# Intended intonation of statements and polar questions in Polish in whispered, semi-whispered and normal speech modes.

*Marzena Żygis*[1], *Daniel Pape*[2], *Luis M.T. Jesus*[2,3], *Marek Jaskuła*[4]

[1] Centre for General Linguistics (ZAS) & Humboldt University, Berlin, Germany
[2] IEETA, University of Aveiro, Portugal
[3] ESSUA, University of Aveiro, Portugal
[4] West Pomeranian University of Technology, Szczecin, Poland

zygis@zas.gwz-berlin.de,danielpape@ua.pt,lmtj@ua.pt, Marek.Jaskula@zut.edu.pl

## Abstract

This paper examines acoustic correlates of intonation in Polish whispered, semi-whispered and normal speech modes. In particular, it investigates correlates of utterance-final rising intonation in polar questions and falling intonation in statements. The paper examines not only properties of vowels but also properties of the following voiceless consonant clusters.

The study is based on measurements of 4608 sibilants (fricatives and affricates) produced by 16 native speakers of Polish. The results point to differences in spectral properties of both utterance-final vowels and consonants where falling intonation in statements contrasts with rising intonation in polar questions. Regarding the consonants, both fricatives and affricates are produced with higher spectral peaks, higher intensity and higher COG and STD values in questions than in statements. Skewness and kurtosis values are lower in questions than in statements. Some spectral differences of sibilants, including spectral slopes, are more distinguishable for questions versus statements in the whispered speech mode than in other speech modes. The more pronounced role of these cues in whispered speech suggests their compensatory function for the fundamental frequency, which is the main correlate of intonation in phonated speech but is completely absent in whispered speech.

In summary, the study shows that speakers produce intended intonation patterns by varying the choice of cues as well as their magnitude in dependence on both (i) speech modes and (ii) intonation patterns.

**Index Terms:** whispered speech, intonation, voiceless clusters, Polish

## 1. Introduction

Interaction between segments and prosody, in general, and between segments and intonation in particular, still belongs to an understudied area of phonetic and phonological research. Relatively little is known about this interaction in whispered speech, and less so in semi-whispered speech; cf. [1] for studies of this interaction in the normal speech mode.

The overall goal of this paper is two-fold. First, it is aimed at providing an insight into the realisation of intended intonation in whispered speech where F0, the main correlate of intonation is completely absent and semi-whispered speech, where F0 is partially absent. Second, the study is intended to contribute to a better understanding of the role of voiceless segments in conveying different intonation patterns in various speech modes.

In particular, the paper addresses two questions:

(1) How are different patterns of intonation produced in whispered speech in comparison to semi-whispered and normal speech modes?

(2) How do voiceless consonant clusters contribute to intended intonation patterns in all three speech modes?

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

## 2. Methods

We recorded eight different items ending in clusters consisting of voiceless retroflex fricatives followed by retroflex affricates. Each item was presented in a polar question (‘Widzi ten blu[s̢t͡s̢]?’ ‘*Does he see the ivy*?’) and a statement (‘Widzi ten blu[s̢t͡s̢]’ ‘*He sees the ivy*’). All words were monosyllabic. The polar question was expected to be produced with a rising intonation and the statement with a falling intonation. In order to compare the realisation of intonation in whispered speech with other speech modes we recorded the questions and statements described above in three different speech modes: normal, whispered and semi-whispered.

Sixteen native speakers of Polish (eigth male), aged 20 to 52 years, took part in the experiment. All speakers were monolingual and spoke Standard Polish. They were asked to read the sentences in all three speech modes, as described above. All recordings were conducted in the sound-proof room at the Electrical Engineering Department of the West Pomeranian University of Technology in Szczecin using a TLM103 Neumann microphone (20cm distance from lips) connected to a ProTools system with a Digi 003 interface (sample rate 44100 Hz). The items were analysed with Praat (version 5.3.57 [2]) and MATLAB (version R2007b [3]). In total, measurements of 4608 sibilants were taken (8 items × 2 sentence types (question, statement) × 3 speech modes (normal, semi-whispered, whispered) × 2 sibilant types (fricative, affricate) × 3 repetitions × 16 speakers).

