Journal of Phonetics 63 (2017) 53–74

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Research Article

Segmental cues to intonation of statements and polar questions in whispered, semi-whispered and normal speech modes

Marzena Żygis [a,b,*], Daniel Pape [c], Laura L. Koenig [d], Marek Jaskuła [e], Luis M.T. Jesus [f]

[a] Leibniz-Centre General Linguistics (Leibniz-ZAS), Schützenstr. 18, 10-117 Berlin, Germany
[b] Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany
[c] Department of Linguistics and Languages, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada
[d] Haskins Laboratories, 300 George Street, New Haven & Adelphi University, Garden City, NY 11530, USA
[e] Faculty of Computer Science and Information Technology, West Pomeranian University of Technology, Zolnierska 52, 71-210 Szczecin, Poland
[f] Institute of Electronics and Informatics Engineering of Aveiro (IEETA), and School of Health Sciences (ESSUA), University of Aveiro, 3810-193 Aveiro, Portugal

ARTICLE INFO

Article history:
Received 18 July 2016
Received in revised form 28 March 2017
Accepted 3 April 2017
Available online 10 May 2017

Keywords:
Segment–prosody interaction
Sibilants
Intonation
Whispered speech
Polish

ABSTRACT

This paper examines how acoustic characteristics of vowels and consonants reflect intonational differences between polar questions and statements in Polish whispered, semi-whispered and normal speech modes, with particular focus on the spectral characteristics of voiceless consonants as a function of intonation, and across speech modes. The results reveal significant differences in spectral properties of both utterance-final vowels and consonants across statements and polar questions. Questions have higher vowel intensities and show differences in formant frequencies that vary with speech mode. Regarding the consonants, both fricatives and affricates are produced with higher intensity, spectral peaks at higher frequencies, and higher Centre of Gravity and Spectral Standard Deviation values in questions than in statements. Conversely, skewness and kurtosis are lower in questions than in statements. Some spectral features of sibilants, including spectral slopes, show greater question-statement differences in the whispered speech mode than in other speech modes. The finding that some cues are more pronounced in whispered speech suggests that they may compensate for the absence of fundamental frequency in this mode. Most generally, the study shows that speakers produce intended intonation patterns by varying the type and magnitude of cues depending on speech mode.

1. Introduction

This paper assesses the acoustic characteristics of vowels and voiceless obstruents in normal, whispered, and semi-whispered speech from speakers of Polish. Our primary questions are how intonational distinctions between yes–no questions and statements are produced in these different speaking modes, and to what extent the voiceless fricatives and affricates reflect the intonational differences in the three modes. To verify the expected intonational patterns and allow for comparison with past work on whispered speech, we also present data on vowel formants and durations and, for the normal voiced condition and, for the normal voiced condition, the fundamental frequency (F0). The results provide evidence for intonational variations in both vowels and voiceless segments across all speaking modes and speak to the variety of ways in which intonation can be manifested in the speech signal aside from the variations in fundamental frequency usually associated with intonational patterns.

The first two sections of the literature review address the relatively new line of work on interactions between segments and prosody in general (Section 1.1) and segments and intonation in particular (Section 1.2). Section 1.3 reviews the acoustic differences between whispered and normal (modal) speech. Semi-whispered speech has not received previous acoustic description. It is included as a third speaking mode here to provide a broader view of how intonational patterns appear in speech conditions where F0 is not reliably available in the speech signal.

* Corresponding author at: Leibniz-Centre for General Linguistics (ZAS), Schützenstr. 18, 10-117 Berlin, Germany. Fax: +49 30 20192 402.
E-mail addresses: zygis@zas.gwz-berlin.de (M. Żygis), lesondumonde@web.de (D. Pape), koenig@haskins.yale.edu (L.L. Koenig), Marek.Jaskula@zut.edu.pl (M. Jaskuła), lmtj@ua.pt (L.M.T. Jesus).

http://dx.doi.org/10.1016/j.wocn.2017.04.001
0095-4470/

1.1. Prosody, segments, and their interaction

One goal of this work is to contribute to the growing literature on interactions between segments and prosody. Traditionally, research into speech and language focused on either segments or prosodic patterns (see Kohler, 2012 for an overview). To our knowledge, the earliest studies of segment–prosody interaction dealt with sonority and syllable structure across languages (Hooper, 1976; Selkirk, 1984). In recent decades, more research has explored interactions between segments and prosody, based on the growing awareness of prosodic levels, as described, for example, in Selkirk (1978, 1986), Nespor and Vogel (1986), and Beckman and Pierrehumbert (1986). In these models, which are known Prosodic Hierarchy models, phonological units combine into increasingly larger units: Segments combine to form syllables, which combine to form prosodic feet, and then phonological words and phrases, and finally utterances. The next paragraph provides examples to demonstrate how broadly this general theoretical framework has been applied; we will then focus on how segments, and particularly consonants, vary as a function of intonation (Section 1.2).

Numerous prosodic phenomena may impact segmental characteristics. Demonstrations of boundary effects can be found, for example, in Shattuck-Hufnagel and Turk (1998), Fougeron (2001), Cho and Keating (2001, 2009), Byrd and Saltzman (2003), and Katsika (2016). Cho and Keating (2009) observed that vowels in CV syllables had higher amplitudes in domain-initial position, i.e., boundary effects may span multiple segments. Studies demonstrating sentential stress effects include the work of Pierrehumbert (1980) and Sluijter (1995). Although much work in this area has emphasized vowel characteristics, a few studies have documented prosodic variation in consonants as well. For instance, Fougeron and Keating (1997) reported more linguo-palatal contact for /n/ at the beginning of higher prosodic domains; in contrast, /o/ had less linguopalatal contact in domain-final syllables compared to initial and medial positions (see also Fougeron, 2001). Along similar lines, Cho and Keating (2001) observed that Korean alveolar consonants had more extensive articulatory contact and longer durations in higher prosodic domains than in lower domains. Cho and McQueen (2005) found that lexical stress, accent and prosodic constituent size all affected consonant durations in Dutch. Cho (2015) provides a summary of timing effects induced by prosody.

1.2. Intonation and segments

In many languages, including Polish, polar questions are characterized by a terminal F0 rise (Wagner, 2008). In the autosegmental framework (e.g., Pierrehumbert, 1980), this is represented as a high boundary tone. Such a description effectively takes F0 to be the primary attribute of intonational differences. On the other hand, Pierrehumbert and Talkin (1992) made the point that F0 need not be the sole carrier of intonational differences, and other authors have recognized that multiple phonetic features may vary as a function of intonation. For example, Ladd (1996:6/2008:4) describes intonation broadly as "the use of suprasegmental phonetic features to convey 'postlexical' or sentence-level pragmatic meanings". The suprasegmental features include fundamental frequency, intensity and duration (Ladd, 1996, p. 6). Grice (2006) also lists multiple phonetic 'channels', including segmental features, which can convey intonation in addition to the 'perceived pitch' (Grice, 2006, p. 779). In line with this perspective, several studies have described interactions between segments and intonation. As with other investigations of segment–prosody interaction (see Section 1.1), these interactions have mainly concentrated on vowels (e.g., Pierrehumbert, 1980; Prieto, van Santen, & Hirschberg, 1995), even in languages with numerous voiceless consonant clusters, such as Polish or Berber (e.g., Dukiewicz, 1978; Gordon & Nafi, 2012; Roettger & Grice, 2015; Steffen-Batogowa, 1966). This is justifiable if we take F0, obviously found only in voiced segments, to be the main correlate of intonation; indeed, most experimental designs on intonation have explicitly avoided voiceless obstruents since they can lead to micro-prosodic perturbations in the F0 contour of adjacent voiced segments and interrupt the smooth patterns of F0 (e.g., Kohler, 1990).

However, recent studies have revealed that voiceless segments are also sensitive to intonational changes, suggesting that excluding them from investigation limits our understanding of intonational variation. Niebuhr (2008) found that the aspiration of German /t/ in utterance-final position under two accent contours distinguished by peak F0 placement differed in duration, intensity and spectral peak frequencies (specifically, burst frequencies were shifted to higher regions in high-F0 conditions). Furthermore, the German fricatives /ʃ/ and /x/ have been seen to vary in the Centre of Gravity (COG) depending on intonation contours: in high-raising (surprised) questions the sibilants were produced with higher COG and compressed COG ranges whereas in falling (concluding) statements they showed lower COG values and higher ranges (Niebuhr, 2009). Finally, Niebuhr, Lill, and Neuschulz (2011) and Niebuhr (2012) investigated the German voiceless sibilants /s/ and /ʃ/ in different intonation contexts. They reported that the fricatives had higher COG values in questions than statements.

In light of these findings, three questions arise: First, to what extent can they be generalized to languages other than German? Second, the previous studies investigated single consonants appearing in coda position. Do longer voiceless sequences display similar characteristics? For example, intonational effects in preceding vowels might carry over into a single following voiceless consonant, but such effects might dissipate over time, i.e., not be as salient in clusters as in singletons. The work of Cho and Keating (2009; see Section 1.1) did indicate that some prosodic effects could span multiple segments, but their results were for boundary effects and not intonation; moreover, vowels and consonants could behave differently in this regard. Finally, all results on intonational variation in consonants have come from voiced speech. It remains unclear to what extent similar relations can be found in speech modes where F0 may be partially absent as in semi-whispered

speech or totally absent as in whispered speech. Past work has suggested that whispered speech, despite lacking an F0 contour, still allows listeners to discern some aspects of intonation (e.g., Heeren & van Heuven, 2009; cf. next section). This observation leads to fundamental questions about how intonational patterns are manifested[1] in whispered speech. The second focus of this work is thus to explore prosodic variation in non-phonated speech modes. Which parameters might allow listeners to differentiate between whispered questions and statements? The next section summarizes previous research on whispered and semi-whispered speech giving special attention to segment–prosody interaction.

1.3. Whispered and semi-whispered speech

Work on whispered speech has established that this unphonated speaking mode conveys considerable information to listeners, including aspects of speaker differences as well as vowel and consonant identity (e.g., Kallail & Emanuel, 1984; Tartter, 1989). The basic acoustic differences between voiced (modal) speech and whisper have also been described in some detail. Overall, whispered vowels have decreased amplitudes compared to voiced vowels; Ito, Takeda, and Itakura (2005) obtained a difference of about 20–25 dB across the spectral envelope. Some studies also observed that whispered vowels are longer than their voiced counterparts (Schwartz, 1968; Sharf, 1967), whereas others found no differences (Heeren, 2015a). One of the most widely-studied aspects of whisper is vowel formant frequencies (e.g., Heeren, 2015a; Heeren & van Heuven, 2011; Higashikawa, Nakai, Sakakura, & Takahashi, 1996; Kallail & Emanuel, 1984; Li & Xu, 2005; Meyer-Eppler, 1957; Morris, 2003; Thomas, 1969; Sharifzadeh, 2010). For example, Ito et al. (2005) reported that formants of Japanese /a, i, u, e, o/ shifted towards higher frequencies in the whispered speech mode as compared to the voiced speech mode. F1 in whispered vowels was about 1.3–1.6 times higher than in the corresponding voiced vowels; for F2 the increase was in the range of 1.0–1.2. Kallail and Emanuel (1984) found systematically higher values of the first three formants in whispered American vowels /i, u, æ, a, ʌ/ compared to the voiced counterparts. The data also revealed that F1 underwent larger changes than F2 or F3. Along similar lines, Eklund and Traunmüller (1996), investigating ten Swedish vowels in stressed positions in whispered and voiced speech, found that F1 was raised more than F2 in whispered speech. Thus, there is a general consensus that vowel formants are raised in whispered speech compared to voiced speech, especially for F1 but possibly for higher formants as well. In a modeling study, Swerdlin, Smith, and Wolfe (2010) demonstrated that the increased glottal areas associated with whisper had the strongest and most consistent effects on F1. However, formant changes in whisper may vary across vowels (Meyer-Eppler, 1957); further, some investigations suggested that increased values of F2 as well as F1 may correlate with pitch percepts in whisper (Higashikawa & Minifie, 1999; Thomas, 1969).

Many studies of whisper investigated individual vowels irrespective of their prosodic position and often produced in isolation. However, some work has attended to how prosody influences the spectral properties of whispered vowels. Higashikawa et al. (1996) asked 12 Japanese informants to produce the vowel /a/ in voiced speech and in whisper at ordinary, high, and low pitches. The results showed that the F1 frequency was significantly higher (i) in ordinary whispering than in voiced speech and (ii) in high-pitched whispering than in low-pitched whispering. F1, F2 and F3 all had a tendency to increase in high-pitched whispering and decrease in low-pitched whispering as compared to ordinary whispering. Heeren (2015a) examined the Dutch vowels /a, i, u/ in CVCs produced at low, mid, and high pitches in whispered and voiced speech. Along with higher F1 and F2 in whisper, the author reported a more extreme difference in spectral balance (defined as the intensity difference between the 0.5–2 kHz and 2–8 kHz bands) as a function of low vs. high pitch targets in whispered speech as compared to modal speech. Intensity was lower in whispered than in modal speech and lower in /i/ and /u/ than /a/ in accordance with intrinsic vowel intensity (Lehiste, 1970). Furthermore, Centre of Gravity values (calculated from 0.05 to 8 kHz) were higher in whispered than in normal speech and COG differences between high vs. low and high vs. medium pitch were larger in whispered than in normal speech. Neither the speech mode nor the pitch target affected the duration of vowels.

The work of Meyer-Eppler (1957) on German suggested that amplitude could play a role in perceived pitch in whisper (see also Thomas, 1969). More recent work has evaluated the role of amplitude in the perception of whispered tones. Whalen and Xu (1992) showed that in the absence of F0 and formant structure, the amplitude information for tones 2, 3 and 4 in Mandarin was fairly distinct and was used by listeners as a cue for tone identification. The authors suggested that differences in amplitude gave rise to weak F0 percepts because their original stimuli, i.e. before removing F0 and formants, were characterized by a strong correlation between amplitude and F0. However, Abramson (1972) did not find clear evidence for a role of amplitude in perceiving whispered Thai tones.

