, e.g. *Piotr*. The realization of the rise-fall-rise on one syllable can only be effected by making *ad-hoc* adjustments to the model so that rise-fall-rises can be treated as primitives. Doing so, however, leads to the unsatisfactory conclusion that a variety of different pitch targets serves the same communicative function. As discussed in detail in Arvaniti and Ladd (2009), this is neither parsimonious nor adequate as it cannot account for the way in which speakers learn, produce and process the melodies of their language; in other words, models like PENTA miss significant generalizations about intonation patterns. In the present instance, PENTA cannot even explain the pattern attested when its primitives match the syllables in number: in trisyllabic words like *Marianna*, the pitch movements are still not evenly distributed one per syllable (fig. 1). Further, the type of full pitch specification advocated by PENTA cannot explain why some elements of the melodies are resistant to change, while others are dispensable: as shown earlier, the rise to a peak is present in the routine call even in monosyllables, but it is dispensed with in the same environment when the urgent call is used. Assuming that *target strength*, a parameter that controls the realization of pitch targets in PENTA, is the same for the routine and urgent call pitch targets (and there is no reason to assume otherwise), this difference remains unexplained. Finally, the Polish results support previous findings, which show clear anticipatory effects due to tonal crowding, such as the earlier occurrence of peaks in words like *Piotr* as compared to longer items like *Magdalena* (cf. fig. 1, 2), yet PENTA does not include an anticipatory mechanism at all. Given the above, it is clear that the present data from Polish add to an increasingly large literature which reports results compatible with the basic tenets of AM but not with those of PENTA.

5. Conclusion

In conclusion, the phonetics of the Polish routine and urgent calls, together with their comparison to other Polish melodies, lead us to propose an autosegmental representation of the former as LH* !H-H% and of the latter as H* L-L%. Our study also revealed small but consistent differences in the realization of the accented vowel as realized with each melody, adding new data to recent evidence about the effects of intonation on the realization of segments. In addition, the present study is of interest from a typological perspective: the urgent call is similar in form to a melody functioning in a comparable manner in Portuguese, while the routine call is similar in function to the vocative chant used in languages like English and Hungarian, but differs from them in form in that it ends in a rise rather than sustained level pitch. This mismatch between form and function when several languages are considered is a point that requires attention within AM: although the intonational primitives are said to be morphemes with pragmatic meaning, the role of meaning is underplayed and melodies are often classified as similar based largely on form. Finally, by using tonal crowding as a diagnostic we were able to detect changes in the phonetics of the melodies when realized on minimal segmental material such as monosyllabic words. The systematic adjustments we found in such contexts advocate in favor of underspecified contours as envisaged by AM but are not readily accountable in models of syllable-by-syllable specification such as PENTA.

Acknowledgments

This research has been supported by University of Kent research funds to Amalia Arvaniti and Marzena Żygis, and by Bundesministerium für Bildung und Forschung (BMBF, Germany) Grant Nr. 01UG1411 to Marzena Żygis. We are grateful to the reviewers and the Editor for their helpful comments and suggestions.

References

Arvaniti A (2011): The representation of intonation; in van Oostendorp M, Ewen CJ, Hume E, Rice K (eds): Companion to Phonology. Wiley-Blackwell, Oxford.

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 7:1–43.

Arvaniti A, Baltazani M (2005): Intonational analysis and prosodic annotation of Greek spoken corpora; in Jun S-A (ed): Prosodic Typology: The Phonology of Intonation and Phrasing. Oxford, Oxford University Press, pp 84–117.

Arvaniti A, Ladd DR, Mennen I (2006): Tonal association and tonal alignment: evidence from Greek polar questions and contrastive statements. Lang Speech 49(pt 4):421–450.

Arvaniti A, Garding G (2007): Dialectal variation in the rising accents of American English; in Cole J, Hualde JI (eds): Papers in Laboratory Phonology 9. Berlin, New York, Mouton de Gruyter, pp 547–576.

Arvaniti A, Ladd DR (2009): Greek wh-questions and the phonology of intonation. Phonology 26:43–74.

Arvaniti A, Ladd DR (2015): Underspecification in intonation revisited: a reply to Xu, Lee, Prom-On & Liu. Phonology 32:537–541.

Baltazani M (2006): Intonation and pragmatic interpretation of negation in Greek. J Pragmatics 38:1658–1676.

Bates D, Maechler M, Bolker B, Walker S, Christensen RHB, Singmann H, Dai B, Grothendieck G (2015): lme4. R package version 1, pp 1–10.

Barr DJ, Levy R, Scheepers Ch, Tily HJ (2013): Random effects structure for confirmatory hypothesis testing: keep it maximal. J Mem Lang 68:255–278.

Beckman ME (1986): Stress and Non-Stress Accent. Dordrecht, Foris.

Barnes J, Shattuck-Hufnagel S, Brugos A, Veilleux N (2006): The domain of realization of the L-phrase tone in American English. Proc Speech Prosody, Dresden.

Borràs-Comes J, Sichel-Bazin R, Prieto P (2015): Vocative intonation preferences are sensitive to politeness factors. Lang Speech 58:68–83.

Brugos A, Shattuck-Hufnagel S, Veilleux N (2006): Transcribing Prosodic Structure of Spoken Utterances with ToBI (MIT Open Courseware). http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6–911-transcribing-prosodic-structure-of-spoken-utterances-with-tobi-january-iap–2006/.

Chahal D, Hellmuth S (2014): The Intonation of Lebanese and Egyptian Arabic; in Jun S-A (ed): Prosodic Typology II. Oxford, Oxford University Press, pp 365–404.