We investigated the following acoustic parameters, see [4] for all parameter details. Note that the parameters displayed in (f), (g) and (h) were calculated at the onset, midpoint and offset of the frication of both sibilants.

(a) duration of the vowel, fricative and affricate;

(b) maximum and mean of F0 of the preceding vowel;

(c) F0 difference between the vowel offset and onset;

(d) formants F1, F2, F3 (using the formant extraction of PRAAT [2]) at the (i) vowel onset, (ii) midpoint and (iii) vowel offset;

10.21437/SpeechProsody.2014-123

(e) mean intensity over the complete duration of the vowel, the fricative and the affricate;

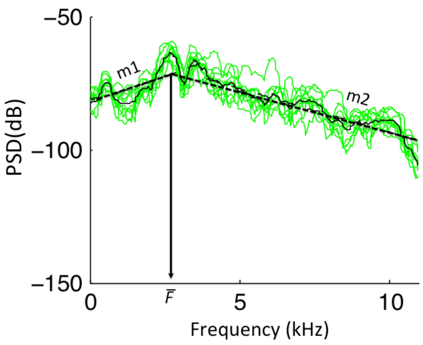
(f) frequency of the highest spectral peak of the frication noise in the range from 20Hz to11kHz;

(g) spectral Centre of Gravity (COG), its standard deviation (STD), skewness, and kurtosis;

(h) the spectral regression slopes, as described in [5]: 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 11025 kHz (see Figure 1).

We computed multitaper spectra with a 12 ms window for the frication noise midpoint (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 [3].

Figure 1 shows an example of 10 different multitaper spectra, from one individual speaker, the overlaid mean spectrum and the computation of the regression lines m1 and m2, with the endpoint/startpoint $\bar{F}$.



(i)

Figure 1: *Multitaper spectra (light colour) with 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* $\bar{F}$.

Regarding statistical analysis, linear mixed effect models were employed for the investigated variables, which were studied as effects of INTONATION (rising vs. falling) and SPEECH MODE (normal, semi-whispered and whispered), as well as their interaction (INTONATION*SPEECHMODE). GENDER (male, female) was included as fixed effect as well. In addition, speaker-specific random slopes for INTONATION TYPE and SPEECH MODE were included into the model. ITEM and SPEAKER were taken as random effects.

All analyses were conducted in the R environment software (version 3.0.2) [6].

## 3. Results

Our results show that different acoustic cues are used to a different extent when producing questions versus statements in various speech modes. In the following we report only those results which – due to their significance – are important for answering the main questions of this study, cf. (1) and (2).

Regarding duration of the vowel, whispered vowels were generally longer than vowels produced in semi-whispered (t=7.304, p<.0001) and normal speech mode (t=5.799, p<.0001). No significant differences were found between statements and questions in all three speech modes.

As was expected, the F0 maximum and the F0 mean of the vowel preceding the sibilant cluster were significantly higher in questions than in statements (F0 max: t=11.21, p <.0001, F0 mean: t=13.41, p <.0001). In addition, we calculated the F0 difference between the vowel offset and onset which in fact points to raising F0 in questions (female:142 Hz, male:81 Hz) and falling F0 in statements (female:-15 Hz, male:-17 Hz) confirming our initial assumption about F0 differences in intonation patterns (t=10.726, p<.0001); cf. also [7], [8]. 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 be reliably extracted) we did not analyse this parameter in those two modes.

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

First, our results point to the importance of formants. The first formant frequency is significantly higher in questions than in statements in whispered (t= 6.23, p<.0001) and normal speech (t=5.174 p<.0001) but lower in semi-whispered speech (t=-3.726 p<.0001). Furthermore, the second formant frequency of the vowel is significantly higher in questions than in statements only in whispered speech (t=2.812, p<.01) but lower in semi-whispered speech (t=-6.485 p<.0001). In normal speech mode the formants do not show any significant difference when questions and statements are performed. Finally, the third formant frequency is higher in questions than in statements in whispered (t=4.15, p<.0001) and normal speech (t=1.76, p<.05).

Regarding the mean intensity of the vowel, the results show that it is significantly higher in questions than in statements across all speech modes: whispered (t=16.76, p<.0001), normal (t=20.67, p<.0001) and semi-whispered (t=14.91, p<.0001).