Little attention has been devoted to consonants in whispered speech. In the case of tonal/intonational differences, this might have been a natural extension of the emphasis on vowels and sonorants in studies of intonation in voiced speech. Nevertheless, the relative neglect of consonants in whisper is somewhat curious insofar as consonants are important carriers of information in the speech signal (cf. Tarttar, 1989). Yet it is also the case that whispered speech looks totally different from voiced speech: Due to the lack of periodic glottal excitation and to the presence of noise excitation voice source harmonics are completely absent and spectrograms are dominated by strong aperiodic energy. Therefore, as stated by Lim (2011, p. 30), “the obvious remaining indicators of the message for whispered speech appear to be largely the formant energies.” It may be that the lack of appropriate techniques to study spectral properties of voiceless consonants has limited investigations of obstruents in whispered speech. One possibility that we will evaluate is that intonation patterns are conveyed in different

[1] We use the word 'manifested' here to emphasize the production-based nature of our work; that is, our question is to what extent acoustic differences are available in the speech signal. Determining the degree to which listeners use such features would require a detailed perception study, beyond the bounds of the current work.

speech modes by means of cue trading (see Repp, 1982, for a review) involving spectral features, particularly of consonants. Cue-trading is defined as the use of different acoustic cues guiding the perception of linguistic distinctions, where several cues (perhaps with differing magnitudes) can integrate to form a robust percept, in contrast to a single major cue being sufficient to generate the perceptual outcome. In the case of intonation patterns, we must specifically ask what other acoustic features may be available to listeners when the F0 is absent, as in whispered and semi-whispered speech.

One of the few acoustic studies of whispered consonants was conducted by Jovičić and Šarić (2008), who investigated duration and average root mean square intensity (RMS) of 25 Serbian consonants. The results revealed that the RMS intensity of voiced consonants was reduced by as much as 25 dB in the whispered mode, whereas voiceless consonants showed almost unchanged RMS intensity. Whispered consonants were, on average, about 10% longer in duration. Schwartz (1972) reported that stop closure durations for /p/ and /b/ were significantly longer in whisper than in voiced speech, whereas durations for /m/ did not differ between the two modes (cf. also Osfar, 2011). Ito et al. (2005) found that 'voiced' consonants in the whispered speech mode had lower energy at low frequencies, up to 1.5 kHz, and their spectra were flatter than in voiced speech. Similar results have been obtained by Lim (2011, p. 70), who pointed out that greater spectral tilt in normal speech, i.e., stronger energy in the low frequency bands as opposed to high frequency bands, can be explained by the low-frequency energy provided by the glottal sound source. Finally, Heeren's (2015b) study investigated duration, intensity, COG and spectral balance, comparing the difference in intensity between the 0.5–2 and 2–8 kHz frequency regions, in spectra of Dutch /f/ and /s/. The parameters were excerpted from nonsense VCV sequences spoken at different pitch targets (low, mid, high) in whispered and voiced speech. The results revealed that in both speech modes durations were longer for lower pitch targets. The Centre of Gravity increased for higher intended pitches in voiced and whispered speech. Intensity decreased with lower pitch targets but the differences were only about 1 dB.

The literature on semi-whisper is quite sparse. A few instrumental studies have evaluated "stage whisper", which we suspect may be similar or comparable to the Polish semi-whisper, but they have concentrated on phonatory characteristics (e.g., Yan, Ahmad, Kenduk, & Bless, 2005); to our knowledge, no studies have evaluated vowels or consonants in semi-whispered speech. This speech mode, as our data shows, is characterized by irregular, i.e., interrupted F0 not only in the case of voiceless segments but also in voiced ones. Its amplitudes are lower than modal speech and higher than whispered speech (see also Section 2 for more informantion). It is worth emphasizing that in some languages, including Polish, the semi-whispered speech mode is captured by a specific word so that speakers immediately know that they should find a way to speak in a mode which lies between modal and whispered ones (see the Methods section for details). This additional non-modal speech mode allows us to broaden our investigation into the ways in which intonational differences may be realized.

1.4. Summary and research questions

Much recent research on voiced speech has highlighted the role that segments play in prosody, and conversely the way prosody affects segments. However, this interdependence has been investigated only to a limited degree thus far with respect to intonation. Very little is known about spectral and other acoustic changes of voiceless consonants under varying intonation patterns, and especially how any such intonational effects on consonants are realized in speech modes partially or totally lacking phonation. Only the work of Heeren (2015b), limited to Dutch, showed that acoustic parameters of whispered fricatives are contingent upon pitch targets. Even for voiced speech, interactions between consonants and intonation have only been investigated for German. Finally, to our knowledge, semi-whispered speech has not been investigated with regard to this interaction at all.

Accordingly, this study pursues three goals.

(1) First, it aims to provide new insights into the realisation of intended intonation in whispered, semi-whispered and modal speech modes in Polish, a Slavic language. Are spectral properties of both vowels and voiceless consonants contingent upon the intended intonation, i.e., questions vs. statements? We hypothesize, following the studies described in Sections 1.2–1.3, that vowels in whispered speech are generally characterized by higher formant frequencies and longer duration, and that formant frequencies are higher in questions than statements. Furthermore, we expect that consonants will be produced with higher Centre of Gravity in questions than in statements (Niebuhr, 2012). With respect to the semi-whispered speech mode, where F0 is partially present, we predict that vowel and consonant properties will tend to fall between those of normal and whispered speech. Since there are no studies on semi-whispered speech our hypothesis is rather intuitive.
(2) Second, we seek a better understanding of the potential role of voiceless segments in conveying intonation patterns across speech modes. We specifically investigate whether voiceless consonants, especially those not adjacent to vowels, vary with intonation patterns and whether question-statement differences in voiceless segments are found in all three speech modes. Is the lack of F0 compensated for in whispered speech and semi-whispered speech? How could acoustic cue-trading be organized when two contrasting patterns of intonation are produced in whispered and semi-whispered speech vs. modal speech? We predict that all three speech modes will show acoustic differences as a function of the intended intonation. In particular, we conjecture that due to the missing F0 in whispered speech, other acoustic cues will take over its function and contribute to expressing differences between statements and questions. We also hypothesize that spectral parameters of voiceless consonants (e.g., Centre of Gravity) will show more pronounced question-statement differences in whispered speech than in voiced speech, which suggests possible compensation for the lack of F0 in whisper. We also anticipate that some cues may distinguish questions vs. statements exclusively in whisper.
(3) Third, do longer voiceless sequences reflect intonational variations? Are there differences between segments within a cluster? *A priori*, we hypothesize that longer voiceless sequences are sensitive to different intonation patterns, similar to what has been found for short voiceless sequences. It might, however, be the case that intonational variations in sequence-final, sentence-final consonants are less pronounced than in preceding consonants.

To address these questions, we carry out detailed acoustic analyses that extend the methods used in previous work. Specifically, we obtain spectral moments and slopes using mulitaper spectra. This sophisticated technique is particularly well-suited for measurements of fricatives but has not been used before for investigating interactions of segments and prosody across speech modes.

The remainder of the paper is organized as follows: In Section 2 we present the methodology of our experiment and in Section 3 its results, focussing on acoustic correlates of intonation encoded in vowels (3.1) and consonants (3.2). A summary of the results is presented in 3.3. Section 4 is devoted to discussion of the results and conclusions.

2. Methods

In order to answer the questions in 1.4 we conducted an acoustic speech production experiment on Polish, a language which provides suitable test material due to its abundance of complex consonant clusters, including in coda position.

For testing our hypotheses we recorded eight different items ending in clusters consisting of a voiceless retroflex fricative followed by a retroflex affricate. Each item was presented in a pair: a polar question *(e.g., Widzi ten blu[ʂtʂ]?* 'Does he see the ivy?') was followed by a statement (*Widzi ten blu[ʂtʂ].* 'He sees the ivy.'). The production of the intonational difference was facilitated by the fact that questions were always followed by statements and the statements were naturally interpreted as answers to the questions. All target words were monosyllabic (see Appendix Table 1). The polar question was expected to be produced with a rising intonation and the statement with a low/falling intonation in accordance with results of previous studies of Polish intonation, e.g., Wagner (2008). Differences between the two types of intonation were independently confirmed by the measurement results provided below (see Section 3). In Polish the sentence stress falls on the sentence final content word, i.e., the last stressed syllable in a sentence (Rubach & Booij, 1985) and the boundary tone is high (H%) in polar questions and low in statements (L%) (Wagner, 2008). In both the statements and polar questions examined in the present study the sentence final intonation contour reflects a combination of pitch accent (sentence stress) and boundary tone.

In order to compare the realisation of intonation across modes we recorded the pairs described above in three different speech modes in a fixed order: normal, whispered and semi-whispered. Since changing the speech mode within an extremely short period of time could be a very challenging and error-prone task for our speakers, we preferred a fixed order with respect to the speech mode. This strategy in fact enabled the speakers to change the modes without any problems. It is also worth emphasizing that our speakers did not have difficulties understanding what we meant by *semi-whispered* speech since in Polish there is a word *pó□szeptem* 'half-whispered'/'semi-whispered' which means that generally one speaks quieter than normally but does not whisper. The fact that Polish, unlike many other languages, has a specific lexical item for this speech mode is another factor that makes it a useful language for investigating the questions at hand, in contrast to languages that have no such word.

Fig. 1 presents the waveforms, spectrograms and F0 curves (in semitones) of the polar question *Widzi ten blu[ʂtʂ]?* 'Does he see the ivy?' produced in normal, whispered and semi-whispered speech-modes. Fig. 2 presents the same polar question sentence (in normal speech mode) along with the corresponding statement (i.e., the answer to the question). As can be seen, the F0 contours for the polar question and statement conditions are typically rising and falling, respectively (see also Wagner, 2008). As noted above, the pitch accent falls on the final word and is conflated with the boundary tone. It can also be seen that in semi-whispered speech the amplitude is lower and F0 is reduced in comparison to the normal speech mode (ca. 100 ms of the vowel portion before the sentence-final consonants is devoiced). Other differences are reported in the Results section.

Sixteen native speakers of Polish (eight male), aged 20–52 years (mean 24.93, standard deviation 9.2), took part in the experiment. All speakers were monolingual, lived in Szczecin and spoke Standard Polish. They were asked to read a list of sentences in a non-randomized speech mode order starting with a normal speech mode, followed by whispered and semi-whispered speech modes so that they could easily pre-plan a given mode. The sentence list was read three times, but each time the items (but not the speech modes) were randomized. All recordings were conducted in a sound-proof room at the Electrical Engineering Department of the West Pomeranian University of Technology in Szczecin using a TLM103 microphone (20 cm distance from lips) connected to a ProTools system with a Digi 003 interface (sampling rate 44100 Hz). The items were analysed with PRAAT (version 5.3.57, Boersma & Weenink, 2014) and MATLAB (version R2007b, MathWorks, 2007). In total, measurements of 4608 sibilants were taken [8 items × 2 intonation types (rising, falling) × 3 speech modes (normal, semi-whispered, whispered) × 2 sibilant types (fricative, affricate) × 3 repetitions × 16 speakers]. Although the focus of the present paper is on the analysis of consonants we also examined vowels in order to gain a more complete insight into the realization of different intonation contours. Thus, we also measured 2304 vowel tokens. The difference in the number of vocalic and sibilant tokens arises from the fact that each lexical item contained one vowel but two sibilants, i.e., a fricative and an affricate. All measured vowels preceded the coda clusters in question.

We examined the following acoustic parameters of both vowels and consonants, as listed in (1). The parameters listed in (f), (g) and (h) were calculated at the midpoint of the sibilant noise for each fricative and affricate. As indicated above, these measures were chosen to follow on previous work and also to employ contemporary methods designed specifically to evaluate fricative acoustics.

(1) Parameters:
Vowels:

(a) Mean intensity over the complete duration of the vowel (Praat algorithm, standard values).
(b) Maximum and mean of F0 of the vowel, for normal voiced speech only.[2]

[2] Due to the complete absence of the F0 in whispered speech and the partial absence of the F0 in semi-whispered speech (where the F0 cannot always be reliably extracted) we did not analyse F0 in semi-whispered and whispered speech modes.

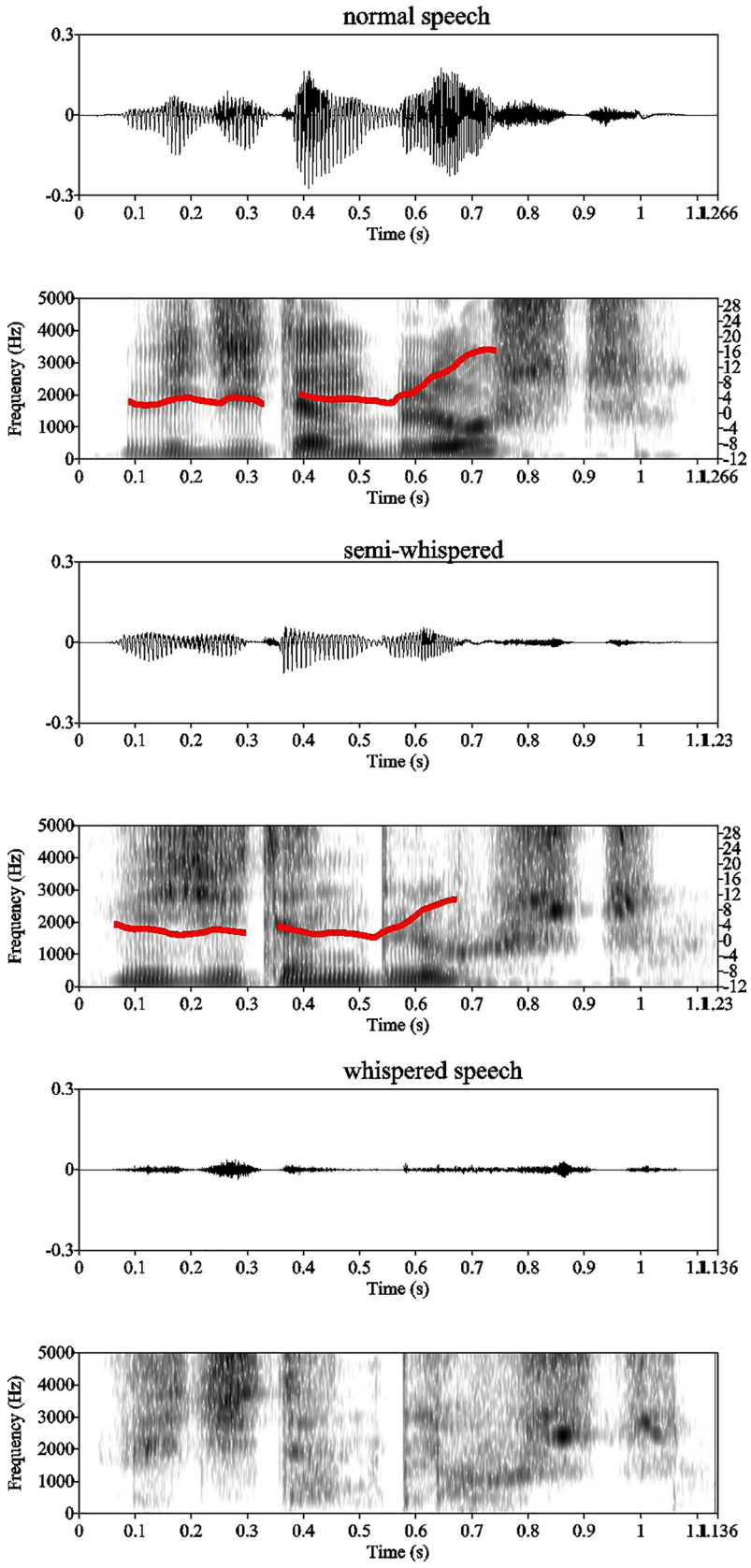

Fig. 1. Waveforms and spectrograms of the polar question *Widzi ten blu[ʂʈ͡ʂ]?* 'Does he see the ivy?' produced in normal (top panels), semi-whispered (mid panels) and whispered speech mode (bottom panels). The red line indicates the F0 tracing over time. The right vertical axis shows the F0 range in semitones.