Demenko G (1999): Analiza cech suprasegmentalnych języka polskiego na potrzeby technologii mowy. Poznań, Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza.

Demenko G (2015): Korpusowe badania języka mówionego. Warszawa, Akademicka Oficyna Wydawnicza Exit.

Dłuska M (1947): Prozodia języka polskiego. Kraków, Polska Akademia Umiejętności.

Dukiewicz L (1978): Intonacja wypowiedzi polskich. Wrocław, Zakład Narodowy imienia Ossolińskich.

Francuzik K, Karpiński M, Kleśta J, Szalkowska E (2004): Nuclear melody in Polish semi-spontaneous and read speech. Evidence from the Polish Intonational Database PoInt. Studia Phonetica Posnaniensia 7:97–128.

Frota S (2014): The intonational phonology of European Portuguese; in Jun S (ed): Prosodic Typology II: The Phonology of Intonation and Phrasing. Oxford, Oxford University Press, pp 6–42.

Frota S, Cruz M, Svartman F, Collischonn G, Fonseca A, Serra C, Pedro O, Vigário M (2015): Intonational variation in Portuguese: European and Brazilian varieties; in Frota S, Prieto P (eds): Intonation in Romance. Oxford, Oxford University Press, pp 235–283.

Gibbon D (1998): Intonation in German; in Hirst D, Di Cristo A (eds): Intonation Systems: A Survey of 20 Languages. Cambridge, Cambridge University Press, pp 78–95.

Gordon M (2002): A phonetically-driven account of syllable weight. Language 78:51–80.

Gordon M (2005): A perceptually-driven account of onset-sensitive stress. Nat Lang Linguist Theory 23:595–653.

Grice M, Ladd DR, Arvaniti A (2000): On the place of 'phrase accents' in intonational phonology. Phonology 17:143–185.

Grice M, Baumann S, Benzmüller R (2005): German intonation in autosegmental-metrical phonology; in Jun S-A (ed): Prosodic Typology: The Phonology of Intonation and Phrasing. Oxford, Oxford University Press, pp 55–83.
Gussenhoven C (2004): The Phonology of Tone and Intonation. Cambridge, Cambridge University Press.
Gussenhoven C (2005): Transcription of Dutch intonation; in Jun S-A (ed): Prosodic Typology: The Phonology of Intonation and Phrasing. Oxford, Oxford University Press, pp 118–145.
Gussenhoven C (2007): The phonology of intonation; in de Lacy P (ed): The Cambridge Handbook of Phonology. Cambridge, Cambridge University Press, pp 253–280.
Gussenhoven C, Rietveld T, Kerkhoff J, Terken J (2003): ToDI second edition. http://todi.let.kun.nl/ToDI/home.htm.
Jassem W (1987): Computer-based classification of basic Polish intonations. Proceedings of the 11th ICPhS, Talin, pp 253–256.
Jassem W, Demenko G (1986): Extracting linguistic information from F0 traces; in Catherin J-L (ed): Intonation in Discourse. Croom & Helm, London.
Karpiński M (2006): Struktura i intonacja polskiego dialogu zadaniowego. Poznań, Wydawnictwo Naukowe UAM.
Kraska-Szlenk I (2003): The phonology of stress in Polish. Lincom, München.
Kuznetsova A, Brockhoff PB, Christensen RHB (2015): Package ‘lmerTest’. R package version 2.0–29.
Ladd DR (2008a): Intonational Phonology. Cambridge, Cambridge University Press.
Ladd DR (2008b): Review of Jun S (ed): (2005). Prosodic Typology: The Phonology of Intonation and Phrasing. Oxford, Oxford University Press. Phonology, vol 25, pp 372–376.
Liberman M (1975): The Intonational System of English. PhD thesis, MIT.
Niebuhr O (2008): Coding of intonational meanings beyond F0: evidence from utterance-final /t/ aspiration in German. J Acoust Soc Am 124:1252–1263.
Niebuhr O (2012): At the edge of intonation: the interplay of utterance-final F0 movements and voiceless fricative sounds. Phonetica 69:7–21.
Pierrehumbert J (1980): The Phonology and Phonetics of English Intonation. PhD Thesis, MIT.
Prieto P (2014): The intonational phonology of Catalan; in Jun S-A (ed): Prosodic Typology II: The Phonology of Intonation and Phrasing. Oxford, Oxford University Press, pp 43–80.
Rubach J, Booji GE (1985): A grid theory of stress in Polish. Lingua 66:281–319.
R Development Core Team (2008): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org.
Sluijter AMC, van Heuven VJ (1996a): Spectral balance as an acoustic correlate of linguistic stress. J Acoust Soc Am 100:2471–2485.
Sluijter AMC, van Heuven VJ, Pacilly JJA (1996b): Spectral balance as a cue in the perception of linguistic stress. J Acoust Soc Am 101:503–513.
Steffen-Batogowa M (1996): Struktura przebiegu melodii języka polskiego ogólnego. Poznań, Sorus.
Varga L (2008): The calling contour in Hungarian and English. Phonology 25:169–497.
Wagner A (2008): A comprehensive model of intonation for application in speech synthesis. PhD dissertation. Adam Mickiewicz University in Poznan.
Xu Y, Lee A, Prom-on S, Liu F (2015): Explaining the PENTA model: a reply to Arvaniti and Ladd. Phonology 32:505–541.
Żygis M, Pape D, Jesus L, Jaskuła M (2014): Intended intonation of statements and polar questions in Polish in whispered, semi-whispered and normal speech modes. Speech Prosody. http://www.speechprosody2014.org/.

Phonetica 2016;73:338–361
DOI: 10.1159/000446001