Besides clear differences in the realisation of whispered vowels in questions and statements, our results point to significant differences in sibilant clusters depending on the intonation type.

Regarding duration, fricatives were longer in whispered speech than normal speech mode (t=2.941, p.<01). Similarly, duration of affricates was longer in whispered speech than semi-whispered (t=1.774, p<.05) and normal speech modes (t=2.493 p<.001). No differences in duration of both fricatives and affricates were found regarding question vs. statement distinction across all three speech modes.

Furthermore, considerable differences were found in spectra of the consonants.

Figure 2 presents multitaper spectra of all recorded items of 8 male speakers obtained at the acoustic midpoint of frication for all three speech modes, where the black lines show statement conditions and the grey lines question conditions.

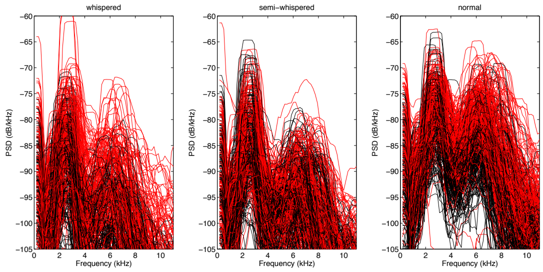


Figure 2: Multitaper spectra of 1145 affricate spectra from 8 male speakers obtained at the acoustic frication midpoint for all three speech modes in separate panels.

With regard to the spectral regression lines slopes of the fricative (m1, m2, cf. Figure 1), m1 exhibits significantly lower values in whispered than in normal speech (t=-2.288, p<.05) and semi-whispered speech (t=-1.997, p<.05). The m2 value is significantly higher in whispered than in normal speech (t=4.01, p<.0001) and semi-whispered speech (t=4.35, p<.0001). Only in whispered fricatives, a distinction in m1 is significant when comparing questions vs. answers, i.e. whispered fricatives display higher m1 values in questions than in statements (t=3.349, p<.001). Regarding m2, questions are produced with a lower m2 in comparison to statements in whispered (t=-8.581, p<.0001) and semi-whispered fricatives (t=-2.19, p<.05).

The spectral slope m1 at the midpoint of the following affricate shows significantly lower values in whispered than in normal speech mode (t=-3.757, p<.0001). Similar to fricatives, a distinction between questions and statements was only found in whispered speech mode (t=3.290, p<.001). Furthermore, with respect to m2, the results show significantly higher m2 values in whispered than in semi-whispered (t=3.915, p<.0001) and normal speech mode (t=-4.67, p<.0001). Again, questions are produced with lower m2 values than statements across all three speech modes differing in t-values: whispered speech mode (t=-12.60, p<.0001), normal speech mode (t=-4.67, p<.0001) and semi-whispered speech mode (t=-4.194, p<.0001). Both slopes are shown in Figure 3; cf. also Figure 1.

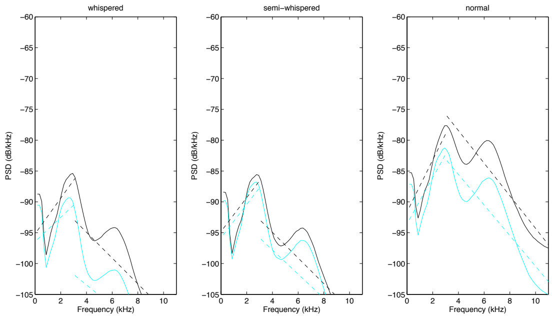


Figure 3: *Multitaper spectra (mean plots over all items and speakers) for the frication midpoint of affricates in whispered, semi-whispered and normal speech mode. Black solid lines correspond to the question and lighter colour to the statement condition. Dotted lines are the spectral regression lines m1 and m2.*

The mean intensity of both fricatives and affricates differs significantly for questions compared to statements: the fricative displays a higher intensity mean for questions than for statements (whispered: t=17.01, p<.0001, normal: t=15.76, p<.0001, semi-whispered: t=5.29, p<.0001). A similar pattern applies for affricates (whispered: t=12.99, p<.0001, normal: t=16.63, p<.0001, semi-whispered: t=4.11, p<.0001).