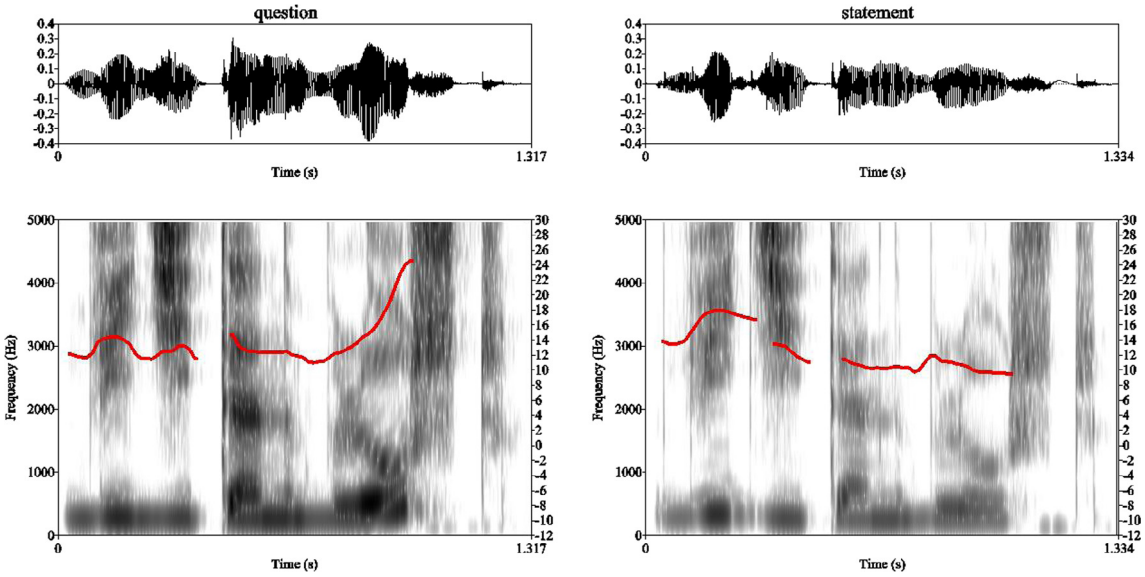

Fig. 2. Waveforms and spectrograms of the polar question *Widzi ten blu[ʂtʂ]?* 'Does he see the ivy?' and the statement *Widzi ten blu[ʂtʂ]* 'He sees the ivy' produced in normal speech. The red line indicates the F0 tracing over time. The right vertical axes show the F0 range in semitones.

(c) F0 difference between the vowel offset and onset.
(d) Formant frequencies F1, F2, F3 at the acoustic vowel midpoint, using the formant extraction algorithm of PRAAT[3], with the following parameter settings: maximum formant frequency: 5000 Hz; number of formants: 3; window length: 25 ms.
(e) Duration of the vowel.

Consonants (all calculated from the multitaper spectra using Matlab):

(f) Spectral Centre of Gravity (COG, first spectral moment), its standard deviation (STD, second spectral moment), skewness (third spectral moment), and kurtosis (fourth spectral moment).
(g) Spectral regression slopes (Jesus & Shadle, 2002; Lousada, Jesus, & Pape, 2012): m1 is the slope of the spectral regression line for the frequency range between 500 Hz and 3000 Hz, and m2 is the slope of the spectral regression line for the range between 3000 Hz and 11,000 Hz (see Fig. 3). The 3000 Hz value was chosen as the reference value because previous work showed this to be the approximate mean frequency of the highest-amplitude spectral peak for the retroflex place of articulation (see Stevens, 1998; Żygis et al., 2012).
(h) Frequency of the highest spectral peak of the frication noise in the range from 2 to 4 kHz.
(i) Mean intensity over the complete phoneme duration.
(j) Phoneme duration of the fricative and affricate.

We computed multitaper spectra with 12 ms windows for the frication noise at midpoint using a 512 point Hamming window. The power spectral density (PSD) was estimated via the Thomson multitaper method (linear combination with unity weights of individual spectral estimates and the default FFT length) available in the MathWorks Signal Processing Toolbox Version 6. All above mentioned acoustic parameters for the consonants have been found to be useful in the analysis of fricatives (Jesus & Shadle, 2002; Żygis et al., 2012). In addition, multitaper analysis provides for acoustic analysis of fricatives that is superior to that based on traditional spectral algorithms. Specifically, compared to standard spectral estimation techniques, multitaper analysis provides a very effective way to reduce the bias of the spectral estimates when calculated over short intervals of the data, and is thus highly suited to examining stochastic parts of the speech signal.

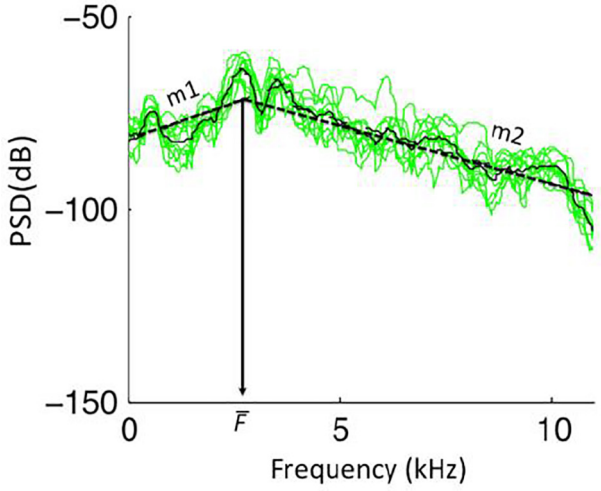

Fig. 3. Plot of 10 multitaper spectra (light colour) with the mean spectrum (black solid) and regression lines (dashed black) used to calculate the low-frequency slope (m1) and high-frequency slope (m2), with the end/starting point at the mean frequency $\overline{F}$, defined here as 3000 Hz (see text). The *y*-axis shows power spectral density.

Fig. 3 shows 10 multitaper spectra from one individual speaker, the overlaid mean spectrum and the computation of the regression lines m1 and m2, with the endpoint/startpoint $\overline{F}$.

All statistical analyses were conducted in the R environment software (version 3.0.2, R Development Core Team, 2010). Linear mixed effect models were employed for studying the influence of Speech Mode [modal, semi-whispered and whispered], Intonation [rise (questions) vs. fall (statements)],

[3] For all productions we manually checked that the formant algorithm neither missed formants nor included spurious peaks that did not represent formants. If necessary we manually corrected the formant values found by the algorithm.

Sex [male, female] and Repetition as well as their interaction on the variables listed in (1). To minimize the Type I error (Barr, Levy, Scheepers, & Tily, 2013) a maximized random structure was included as well: random intercepts for participants and items as well as their slopes for Speech Mode, Intonation, Repetition and their interactions were added to the initial model. Very high correlations found between random-effects terms were eliminated. (No high correlations between fixed effects were observed). The maximized models were tested against less complex models by means of likelihood ratio tests and the best fit model was taken as the final model. Finally, we also corrected for multiple comparisons by using the Tukey test.

All *p*-values reported in the paper are based on the Satterthwaite approximation available in the package 'lmerTest' (Kuznetsova, Brockhoff, & Christensen, 2015) which includes tests for linear mixed-effect models implemented in the 'lme4' package (Bates, Maechler, Bolker, & Walker, 2015). The results section includes mean values and *p*-values for significant findings; full LME output is provided for significant effects on all dependent measures in the Appendix tables.

3. Results

Our results indicate that several acoustic cues are used to differing extents when questions versus statements are produced in various speech modes. The presentation of the results follows the parameter listing displayed in (1).

3.1. Vowels

To ensure that participants produced expected differences between normal, semi-whispered and whispered speech we calculated the mean intensity of the vowels. The results point to the lowest mean intensity in whispered speech (47 dB) in comparison to semi-whispered (62 dB, $p < 0.001$) and modal speech (71 dB, $p < 0.001$; see Table 2 in the Appendix). The results also reveal that mean intensity is significantly higher in questions (Q) than in statements (S) in all three speech modes (whispered: Q 49 dB vs. S 44 dB, $p < 0.001$, semi-whispered: Q 64 dB vs. S 60 dB, $p < 0.001$) and modal speech mode (Q 74 dB vs. S 68 dB, $p < 0.001$). Furthermore, interactions between phonation and intonation type are highly significant ($p < 0.001$); see Fig. 4.

Note that the boxes in Fig. 4 and all remaining box plots correspond to the 25th to 75th percentile range, black lines in the boxes represent medians, and whiskers correspond to ±1.5 inter-quartile range; outliers, i.e., data above or below this range are represented as points in the graph.

To verify expected intonational patterns in questions and statements, we investigated the F0 maximum and the F0 mean of the vowel preceding the sibilant cluster in modal speech (cf. (1b)). The results, presented in semitones,[4] indeed show that average values for both F0 maximum and F0 mean are significantly higher in questions than in statements (F0 max: Q 18.66 vs. S 7.42, $p < 0.001$, F0 mean: Q 14.29 vs. S 5.78, $p < 0.001$, see Appendix Tables 3,4 and Fig. 5). As expected, average values for F0 maximum and F0 mean are lower for male than female speakers (F0 max: male 8.39 vs. female 17.10, $p < 0.001$; F0 mean: male 5.21 vs. female 14.09, $p < 0.001$; see Tables 3,4). Recall that due to the complete absence of the F0 in whispered speech and the partial absence of the F0 in semi-whispered speech we did not analyse this parameter in those two modes since we could not estimate it reliably.

We also calculated the F0 difference between the vowel offset and onset (cf. (1c)). This measure likewise points to an average F0 increase in questions (female 8.23 vs. male 7.86, $p < 0.001$) and average falling F0 in statements (female −1.61 vs. male −2.73, $p < 0.001$) confirming our initial assumptions about F0 differences in Polish intonation patterns; cf. also Wagner (2008).

The absence of F0 in whispered speech leads us to a key question for the present study, namely: How can an intonational distinction between questions and statements be produced in whispered speech if the F0, the most important correlate of intonation, is not available to play a distinguishing role?

First, our results point to the importance of formants. Since the vowel preceding the clusters was not always the same due to lexical restrictions (cf. Appendix) we will present results on the two most distinctive and most frequent vowels in our data base, viz., the low vowel /a/ and the high back vowel /u/.

The data show that the mean F1 is significantly higher in whispered speech than in semi-whispered and modal speech (whispered 954 Hz vs. semi-whispered 478 Hz, $p < 0.001$, modal 595 Hz, $p < 0.001$; see Appendix Table 5). Questions are produced with higher mean F1s than statements in whispered speech (Q 992 Hz vs. S 914 Hz, $p < 0.001$) and modal speech (Q 628 Hz vs. S 560 Hz, $p < 0.001$). In contrast, questions in semi-whispered speech have lower average F1s than statements (Q 466 Hz vs. S 491 Hz, $p < 0.05$). As expected, female speakers produce higher mean F1s than male speakers (female 735 Hz vs. male 619 Hz, $p < 0.001$). Also as expected, the average F1 of /u/ is significantly lower than that of /a/ (/u/ 829 Hz vs. /a/ 1071 Hz, $p < 0.05$). (Note that the values of F1 for /u/ are extremely high in whispered speech as compared to semi-whispered and modal speech.) Lastly, the interaction of phonation and intonation type is highly significant ($p < 0.001$). Fig. 6 presents the results.

Similar to F1, the mean second formant frequency of vowels is significantly higher in whispered speech as compared to semi-whispered and modal speech (whispered 1436 Hz vs. semi-whispered 1165 Hz, $p < 0.001$, modal 1155 Hz, $p < 0.001$; see Appendix Table 6). Questions are produced with higher average F2s than statements in whispered speech (Q 1464 Hz vs. S 1407 Hz, $p < 0.001$), and with lower average F2s in semi-whispered speech (Q 1135 Hz vs. S 1195 Hz, $p < 0.01$). No F2 differences are found in modal speech. Female speakers have higher mean F2s than male speakers (female 1358 Hz vs. male 1185 Hz, $p < 0.001$). Finally, the interaction of phonation and intonation type is highly significant ($p < 0.001$). Fig. 7 presents the results.

The third formant frequency is higher on average in whispered than in modal speech (whispered 2693 vs. modal 2587, $p < 0.01$; see Appendix Table 7) but does not differ from semi-whispered speech. Questions are produced with higher mean F3 than statements in whisper (Q 2748 Hz vs. S 2636 Hz, $p < 0.001$) and modal speech (Q 2623 Hz vs. S

[4] The calculations were based on the reference value of 100 Hz, which is one of the standards in calculating semitones from Hz frequency values.

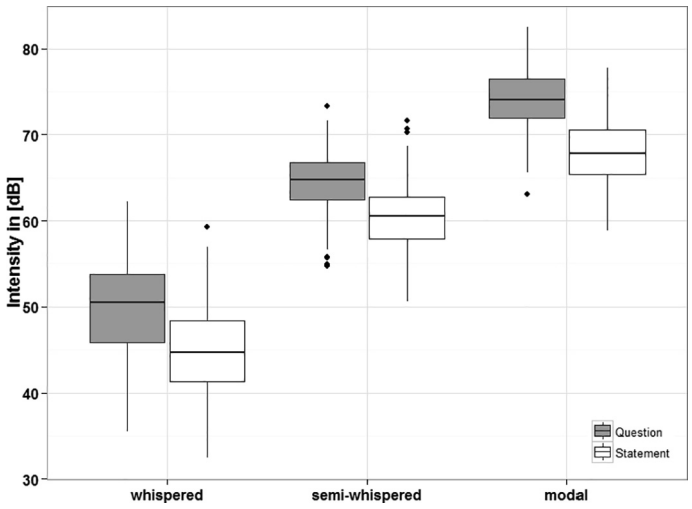

Fig. 4. Mean intensity values for questions and statements in vowels across whispered, semi-whispered and modal speech mode.

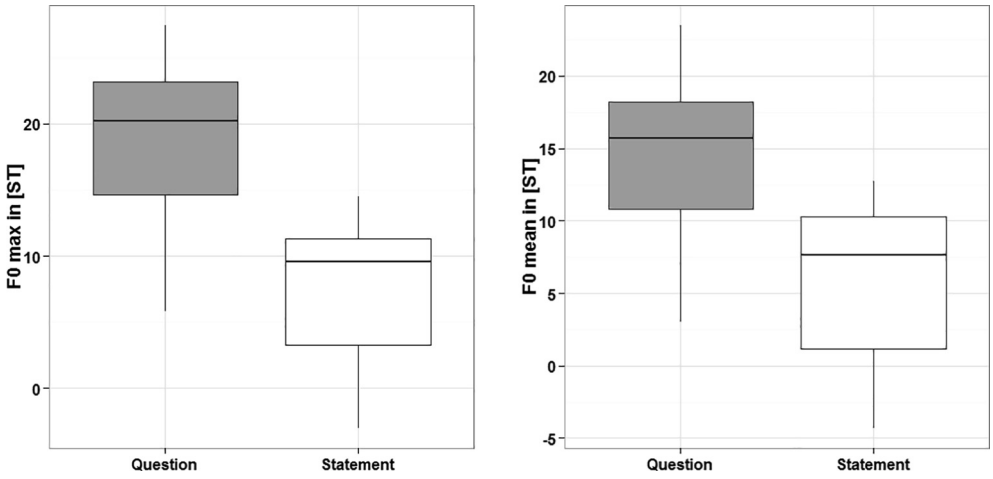

Fig. 5. Boxplots for F0 maximum and mean of the vowels in questions and statements in modal speech.

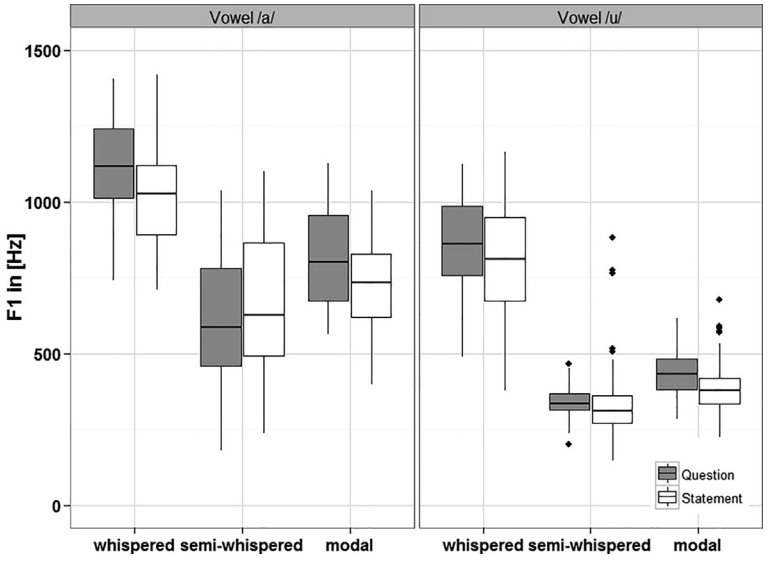

Fig. 6. Boxplots for F1 frequency values at the vowel midpoint of /a/ (left) and /u/ (right) for questions and statements across whispered, semi-whispered and modal speech modes.

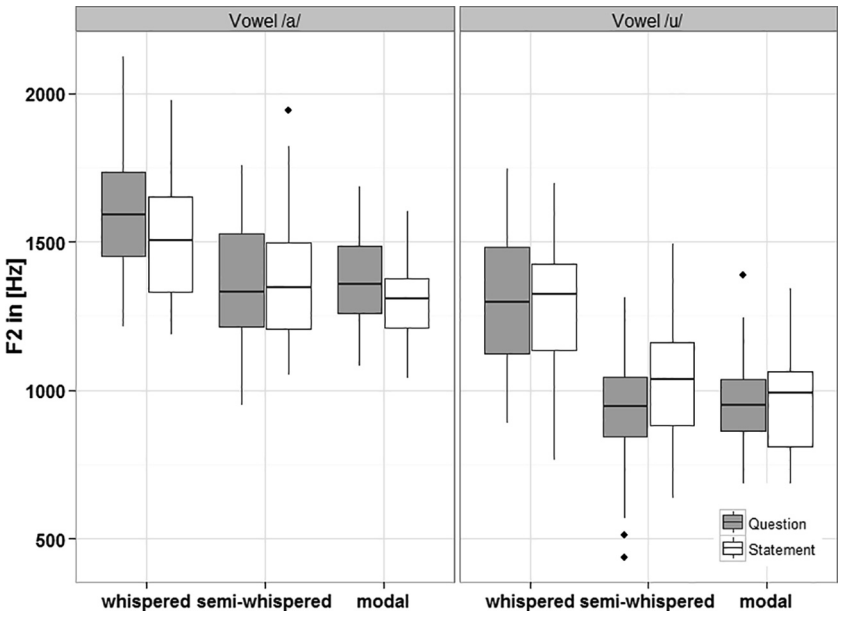

Fig. 7. F2 frequency values at the vowel midpoint of /a/ (left) and /u/ (right) for questions and statements across whispered, semi-whispered and modal speech modes.

2552 Hz, $p < 0.01$). No differences between male and female speakers are observed. Fig. 8 illustrates the results.

Regarding vowel duration (cf. (1d)), whispered vowels are generally longer than vowels produced in semi-whispered speech (mean log whispered −0.83 vs. mean log semi-whispered −0.89, $p < 0.001$; see Table 8) and modal speech mode (mean log −0.89, $p < 0.001$). No significant differences in vowel duration are found between statements and questions in any speech mode.

3.2. Fricatives and affricates

3.2.1. Spectral moments

First we will provide results for spectral moments, the parameters that have been used most frequently for sibilant description in past work. Next we will show the measures that, in our view, provide a clearer and more accurate representation of frication noise characteristics, i.e., spectral slopes and locations of specific spectral peaks (cf. Fig. 3).

The first spectral moment, i.e., COG, is significantly lower on average in whispered speech than in semi-whispered speech (whispered 3339 Hz vs. semi-whisper 3466 Hz, $p < 0.001$; see Table 9) and normal speech (4540 Hz, $p < 0.001$). The mean COG values are significantly higher for questions than for statements across all speech modes (whispered: Q 3579 Hz vs. S 3088 Hz, $p < 0.001$; semi-whispered: Q 3582 Hz vs. S 3354 Hz, $p < 0.001$; modal: Q 4823 Hz vs. S 4256 Hz, $p < 0.001$). Fricatives exhibit higher mean COG values than affricates (fricatives 4127 Hz vs. affricates 3434 Hz, $p < 0.001$). Furthermore, repetition has a significant effect on average COG values, which were higher in later repetitions ($p < 0.001$).[5] Finally, it should be noted that the interaction between phonation and intonation type was highly significant ($p < 0.001$) with respect to COG and all other parameters presented below apart from duration. The Appendix tables show the statistical details. The results for COG are illustrated in Fig. 9.

For the second spectral moment, whispered speech displays a significantly higher mean standard deviation (STD) than semi-whispered (whispered 1688 vs. semi-whispered 1796, $p < .01$) and modal speech (1995, $p < .001$; see Table 10). Only in whispered and semi-whispered speech does average STD differ in the production of questions and statements, being higher in questions (whispered: Q 1758 vs. S 1615, $p < .001$; semi-whispered: Q 1830 vs. S 1763, $p < .05$). Fricatives display higher STD than affricates (fricatives 1990 vs. affricates 1753, $p < .001$).

With respect to skewness, the third spectral moment, our results indicate significant differences between whispered and other speech modes. Average skewness is significantly higher in whispered speech compared to semi-whispered (whispered 1.82 vs. semi-whispered 1.66, $p < 0.001$; see Table 11) and modal speech (0.89, $p < 0.001$). The data show lower mean skewness values for questions than statements in whispered speech only (Q 1.56 vs. S 2.11, $p < 0.001$). Lastly, frication in fricatives displays lower mean skewness values than in affricates (fricatives 1.23 vs. affricates 1.67, $p < 0.001$). Fig. 10 presents the results.

The fourth spectral moment, kurtosis, has significantly higher average values in whispered than in semi-whispered speech (whispered 19.59 vs. semi-whispered 11.05, $p < 0.001$; see Table 12) and normal speech mode (3.23, $p < 0.001$). As for comparing questions to statements, average kurtosis differs in whispered and semi-whispered speech: it is lower in questions than in statements (whispered Q 10.76 vs. S 18.80, $p < 0.001$; semi-whispered: Q 8.92 vs. S. 11.35). Lastly, mean kurtosis is significantly lower in fricatives than affricates (fricatives 4.84 vs. affricates 8.55, $p < 0.001$). The results are illustrated in Fig. 11.

[5] The repetition effect for COG does not yield to simple explanation, and no other measure showed a significant effect of this factor.

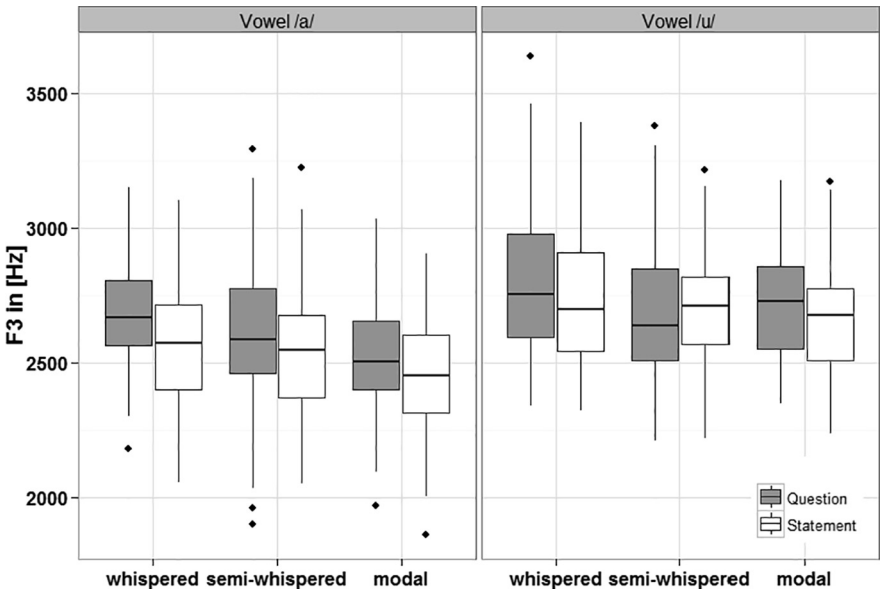

Fig. 8. F3 frequency values at the vowel midpoint of /a/ (left) and /u/ (right) for questions and statements across whispered, semi-whispered and modalspeech modes.

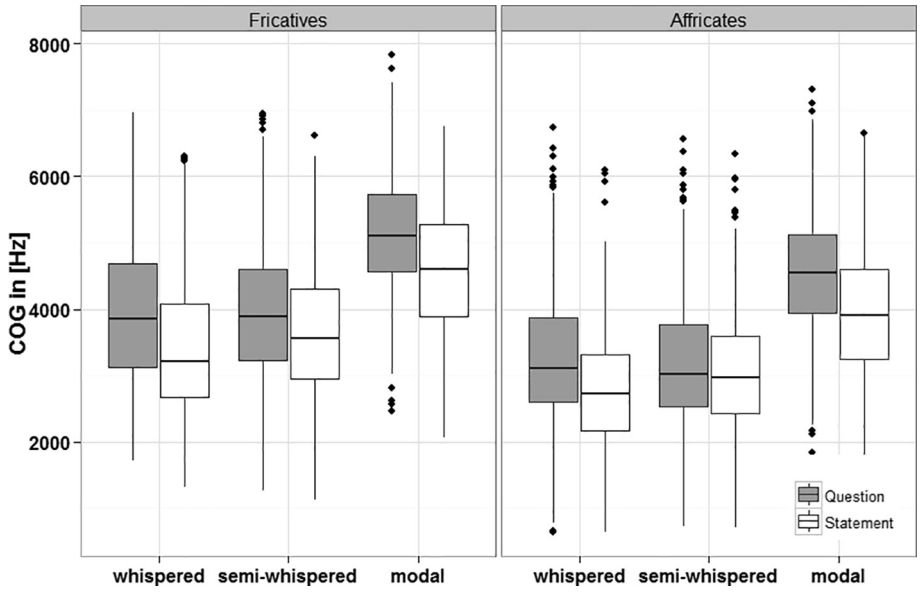

Fig. 9. Boxplots for Centre of Gravity at the midpoint of the frication of fricatives (left) and affricates (right).