The frequency of the highest peak of the fricative at its midpoint is higher in questions as opposed to statements for all three speech modes: whispered (t=2.737, p<.01), semi-whispered (t=3.259 p<.001) and normal (t=4.886, p<.0001). Similar observations apply to the affricate midpoint, but exclusively to whispered (t=3.149, p<.001) and normal speech (t=8.726, p<.0001).

Significant differences are also found for the four spectral moments (COG, STD, skewness and kurtosis) at the sibilants' midpoint. The COG values are significantly higher for questions than for statements across all speech modes. This conclusion holds true for both fricatives and affricates whereby the t-values are higher for whispered and normal speech in comparison to the semi-whispered speech mode. The results are presented in Table 1 together with the COG's standard deviation (SD) at the midpoints of both sibilants.

Table 1: Comparison of COG and SD values for the difference between questions and statements across three different speech modes.

| | whispered | semi-whisp. | normal |
|---|---|---|---|
| Fricative (COG) | t=5.316<br>p<.0001 | t=2.977<br>p<.01 | t=4.99<br>p<.0001 |
| Affricate (COG) | t=6.101<br>p<.0001 | t=1.574<br>n.s. | t=7.88<br>p<.0001 |
| Fricative (STD) | t=5.923<br>p<.0001 | t=1.786<br>p<.05 | t=0.82<br>n.s. |
| Affricate (STD) | t=3.375<br>p<.001 | t=1.153<br>n.s | t=2.43<br>p<.01 |

Regarding skewness, the third spectral moment, our results indicate significant differences between all three speech modes. For fricatives, skewness is significantly higher in whispered speech compared to normal (t=4.418, p<.0001) and semi-whispered speech (t=3.297, p<.0001). In the same vein, frication in affricates displays higher skewness values in whispered than in normal (t=7.096, p<.0001) or semi-whispered speech mode (t=4.343, p<.0001). If we compare the production of questions in comparison statements, the results indicate lower skewness values for questions as compared to statements for fricatives in whispered speech mode only (t=-3.949, p<.0001). In affricates, the skewness is lower in questions than in statements for both whispered (t=-6.636, p<.0001) and normal speech (t= -1.975, p<.05).

The fourth spectral moment, kurtosis, is significantly higher for fricatives in whispered than in semi-whispered (t=3.290, p<.001) and normal speech mode (t=3.842, p<.0001). However, when comparing questions to statements, kurtosis values are significantly smaller in whispered speech (t=-5.656 p<.0001). Regarding affricates, the results show that kurtosis was significantly higher in whispered than in semi-whispered (t=3.968, p<.0001) and normal speech mode (t= 6.066, p<.0001). However, a significant difference in the production of statements vs. questions is found only in whispered speech, where kurtosis is lower in questions (t=-8.163, p<.0001).

## 4. Discussion

The results show that, in Polish, different intonation patterns between questions and statements are produced by means of various acoustic cues which are, in turn, dependent on the speech mode. In whispered speech, where the F0 is entirely absent, intended rising intonation in questions is produced by both vocalic and consonantal cues. Regarding the former, the results point to a higher F1 and F2 in the utterance-final vowel, a higher amplitude in questions only and no difference in vowel duration between questions and statements.

These results are in accordance with findings reported for the Dutch language where it was shown that in whispered speech F1 and F2 are higher in questions than statements for /ə/ [9]. However, a higher F1 was not found when other Dutch vowels were investigated [10]. Similarly to the results of the present study, the amplitude of the vowel was higher in questions as opposed to statements [9], [10]. In addition, no difference in vowel duration was found between whispered questions and statements [9].

Whereas the majority of previous studies have focused on the properties of vowels when investigating intonation/pitch [11], [12], [13], the results of the present study also point to spectral differences of utterance-final consonants [14], [15]. Significant differences in virtually all spectral parameters are found when comparing questions to statements. The former are produced with higher spectral COG values. This finding is in line with [14], where higher COG values of voiceless fricatives were reported for questions in German phonated speech mode. In the present study, higher COG values are also found for both fricatives and affricates in whispered and semi-whispered questions as compared to statements. The higher COG values are accompanied by higher SD values (with the exception of affricates in semi-whispered speech mode).