3.2.2. Spectral tilt and peak frequencies

Fig. 12 shows spectra, averaged over items and speakers, for the fricatives and affricates in the three speech modes and two intonation contexts. (The double-peaked nature of the spectra is discussed below, as are amplitude differences). The figure also shows the m1 and m2 slopes (cf. Fig. 3 above). The m1 slope measure reflects the balance of low-frequency energy in the spectrum relative to the lowest-frequency spectral peak which in these data is generally around 2–3 kHz. (Recall that this slope measure was taken over the frequency range 500–3000 Hz, so that very low-frequency energy is excluded.) The statistical results show that mean m1 turns out to be significantly lower, i.e., less steep, in whispered than in modal speech mode (whispered 4.82 vs. modal 5.71, $p < 0.001$; see Table 13) and semi-whispered speech mode (4.97, $p < 0.001$). This may reflect the stronger balance of frequencies below 1000 Hz, i.e., below the spectral minimum, for whisper (see spectra). Only in whispered speech is a significant difference observed between questions and statements; namely whispered fricatives display higher average m1 values in questions than in statements (Q 5.35 vs. S 4.28, $p < 0.001$). A steeper slope in whispered questions suggests that this combination of intonation and speaking mode leads to increased excitation of the main spectral peak (compare the relative amplitudes of questions and statements across modes in Fig. 12). In addition, fricatives are produced with a higher mean m1 than affricates (fricatives 5.91 vs. affricates 4.43, $p < 0.05$).

The spectral slope m2 at the midpoint of the sibilant is, on average, less steep in whispered than in modal speech (whis-

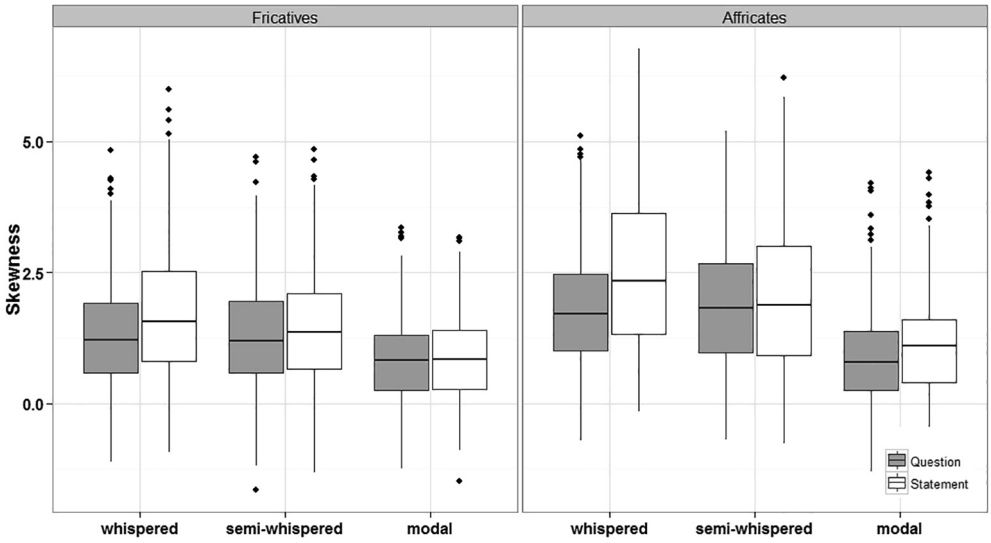

Fig. 10. Boxplots for skewness of fricatives (left) and affricates (right) measured at the midpoint of frication.

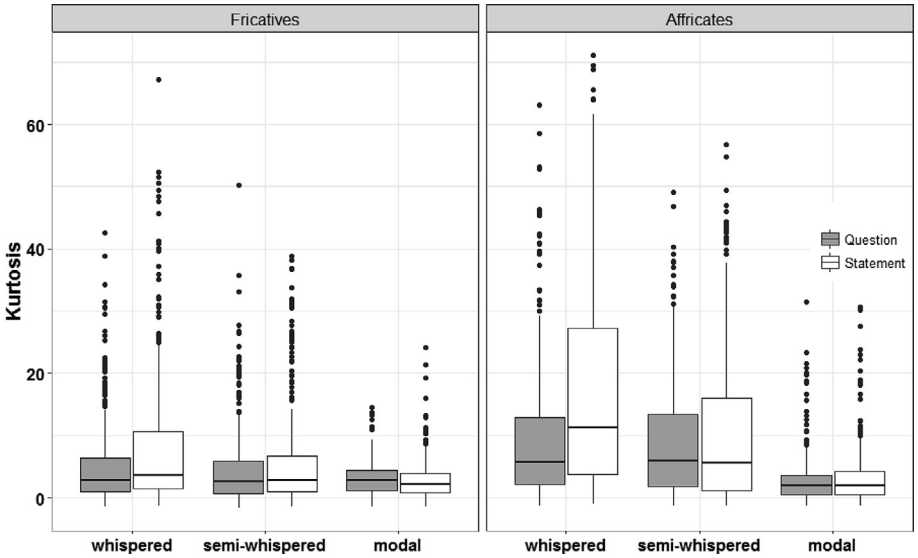

Fig. 11. Boxplots for kurtosis of fricatives (left) and affricates (right) measured at the midpoint of frication.

pered −2.18 vs. modal −2.58, $p < 0.001$; see Table 14) and in semi-whispered speech (−2.25, $p < 0.001$). Questions are produced with a steeper m2 in comparison to statements in whispered (Q −2.32 vs. S −2.02, $p < 0.001$), semi-whispered (Q −2.3 vs. S −2.2, $p < 0.001$) and modal speech (Q −2.62 vs. S −2.54, $p < 0.01$). In Fig. 12 this is evident as the greater amplitude difference between the question and statement for the second spectral peak compared to the first, consistent across all plots. Fricatives have steeper m2 values than affricates (fricatives −2.46 vs. affricates −2.21, $p < 0.001$).

Since we found that nearly all retroflex spectra had two major peaks (see Fig. 12), we refined the calculation of the highest amplitude spectral peak to the range from 2 kHz to 4 kHz, i.e., to the frequency region where according to previous studies the major broad (and only) spectral peak is expected for this retroflex place of articulation.[6] In this selected frequency range, the highest peak has a lower average value in whispered speech than in modal speech (whispered 2651 Hz vs. modal 2932 Hz, $p < 0.001$; see Table 15). Questions are produced with a higher mean peak than statements in whispered (Q 2749 Hz vs. S 2549 Hz, $p < 0.001$), semi-whispered (Q 2690 Hz, vs. S 2586 Hz, $p < 0.001$) and modal speech (Q 3026 Hz vs. S 2836 Hz, $p < 0.001$). There is no difference between fricatives and affricates with respect to the frequency of the highest spectral peak. The results are shown in Fig. 13.

[6] As stated in Section 2, for this place of articulation the literature describes one broad spectral peak around 3 kHz. We are not sure at the moment what phenomenon causes the additional spectral peak. For this reason we limited the frequency range for the computation of the major broad peak which is associated with the resonance of the front cavity and thus codes how anteriorly the fricatives are articulated.

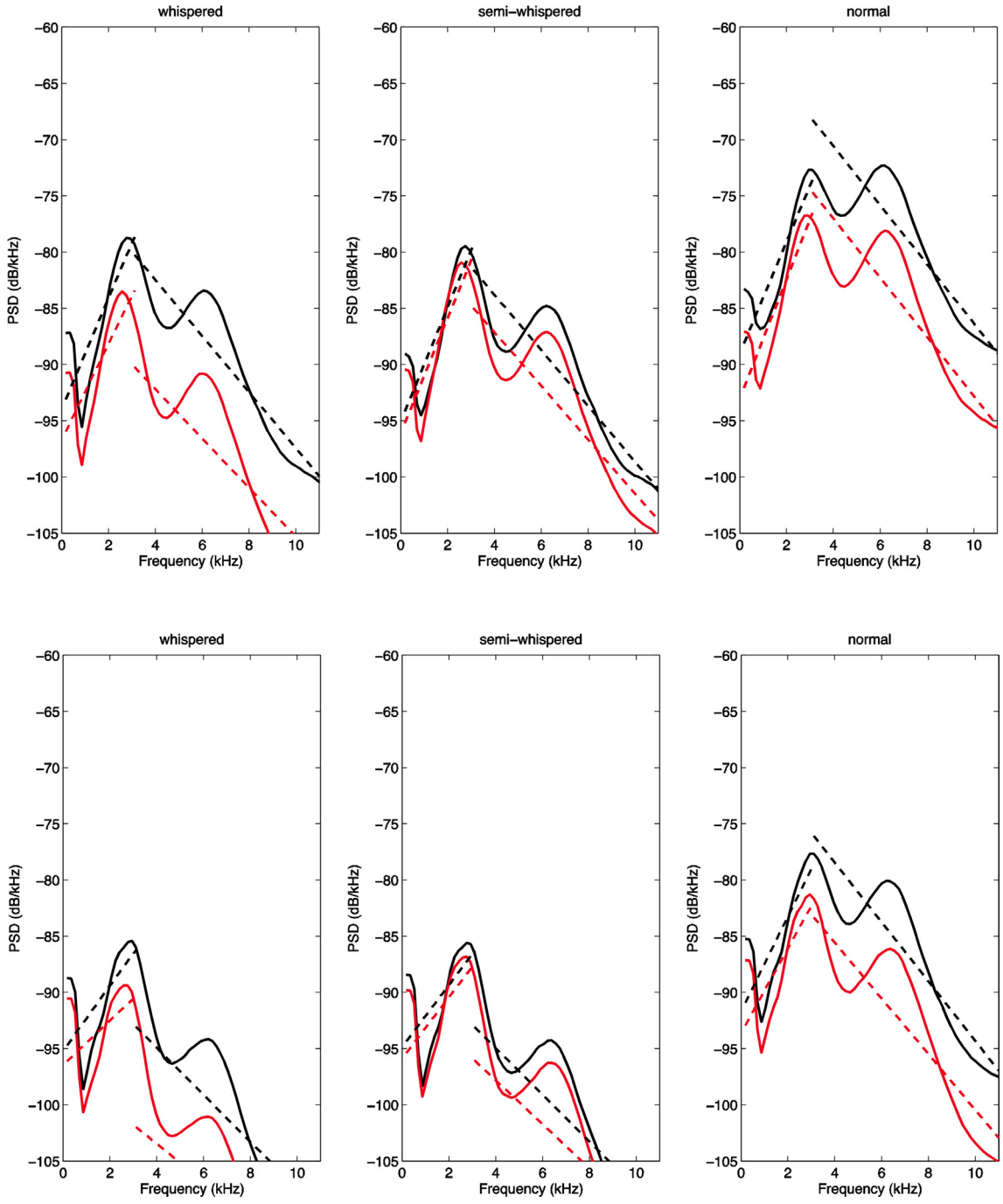

Fig. 12. Multitaper spectra (mean plots over all items and speakers) for the frication midpoint of fricatives (top) and affricates (bottom) in whispered, semi-whispered and modal speech modes. Black solid lines correspond to the question and lighter colour to the statement condition. Dashed lines are the spectral regression lines m1 and m2. The y-axes show power spectral density (PSD).

As with vowel intensity (Section 3.1), the fricative intensity measure also clearly indicates differences across the three speech modes: average values are lower for whispered than semi-whispered (whispered 49.84 dB vs. semi-whispered 50.10 dB, $p < 0.001$; see Table 16 and Fig. 14) and modal speech modes (57.30 dB, $p < 0.001$, see Fig. 14). Furthermore, values are higher on average in questions than in statements across all three speech modes (whispered speech: Q 51.90 dB vs. S 47.70 dB, $p < 0.001$; semi-whispered: Q 50.94 dB vs. S 49.43 dB, $p < 0.001$; modal: Q 59.76 dB vs. S 54.95 dB, $p < 0.001$). Fricatives display a higher mean intensity than affricates (54.60 dB vs. 50.20 dB, $p < 0.001$). The magnitude of the intensity difference between whisper and semi-whisper is much smaller for the fricative noise than that observed for the vowels (cf. Fig. 4 above).

Lastly, the mean log transformed sibilant duration is longer in whispered speech than in semi-whispered speech (whispered −0.84 vs. semi-whispered −0.85, $p < 0.001$; see Table 17) and modal speech (−0.86, $p < 0.001$). However, the sibilants show similar durations in questions and state-

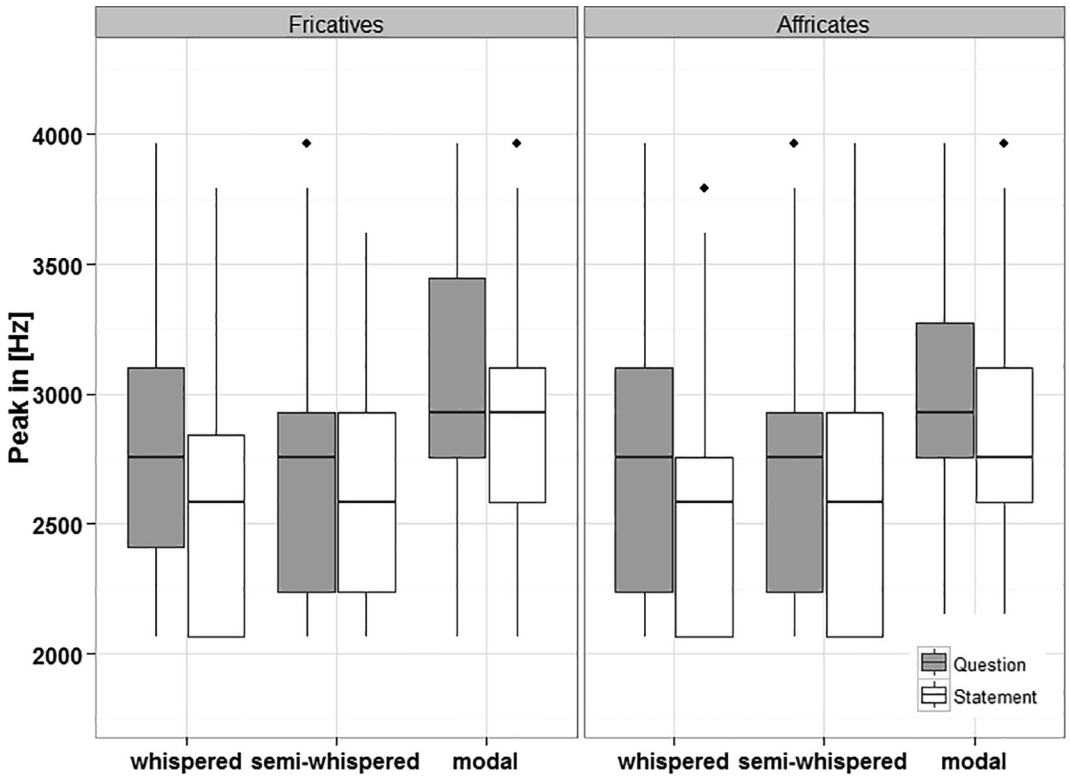

Fig. 13. Boxplots for the highest spectral peak from 2 kHz to 4 kHz of fricatives (left) and affricates (right) measured at the midpoint of frication.

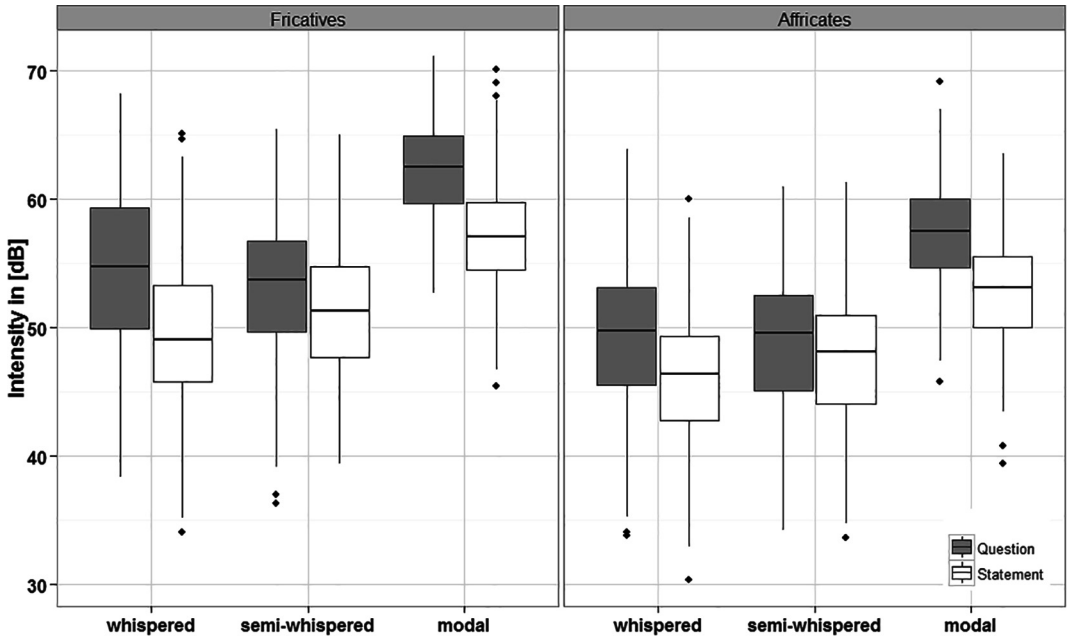

Fig. 14. Boxplots for the intensity of fricatives (left) and affricates (right) measured at the midpoint of frication.

ments across all three speech modes. As expected, fricatives have shorter mean durations than affricates (fricatives −0.89 vs. affricates −0.82, $p < 0.001$), see Fig. 15.

3.3. Summary of the results

The results are generally in line with past studies of acoustic differences between whisper and voiced speech. For example, we observed, as many past authors, that intensities are lower, durations are longer, and vowel formants are higher in whisper than in modal speech (see Section 1.3 above). With the exception of F3, whisper vs semi-whisper comparisons showed the same general patterns as whisper-modal differences. Of most interest here, the data also indicate that, in Polish, intonational differences between

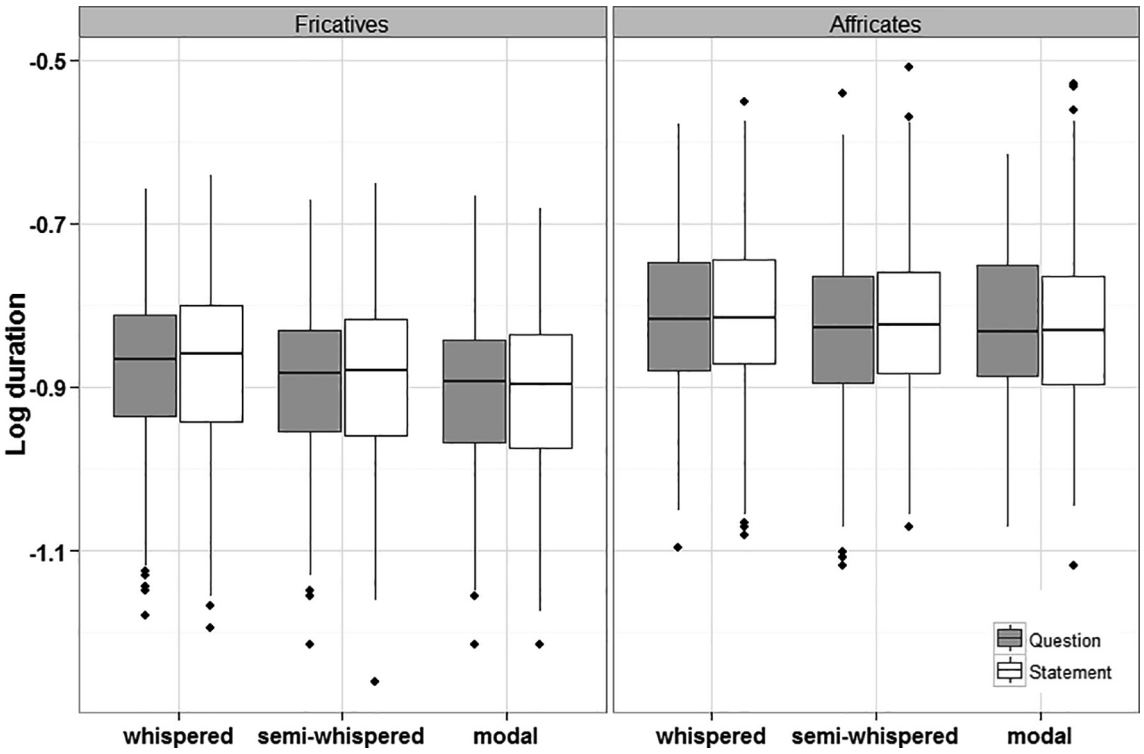

Fig. 15. Boxplots for duration of fricatives (left) and affricates (right).

questions and statements are reflected in the acoustic characteristics of both vowels and consonants.

With respect to vowels, the results are largely consistent with outcomes of previous intonational research. The intensity at the acoustic midpoint of the vowel is higher in questions as opposed to statements (Heeren & van Heuven, 2009, 2011). No question-statement difference in vowel duration was found in any speech mode, comparable with the findings obtained by Heeren and van Heuven (2009) for whispered Dutch. Questions were produced with a higher F1 and F3 than statements in whispered and modal speech modes. Questions had higher F2s than statements in whispered speech only. The results are also in accordance with Heeren and van Heuven (2009) who found that /ə/ in whispered Dutch had higher F1 and F2 frequencies in questions than in statements. The results for F1 and F2 could suggest that when producing questions the articulatory vowel settings are more open and more fronted than when producing statements, especially in whisper. The reversed direction of the formant differences between questions and statements in semi-whisper is difficult to interpret given the virtual absence of previous acoustic work on this mode; the finding does, however, contribute to the general conclusion that intonational patterns may be realized differently across speaking modes, and suggests that further work on semi-whispered speech is warranted.

For consonants the results clearly indicate that:

(1) Spectral properties of consonants are contingent upon the intended intonation.
(2) While the differences between questions and answers are found in all three speech modes, some are more pronounced in whispered speech, and some appear only in whispered speech.
(3) Even the second phoneme in a consonant cluster is sensitive to intonational changes, which is a previously unreported result.

Several parameters displayed differences between questions and statements and seem to be robust as they are found in all three speech modes. A higher-frequency spectral peak and a higher COG were found in questions as opposed to statements. This latter finding is in line with results presented for German (Niebuhr, 2012; Ritter & Roettger, 2014) where higher COG values for voiceless fricatives were reported for questions in German modal speech and in accordance with the study by Heeren (2015b) where higher COG characterized /s/ and /f/ in higher pitch targets in whispered speech. Questions were produced with a lower spectral regression line slope m2 (from 3 kHz to 11 kHz) in comparison to statements indicating that the spectra fell more steeply above 3 kHz for questions than for statements. In addition, questions were distinguished from answers by higher intensity, although the difference in consonantal intensity was smaller in semi-whisper than in the other modes.

Apart from these robust differences, some parameters differentiate questions from answers exclusively in whispered speech. These include the spectral slope m1 (from 500 Hz to 3 kHz), being higher in questions than in statements indicating that questions are produced with steeper low-frequency slope. Furthermore, the two spectral moments skewness and kurtosis were found to be lower in whispered questions than in whispered statements. It is worth noting that the significantly higher skewness in whispered speech indicates that the mass of the spectral distribution is towards lower frequency values as compared to the other speech modes. However, for whispered questions, the mass of the

spectral distribution moves towards higher frequencies as compared to whispered statements. In a similar vein, the higher kurtosis (peakedness; width of peak) in whispered speech indicates a narrower spectral peak for these sibilants in comparison to their semi-whispered and normal speech counterparts. The peak is, however, broader in whispered and semi-whispered questions than in whispered and semi-whispered statements.

Finally, all segments in these long voiceless sequences, i.e., in the clusters, undergo changes in accordance with the pitch target. Our results indicate sensitivity to intonational differences not only in the fricatives directly adjacent to vowels but also the frication portions of affricates following voiceless fricatives and in absolute utterance-final position. Although the spectral properties of both fricatives and affricates are influenced by intended intonation, there are also differences between the segments independent of the pitch target: fricatives are produced with a higher m1 and lower m2 than affricates. Fricatives are also characterized by higher intensities, COG and STD than affricates, but lower skewness and kurtosis. The fact that the spectra of fricatives are different from those of affricates may be due to (i) the co-articulatory effects, i.e., the influence of a preceding vowel on fricatives, (ii) their position, i.e., the very final sentence position for affricates influenced by F0 declination and/or (iii) inherent differences between fricatives and affricates.

4. Discussion and conclusions

The results indicate that varying intonation patterns affect not only vowels, as reported in much previous research, but consonants as well. Their spectral properties differ in line with the intended intonation of the speaker. Moreover, these spectral differences are more pronounced in whispered speech than in semi-whispered and normal speech. These results contribute to our understanding of different speaking modes and the encoding of intonational differences, and have several linguistic and non-linguistic applications.

First of all, they provide further demonstration of an interdependent relationship between segments and intonation. Segmental properties are contingent upon intonation, independently of whether F0, the main correlate of intonation, is present or not and conversely segments, including voiceless sounds, show intonational differences. This conclusion has at least three important consequences for linguistically-oriented research. First, in general terms, intonation should not be viewed as a concept exclusively related to F0 but rather as an output of the interaction of multiple cues, including spectral properties of consonants. This conclusion supports Ladd's (1996:6/2008:4) view according to which *intonation* refers to "suprasegmental phonetic features" including fundamental frequency, intensity and duration. Our study shows that intonation is also encoded in other measures such as spectral moments and such correlates can be found in consonants. This conclusion is also in line with recent studies on intonation, see e.g., recent studies in Żygis and Malisz (2016) where it has been shown that intonation does not only refer to F0 but it is an output of an interplay of several acoustic parameters. Second, voiceless segments, which have been generally avoided in research on intonation due to the F0 absence, should be considered as units which do encode information on intonation. Third, the results call into question the typical practice of disregarding intonational conditions in studies of segmental properties. The data indicate that intonation patterns significantly affect both vocalic and consonantal properties. This conclusion applies to all speech modes: whispered, semi-whispered and normal.

Some of the spectral differences observed in consonants across conditions presumably arise from basic physiological processes. We focus here on the averaged spectra and the spectral slope measures (the double-peaked nature of these fricative spectra calls into question the use of the spectral moments analysis). For example, it is evident in the averaged spectra (Fig. 12) that the affricates have lower amplitudes than the fricatives. This may reflect the effects of decreasing subglottal pressure at utterance end. Lower driving pressures should yield reduced excitation of spectral poles, providing an explanation for flatter (less positive) m1 slope values in the affricates. Jesus and Shadle (2002) predicted that steeper m1 slopes should result from more posterior places of articulation, more localized sources, and higher source strengths. The higher m1 results in whispered questions compared to statements could therefore arise from any of those conditions (we note that a more posterior place of articulation would conflict with the vowel results, where higher F2s were observed). Conversely, lower m1 values in whisper compared to semi-whisper and modal speech could correspond to more anterior articulation, a more diffuse source (suggesting a difference in lingual configuration), and a lower source strength in this speech mode. Given that numerous articulatory characteristics can yield the same acoustic effect, articulatory data will be needed to distinguish among these possibilities.

Whether the changes across speech modes are caused by internal or external factors, the end result is the same; the production system maintains its functions, i.e., to produce questions and statements, because the system of spoken language is robust (see e.g., Kingston & Diehl, 1995; Winter, 2014). Evidence for this property has been established, for example, by the wealth of experimental evidence derived from perturbation studies (see e.g., Brunner, Hoole, & Perrier, 2011; Weismer & Bunton, 1999) in which speakers, despite articulatory perturbations, found a way to achieve the desired output. The robustness of spoken language is possible due to a considerable repertoire of different and partially redundant cues which can enter into trading relations. Conversely, one cue can also serve different functions. For example, lower-frequency spectral peaks characterize whispered as compared to modal speech, and also differentiate questions vs. statements in whispered speech (see a discussion on various principles of robustness shared by speech and biological systems in Winter, 2014).

The study also reports, to our knowledge for the first time, results on vowel and consonant characteristics in semi-whispered speech. The vowel intensity data (along with auditory impression) indicate that speakers produced this mode as expected in regard to the overall amplitude characteristics, in between the extremes of voiced and whispered speech. In other measures, however—most notably the formants—values for semi-whisper were not intermediate between those of the

other two modes. Conceivably the lower F1s observed in semi-whisper compared to the other conditions reflect a strategy for producing softer speech by limiting articulatory opening. In contrast, the reversal of the formant differences between questions and statements, with lower values in questions than statements for F1 and F2 in semi-whisper, does not lend itself to simple interpretation. It is clear that more work is required to understand the nature of semi-whispered speech, but on the basis of the present data it appears that semi-whispered speech is not strictly identifiable in terms of presence vs. absence of particular cues but rather is characterized by an interplay of various cues, which may be subject to extensive inter-speaker variation. The distinct patterns of question-statement differences in semi-whispered speech observed here do suggest that this mode may provide useful material for testing interactions between speaking mode and intonation patterns, as well as cue weighting patterns in perception.

We have presented a range of evidence indicating that various spectral cues encode prosodic differences, viz., rising intonation in questions and falling intonation in statements. The fact that some spectral differences between questions and statements are found exclusively, or to the greatest extent, in whispered speech emphasises the potential relevance of these cues for this particular speech mode and also suggests that they compensate for the lack of F0 which plays the most distinctive intonational role in the phonated speech mode. Such relations, in which one cue compensates for the absence or reduced occurrence of another cue, are widely known in the literature as trading relations (e.g., Diehl, 2011; Parker, Diehl, & Kluender, 1986). It is often the case that redundant cues enter into trading relationships wherein the magnitude of some cues is increased whereas others are decreased (see Pape & Jesus, 2014, 2015 for a discussion of the cue-trading aspect of the production-perception link). In the case at hand it is evident that some spectral cues are more pronounced in whispered speech than semi-whispered or voiced speech; the same cues may be less important or even redundant in voiced speech. The specific cues that are most important for robust perception of prosodic contrasts, and the degree to which such cues and cue-trading relationships are listener-dependent, remain to be determined by means of perception experiments.

The study additionally may have implications for our understanding of speech adaptation processes. Previous studies have demonstrated that language users are able to quickly alter their output with the onset of environmental noise (e.g., Grynpas, Baker, & Hazan, 2011; see Winter & Christiansen, 2012 for an extensive discussion). The current results suggest that the feedback mechanism which detects changes in one's perceived output and quickly adapts the speech production mechanism in order to achieve more intelligible speech is found in whispered speech as well. Our speakers varied the acoustic characteristics in accordance with which of the three speech modes they used. Thus, they did not adapt to external circumstances but to changes in their own speech patterns. In this context it would be interesting to see whether and how speech may differ in conditions where speakers talk to others, e.g., in dialogues (see a discussion of speech accommodation in terms of convergence and divergence in Winter & Christiansen, 2012).

Finally, the results of the present study might be useful for speech synthesis where F0 reconstruction in whispered speech is a highly topical issue (e.g., McLoughlin, Sharifzadeh, Tan, Li, & Song, 2015; Sharifzadeh & McLoughlin, 2011; Toda, Nagiri, & Shikano, 2012). Several attempts to reconstruct F0 appear to have been insufficient because the algorithms were either based on small units (see e.g., MELP by Morris, 2003 or CELP by Sharifzadeh, 2010) or they were trying to derive F0 from vowel formants (McLoughlin et al., 2015). In our view, fine-grained spectral consonantal properties contingent upon intonational patterns, obtained by means of sensitive techniques, may provide additional parameters for reconstructing the intended intonation of a given sentence. Additional analyses of semi-whispered speech, in which F0 is partially present, could provide further insight into the variety of ways that prosodic variation might be synthesized.

Acknowledgements

This research has been supported by Bundesministerium für Bildung und Forschung (BMBF, Germany) Grant Nr. 01UG1411 to Marzena Żygis. This study has also been financed by FCT (Portugal), through IEETA (Portugal), within project UID/CEC/00127/2013 and the post-doctoral fellowship from FCT (Portugal) SFRH/BPD/48002/2008 to Daniel Pape.

Appendix A.

Table 1
Words used in the experiment.

Orthography	IPA	Gloss
bluszcz	[blu$\underline{ʂ}$t$\underline{ʂ}$]	"ivy"
deszcz	[dɛ$\underline{ʂ}$t$\underline{ʂ}$]	"rain"
dreszcz	[drɛ$\underline{ʂ}$t$\underline{ʂ}$]	"shiver"
gąszcz	[gɔ̃$\underline{ʂ}$t$\underline{ʂ}$]	"thicket"
haszcz	[ha$\underline{ʂ}$t$\underline{ʂ}$]	"bush"
moszcz	[mɔ$\underline{ʂ}$t$\underline{ʂ}$]	"grape must"
p□aszcz	[pwa$\underline{ʂ}$t$\underline{ʂ}$]	"coat"
t□uszcz	[twu$\underline{ʂ}$t$\underline{ʂ}$]	"fat"

Table 2
Intensity of vowels across speech modes (dB), showing significant effects and interactions only.

Intensity		mean	s.d.	β	SE	df	t	p
Ref: whispered		47	5.9					
Semi-whispered		62	3.9	14.73	0.74	16	19.78	<0.001
Modal		71	4.6	24.41	1.07	15	22.79	<0.001
Ref. whispered	Question	49	5.6					
	statement	44	4.9	−5.05	0.31	31	−16.27	<0.001
Semi-whispered	Question	64	3.2					
	Statement	60	3.5	−3.92	0.3	30	−12.69	<0.001
Modal	Question	74	3.2					
	Statement	68	3.8	−6.11	0.3	30	−19.86	<0.001
Statement: modal				−1.05	0.29	2065	−3.60	<0.001
Statement: semi-whispered				1.13	0.29	2066	3.85	<0.001

Table 3
F0 maximum in modal speech (semitones), showing significant effects and interactions only.

F0 maximum	mean	s.d.	β	SE	df	t	p
Ref: question	18.66	5.44					
Statement	7.42	4.51	−11.26	0.49	15	−22.91	<0.001
Ref: male	8.39	6.17					
Female	17.10	6.10	8.04	0.85	13	9.45	<0.001

Table 4
F0 mean in modal speech (semitones), showing significant effects and interactions only.

F0 mean	mean	s.d.	β	SE	df	t	p
Ref: question	14.29	4.93					
Statement	5.78	4.70	−8.49	0.55	18	−15.31	<0.001
Ref: male	5.21	5.07					
Female	14.09	4.36	−8.97	0.78	13	11.39	<0.001

Table 5
F1 values across speech modes (Hz), showing significant effects and interactions only.

F1		mean	s.d.	β	SE	df	t-Value	p-Value
Ref: whispered		954	208					
Semi-whispered		478	222	−401	14.39	1062	−27.85	<0.001
Modal		595	221	−345	14	1062	−24.43	<0.001
Ref: whispered	Question	992	203					
	Statement	914	206	−88.1	14.05	1062	−6.26	<0.001
Semi-whispered	Question	466	198					
	Statement	491	244	28.4	14.4	1062	1.97	<0.05
Modal	Question	628	231					
	Statement	560	206	−69.4	13.9	1062	−4.97	<0.001
Ref: male		619	270					
Female		735	310	113	20.65	14	5.5	<0.001
Ref: /a/		1071	166					
/u/		829	174	−305	34.1	2	−8.95	<0.05
Question: semi-whispered				−116	20.12	1062	−5.79	<0.001

Table 6
F2 values across speech modes (Hz), showing significant effects and interactions only.

F2		mean	s.d.	β	SE	df	t-Value	p-Value
Ref: whispered		1436	242					
Semi-whispered		1165	270	−175	17	1007	−10.3	<0.001
Modal		1155	246	−239	17	1007	−14.06	<0.001
Ref: whispered	Question	1464	249					
	Statement	1407	233	−77	17.2	1007	−4.5	<0.001
Semi-whispered	question	1135	276					
	statement	1195	260	49.81	16	1007	3.01	<0.01
Ref: male		1185	204					
Female		1358	290	154	25.24	14	6.13	<0.001
Question: modal				−47.14	23.74	1007	−1.985	<0.001
Question: semi-whispered				−127.60	23.90	1007	−5.337	<0.001

Table 7
F3 values across speech modes (Hz), showing significant effects and interactions only.

F3		mean	s.d.	β	SE	df	*t*-Value	*p*-Value
Ref: whispered		2693	254					
Modal		2587	236	−88.4	31.34	22	−2.82	<0.05
Ref: whispered	Question	2748	243					
	Statement	2636	252	−120	28.13	30	−4.28	<0.001
Modal	Question	2623	237					
	Statement	2552	230	−83.7	28.05	30	2.98	<0.01

Table 8
Vowel duration (log), showing significant effects and interaction only.

Vowel duration	mean	s.d.	β	SE	df	*t*-Value	*p*-Value
Ref: whispered	−0.83	0.12					
Semi-whispered	−0.89	0.13	−0.062	0.005	2000	−11.35	<0.001
Modal	−0.89	0.13	−0.066	0.005	2000	−12.12	<0.001

Table 9
Centre of gravity, showing significant effects and interactions only.

Centre of Gravity (COG)		mean	s.d.	β	SE	df	*t*-Value	*p*-Value
Ref: whispered		3339	1133					
Semi-whispered		3466	1105	282	66.45	23	4.24	<0.001
Modal		4540	1029	1180	126.83	17	9.3	<0.001
Ref: whispered	Question	3579	3088					
	Statement	3088	1074	−502	74.29	26	−6.76	<0.001
Semi-whispered	Question	3582	1141					
	Statement	3354	1057	−217	76.89	23	−2.83	<0.001
Modal	Question	4823	952					
	Statement	4256	1025	−564	76.89	23	−7.34	<0.001
Ref: fricatives		4127	1146					
Affricates		3434	1184	−688	93.73	15	−7.34	<0.001
Repetition				50.39	14.51	4430	3.47	<0.001
Question: modal				52.90	57.65	4694	0.918	<0.001
Question: semi-whispered				−293	57.65	4430	−5.09	<0.001

Table 10
Standard deviation, showing significant effects and interactions only.

Standard deviation		mean	s.d.	β	SE	df	*t*-Value	*p*-Value
Ref: whispered		1688	508					
Semi-whispered		1796	495	66.41	20.33	44000	3.26	<0.01
Modal		1995	390	261.51	20.22	4400	12.92	<0.001
Ref: whispered	Question	1758	489					
	Statement	1614	517	−147.58	30.26	31	−30.26	<0.001
Ref: semi-whispered	Question	1830	486					
	Statement	1763	501	−62.50	30.24	31	−2.06	<0.05
Ref: fricatives		1900	476					
Affricates		1753	482	−149.09	44.28	15	−3.36	<0.01
Question: modal				−92.93	28.80	4628	−3.22	<0.01
Question: semi-whispered				−85.08	28.80	4628	−2.95	<0.01

Table 11
Skewness, showing significant effects and interactions only.

Skewness		mean	s.d.	β	SE	df	*t*-Value	*p*-Value
Ref: whispered		1.82	1.32					
Semi-whispered		1.66	1.22	−0.42	0.05	4212	−7.95	<0.001
Modal		0.89	0.85	−1.18	0.05	4215	−22.41	<0.001
Ref: whispered	Question	1.56	1.13					
	Statement	2.11	1.43	0.58	0.1	22	5.49	<0.001
Ref: fricatives		1.23	1.07					
Affricates		1.67	1.30	0.46	0.05	7	8.06	<0.001
Question: Modal				0.43	0.07	4214	5.84	<0.001
Question: Semi-whispered				0.45	0.07	4215	6.12	<0.001

Table 12
Kurtosis, showing significant effects and interactions only.

Kurtosis		mean	s.d.	β	SE	df	*t*-Value	*p*-Value
Ref: whispered		19.59	42.29					
Modal		3.23	5.71	−4.60	0.43	4274	−10.52	<0.001
Ref: whispered	Question	10.76	18.53					
	Statement	18.80	26.96	5.57	0.45	4275	12.19	<0.001
Ref: semi-whispered	Question	8.92	14.06					
	Statement	11.35	18.25	1.16	0.44	4274	2.61	<0.01
Ref: fricatives		4.84	7.07					
Affricates		8.55	11.65	4.01	0.25	4274	15.60	<0.001
Question: modal				5.55	0.62	4274	8.82	<0.001
Question: modal				4.40	0.63	4274	6.89	<0.001

Table 13
m1, showing significant effects and interactions only.

m1		mean	s.d.	β	SE	df	*t*-Value	*p*-Value
Ref: whispered		4.82	5.1					
Semi-whispered		4.97	4.81	0.64	0.17	4464	3.61	<0.001
Modal		5.71	3.9	1.40	0.17	4462	7.87	<0.001
Ref: whispered	Question	5.35	4.83					
	Statement	4.28	5.32	−1.06	0.24	36	4.32	<0.001
Ref: fricatives		5.91	4.49					
Affricates		4.43	4.71	−1.46	0.60	15	−2.42	<0.05
Question: modal				−0.98	0.25	4463	−3.93	<0.001
Question: semi-whispered				−0.89	0.25	4464	−3.54	<0.001

Table 14
m2, showing significant effects and interactions only.

m2		mean	s.d.	β	SE	df	*t*-Value	*p*-Value
Ref: whispered		−2.18	0.49					
Semi-whispered		−2.25	0.41	−0.18	0.01	4323	−11.66	<0.001
Modal		−2.58	0.25	−0.54	0.01	4324	−33.35	<0.001
Ref: whispered	Question	−2.32	0.43					
	Statement	−2.02	0.51	0.32	0.02	27	12.30	<0.001
Semi-whispered	Question	−2.3	0.38					
	Statement	−2.2	0.42	0.10	0.02	26	3.88	<0.001
Modal	Question	−2.62	0.24					
	Statement	−2.54	0.27	0.74	0.02	26	2.85	<0.01
Ref: fricatives		−2.46	0.35					
Affricates		−2.21	0.47	0.25	0.02	7	9.57	<0.001
Question: modal				0.25	0.02	4324	11.08	<0.001
Question: semi-whispered				0.24	0.02	4324	9.80	<0.001

Table 15
The frequency of the highest peak (2–4 kHz), showing significant effects and interactions only.

The highest peak		mean	s.d.	β	SE	df	*t*-Value	*p*-Value
Ref: whispered		2651	465					
Modal		2932	486	284	55.31	17	5.14	<0.001
Ref: whispered	Question	2749	481					
	Statement	2549	426	−200	19.95	59	−10.04	<0.001
Semi-whispered	Question	2690	464					
	Statement	2586	412	−97.28	19.93	59	−4.88	<0.001
Modal	Question	3026	483					
	Statement	2836	412	−184	20.05	60	−9.18	<0.001
Question: modal				−16.19	24.43	4444	−0.66	<0.001
Question: semi-whispered				−102	24.34	4443	−4.23	<0.01

Table 16
Fricative intensity, showing significant effects and interactions only.

Intensity		mean	s.d.	β	SE	df	*t*-Value	*p*-Value
Ref: whispered		49.84	6.51					
Semi-whispered		50.10	5.60	1.93	0.51	16	3.72	<0.001
Modal		57.30	5.16	7.39	0.82	16	8.9	<0.001
Ref: whispered	Question	51.90	6.47					
	Statement	47.70	5.83	−4.44	0.25	26	−17.73	<0.001
Semi-whispered	Question	50.94	5.74					
	Statement	49.43	5.51	−1.42	0.25	26	−5.67	<0.001
Modal	Question	59.76	4.51					
	Statement	54.95	4.63	−4.8	0.25	26	−19.17	<0.001
Ref: fricatives		54.60	6.60					
Affricates		50.20	6.15	−4.39	0.45	18	−9.7	<0.001
Question: modal				0.35	0.21	4403	1.68	<0.001
Question: semi-whispered				−3.02	0.21	4404	−14.29	<0.001

Table 17
Duration, showing significant effects and interactions only.

Duration	mean	s.d.	β	SE	df	*t*-Value	*p*-Value
Ref: whispered	−0.84	0.10					
Semi-whispered	−0.85	0.10	−0.013	0.002	4200	−6.12	<0.001
Modal	−0.86	0.10	−0.021	0.002	4200	−9.91	<0.001
Ref: fricatives	−0.89	0.09					
Affricates	−0.82	0.09	0.07	0.019	15	3.73	<0.001

References

Abramson, A. S. (1972). Tonal experiments with whispered Thai. In Valdman (Ed.), *Papers in linguistics and phonetics to the memory of Pierre Delattre* (pp. 31–44). The Hague: Mouton.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-9., https://CRAN.R-project.org/package=lme4.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology, 3*(01), 255–309.

Boersma, P., & Weenink, D. (2014). *Praat: doing phonetics by computer* [Computer program]. Version 5.3.57, retrieved 27 October, 2013. from http://www.praat.org/.

Brunner, J., Hoole, P., & Perrier, P. (2011). Adaptation strategies in perturbed /s/. *Clinical Linguistics and Phonetics, 25*(8), 705–724.

Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics, 31*(2), 149–180.

Cho, T., & Keating, P. (2001). Articulatory and acoustic studies of domain-initial strengthening in Korean. *Journal of Phonetics, 29*, 155–190.

Cho, T., & Keating, P. (2009). Effects of initial position versus prominence in English. *Journal of Phonetics, 37*(4), 466–485.

Cho, T., & McQueen, J. (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics, 33*(2), 121–157.

Cho, T. (2015). Language effects on timing at the segmental and suprasegmental levels. In M. A. Redford (Ed.), *The handbook of speech production* (pp. 505–529). Hoboken. NJ: Wiley-Blackwell.

Diehl, R. L. (2011). On the robustness of speech perception. In *International congress of phonetic science* (pp. 1–8). University of Hong Kong.

Dukiewicz, L. (1978). *Intonacja wypowiedzi polskich*. Wroc□aw: Ossolineum.

Eklund, I., & Traunmüller, H. (1996). Comparative study of male and female whispered and phonated versions of the long vowels of Swedish. *Phonetica, 54*, 1–21.

Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics, 29*, 109–135.

Fougeron, C., & Keating, P. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America, 101*, 3728–3740.

Gordon, M., & Nafi, L. (2012). The acoustic correlates of stress and pitch accent in Tashlhiyt Berber. *Journal of Phonetics, 40*, 706–724.

Grice, M. (2006). Intonation. *Encyclopedia of Language and Linguistics* (2nd ed.) (vol 5, pp. 778–788). Oxford: Elsevier.

Grynpas, J., Baker, R., & Hazan, V. (2011). Clear speech strategies and speech perception in adverse listening conditions. *International congress of phonetic science*, Hong Kong, pp. 779–782.

Heeren, W. F. L. (2015a). Vocalic correlates of pitch in whispered versus normal speech. *Journal of the Acoustical Society of America, 138*, 3800–3810.

Heeren, W. F. L. (2015b). Coding pitch differences in voiceless fricatives: Whispered relative to normal speech. *Journal of the Acoustical Society of America, 138*, 3427–3438.

Heeren, W. F. L., & van Heuven, V. J. J. P. (2009). Perception and production of boundary tones in whispered Dutch. *Proceedings of Interspeech, 2009*, 2411–2414.

Heeren, W. F. L., & van Heuven, V. J. J. P. (2011). Acoustics of whispered boundary tones: Effects of vowel type and tonal crowding. *Proceedings of the ICPhS XVII*, 851–854.

Higashikawa, M., & Minifie, F. D. (1999). Acoustical-perceptual correlates of "whisper pitch" in synthetically generated vowels. *Journal of Speech, Language, and Hearing Research, 42*, 583–592.

Higashikawa, M., Nakai, K., Sakakura, A., & Takahashi, H. (1996). Perceived pitch of whispered vowels – Relationship with formant frequencies: A preliminary study. *Journal of Voice, 10*, 155–158.

Hooper, J. B. (1976). *An introduction to natural generative phonology*. New York: Academic Press.

Ito, T., Takeda, K., & Itakura, F. (2005). Analysis and recognition of whispered speech. *Speech Communication, 45*, 139–152.

Jesus, L. M. T., & Shadle, C. H. (2002). A parametric study of the spectral characteristics of European Portuguese fricatives. *Journal of Phonetics, 30*(3), 437–464.

Jovičić, S. T., & Šarić, Z. (2008). Acoustic analysis of consonants in whispered speech. *Journal of Voice, 22*(3), 263–274.

Kallail, K. J., & Emanuel, F. W. (1984). Formant frequency differences between isolated whisper and phonated vowel samples produced by adult female subjects. *Journal of Speech and Hearing Research, 27*, 245–251.

Katsika, A. (2016). The role of prominence in determining the scope of boundary-related lengthening in Greek. *Journal of Phonetics, 55*, 149–181.

Kingston, J., & Diehl, R. L. (1995). Intermediate properties in the perception of distinctive feature values. In B. Connell & A. Arvaniti (Eds.), *Phonology and phonetics: Papers in laboratory phonology IV* (pp. 7–27). Cambridge: Cambridge University Press.

Kohler, K. J. (1990). Macro and micro F0 in the synthesis of intonation. In J. Kingston & M. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and the physics of speech* (pp. 115–138). Cambridge: Cambridge University Press.

Kohler, K. J. (Ed.). (2012). Bridging the segment-prosody divide in speech production and perception. *Phonetica, 69*(1–2) (special issue).

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package "lmerTest". R package version 2.0-29.

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

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Li, X.-L., & Xu, B.-L. (2005). Formant comparison between whispered and voiced vowels in Mandarin. *Acta Acoustica, 91*(6), 1079–1085.

Lim, B. P. (2011). *Computational differences between whispered and non-whispered speech* (Ph.D. dissertation). University of Illinois at Urbana-Champaign.

Lousada, M., Jesus, L. M. T., & Pape, D. (2012). Estimation of stops' spectral place cues using multitaper techniques. *DELTA, 28*(1), 1–26.

MathWorks (2007). *Signal Processing Toolbox 6 User's Guide*. Natick: MathWorks.

McLoughlin, I., Sharifzadeh, H. R., Tan, S. L., Li, J., & Song, Y. (2015). Reconstruction of phonated speech from whispers using formant-derived plausible pitch modulation. *ACM Transactions on Accessible Computing, 6*(4), 1–21.

Meyer-Eppler, W. (1957). Realization of prosodic features in whispered speech. *Journal of the Acoustical Society of America, 29*(1), 180–182.

Morris, R. W. (2003). *Enhancement and recognition of whispered speech* (Ph.D. dissertation). Atlanta, GA: Georgia Institute of Technology.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.

Niebuhr, O. (2008). Coding of intonational meaning beyond F0: Evidence from utterance-final /t/ aspiration in German. *Journal of the Acoustical Society of America, 124*(2), 1252–1263.

Niebuhr, O. (2009). Intonation segments and segmental intonation. In *Proceedings of Interspeech, 6–10 September, Brighton UK*. pp. 2435–2438.

Niebuhr, O. (2012). At the edge of intonation: The interplay of utterance-final F0 movements and voiceless fricative sounds. *Phonetica, 69*(1–2), 7–27.

Niebuhr, O., Lill, C., & Neuschulz, J. (2011). At the segment–prosody divide: The interplay of intonation, sibilant pitch and sibilant assimilation. *Proceedings of the ICPhS XVII*, 1478–1481.

Osfar, O. M. (2011). *Articulation of whispered alveolar consonants* (M.A. thesis). University of Illinois at Urbana-Champaign.

Pape, D., & Jesus, L. M. T. (2014). Cue-weighting in the perception of intervocalic stop voicing in European Portuguese. *Journal of the Acoustical Society of America, 136*(3), 1334–1343.

Pape, D., & Jesus, L. M. T. (2015). Stop and fricative devoicing in European Portuguese, Italian and German. *Language and Speech, 58*(2), 224–246.

Parker, E. M., Diehl, R. L., & Kluender, K. R. (1986). Trading relations in speech and nonspeech. *Perception & Psychophysics, 39*, 129–142.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation* (Ph.D. dissertation). MIT (distributed 1988, Indiana University Linguistics Club).

Pierrehumbert, J., & Talkin, D. (1992). Lenition of /h/ and glottal stop. In G. Docherty & D. R. Ladd (Eds.), *Papers in laboratory phonology II: Segment, gesture, tone* (pp. 90–117). Cambridge: Cambridge University Press.

Prieto, P., van Santen, J., & Hirschberg, J. (1995). Tonal alignment patterns in Spanish. *Journal of Phonetics, 4*, 429–451.

R Development Core Team (2010). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. http://www.R-project.org/, version 3.0.2., retrieved 25.09.2013.

Repp, B. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin, 92*(1), 81–110.

Ritter, S., & Roettger, T. (2014). Speakers modulate noise-induced pitch according to intonational context. In *Proceedings of 7th international conference on speech prosody*. pp. 890–894.

Roettger, T. B., & Grice, M. (2015). The role of high pitch in Tashlhiyt Tamazight (Berber): Evidence from production and perception. *Journal of Phonetics, 51*, 36–49.

Rubach, J., & Booij, G. E. (1985). A grid theory of stress in Polish. *Lingua, 66*, 281–319.

Schwartz, M. F. (1968). Effect of masking noise upon syllable duration in oral and whispered reading. *Journal of the Acoustic Society of America, 43*, 169–170.

Schwartz, M. F. (1972). Bilabial closure durations for /p/, /b/, and /m/ in voiced and whispered vowel environments. *Journal of the Acoustical Society of America, 51*, 2025–2029.

Selkirk, E. O. (1986). On derived domains in sentence phonology. *Phonology, 3*, 371–405.

Selkirk, E. O. (1984). On the major class features and syllable theory. In M. Arnoff & R. T. Oehre (Eds.), *Language and sound structure*. Cambridge: MIT Press.

Selkirk, E. O. (1978). On prosodic structure and its relation to syntactic structure. In T. Fretheim (Ed.), *Nordic Prosody II* (pp. 111–140). Trondheim: Tapir.

Sharf, D. J. (1967). Vowel duration in whispered and in normal speech. *Language and Speech, 7*, 89–97.

Sharifzadeh, H. R. (2010). *Reconstruction of natural sounding speech from whispers* (Ph.D. dissertation). Singapore: Nanyang Technological University.

Sharifzadeh, H. R., & McLoughlin, I. (2011). *Reconstruction of normal sounding speech for laryngectomy patients through a modified CELP codec*. Nanjing, China: EPS International Forum on Rehabilitation Medicine.

Shattuck-Hufnagel, S., & Turk, A. (1998). The domain of phrase-final lengthening in English. *The sound of the future: A global view of acoustics in the, 21st century, proceedings of the 16th international congress on acoustics and 135th meeting Acoustical Society of America*. pp. 1235–1236.

Sluijter, A. M. C. (1995). *Phonetic correlates of stress and accent* (Ph.D. thesis). Leiden: Holland Institute of Generative Linguistics.

Steffen-Batogowa, M. (1966). Versuch einer strukturellen Analyse der polnischen Aussagemelodie. *Zeitschrift für Phonetik und Allgemeine Sprachwissenschaft, 19*, 398–440.

Stevens, K. N. (1998). *Acoustic phonetics*. Cambridge, MA: MIT Press.

Swerdlin, Y., Smith, J., & Wolfe, J. (2010). The effect of whisper and creak vocal mechanisms on vocal tract resonances. *Journal of the Acoustical Society of America, 127*, 2590–2598.

Tartter, V. C. (1989). What's in a whisper? *Journal of the Acoustical Society of America, 86*(5), 1678–1683.

Thomas, I. B. (1969). Perceived pitch of whispered vowels. *Journal of the Acoustical Society of America, 46*(2B), 468–470.

Toda, T., Nagiri, M., & Shikano, K. (2012). Statistical voice conversion techniques for body-conducted unvoiced speech enhancement. *IEEE transactions on Audio, Speech, and Language Processing, 20*(9), 2505–2517.

Wagner, A. (2008). *A comprehensive model of intonation for application in speech synthesis* (Ph.D. dissertation). Poznań: Adam Mickiewicz University.

Weismer, G., & Bunton, K. (1999). Influences of pellet markers on speech production behavior: acoustical and perceptual measures. *Journal of the Acoustical Society of America, 105*, 2882–2894.

Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica, 49*, 25–47.

Winter, B. (2014). Spoken language achieves robustness and evolvability by exploiting degeneracy and neutrality. *BioEssays, 36*, 960–967.

Winter, B., & Christiansen, M. H. (2012). Robustness as a design feature of speech communication. In T. C. Scott-Phillips, M. Tamariz, E. A. Cartmill, & J. R. Hurford (Eds.), *Proceedings of the 9th international conference on the evolution of language* (pp. 384–391). New Jersey: World Scientific.

Yan, Y., Ahmad, K., Kenduk, M., & Bless, D. (2005). Analysis of vocal-fold vibrations from high-speed laryngeal images using a Hilbert transform-based methodology. *Journal of Voice, 19*, 161–175.

Żygis, M., & Malisz, Z. (Eds.). (2016). Slavic perspectives on prosody. *Phonetica, 73* (special issue, vol. 3 (Slavic Prosody) & 4 (Interfaces in Slavic Prosody)).