The third and the fourth spectral moments are of special importance as they differ considerably between questions and statements across all speech modes. The significantly higher skewness in whispered speech indicates that the mass of the spectral distribution is towards lower frequency values in whispered speech in contrast 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 affricates, the latter difference applies to normal speech as well, but it is considerably less pronounced in normal speech than in whispered speech.

The higher kurtosis (peakedness; width of peak) in whispered speech indicates a narrower spectral peak for fricatives and affricates in comparison to their semi-whispered and normal speech counterparts. The width of peak in frication of both fricatives and affricates is wider in questions than in statements in whispered speech. In affricates, the width of peak is wider in questions than in statements in both speech modes. Thus, the question vs. statement differences with regard to the third and fourth spectral moment are found exclusively in whispered speech.

Furthermore, intensity is higher in questions than in statements in both sibilants across all three speech modes. Similarly, the highest spectral peak is found at higher frequencies in questions than statements (with the exceptions of affricates in semi-whispered speech).

Finally, regarding the spectral regression line slope m1, the spectra of semi-whispered and normal speech rise more steeply in the frequency range of 500-3000Hz than the spectra of whispered speech. However, a significant m1 difference between questions and statements is found in whispered fricatives and affricates where questions are produced with steeper m1 slopes in comparison to statements. Regarding the spectral regression line slope m2 (from 3kHz to 11 kHz), the spectra for questions in whispered and semi-whispered fricatives fall more steeply above 3kHz compared to the spectra of statements. For affricates, the spectra of questions also fall more steeply above 3kHz than statements but the difference in steepness was largest in whispered followed by normal and then semi-whispered speech modes.

## 5. Conclusions

In summary, our results point to differences in spectral properties of not only vowels but also consonants when statements and polar questions are produced. These differences, especially with regard to consonants, are intensified when speakers switch from normal speech to whispered speech. In particular, questions are not only realized with a higher F1 and F2 of the vowel but also with a higher intensity, higher spectral peak frequency, higher COG, and SD as well as lower kurtosis and skewness of the consonant. Further differences are found in the spectral regression line slopes.

Some spectral differences between the production of questions and statements are found exclusively, or to the greatest extent, in whispered speech, emphasising a relevance of these cues for this particular speech mode. Moreover, the more pronounced role of these cues in whispered speech suggests a compensation for the distinguishing role of F0 in phonated speech mode. The perceptual relevance of this finding requires further investigation.

This study also sheds further light on the interaction of segments and intonation. It shows, in fact, that taking intonation patterns into consideration seems to be indispensable if spectral properties of segments - in this case sibilants - are to be investigated. Differences in intonation patterns significantly change consonantal properties. This conclusion applies to all speech modes: whispered, semi-whispered and normal.

Finally, due to the investigation of different speech modes, the study provides new aspects for a discussion of the phonological concept of intonation. It shows that speakers produce the intended intonation patterns, encoded at the underlying level, by varying a choice of acoustic cues, i.e., cues-trading as well as their magnitude in dependence on both (i) speech modes and (ii) intonation patterns.

## 6. Acknowledgements

This research has been supported by Bundesministerium für Bildung und Forschung (BMBF, Germany) Grant Nr. 01UG1411 to Marzena Żygis. The study was also funded by FEDER through the Operational Program Competitiveness Factors – COMPETE and by Portuguese National Funds through FCT – Foundation for Science and Technology in the context of the project FCOMP-01-0124-FEDER-022682 (FCT reference PEst-C/EEI/UI0127/2011) to IEETA to Luis Jesus and Daniel Pape, and the post-doctoral fellowship from FCT (Portugal) SFRH/BPD/48002/2008 to Daniel Pape.

## 7. References

[1] K. Kohler (ed.), "Bridging the Segment-Prosody Divide in Speech Production and Perception", *Phonetica,* vol. 69, 2012.

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

[3] MathWorks, "Signal Processing Toolbox 6 User's Guide", N. MathWorks, Ed., 2007.

[4] M. Żygis, D. Pape, and L. M. T. Jesus, "(Non)retroflex Slavic affricates and their motivation. Evidence from Czech and Polish", *Journal of the International Phonetic Association,* 42: 281-329, 2012.

[5] L. M. T. Jesus and C. Shadle, "A parametric study of the spectral characteristics of fricatives," *Journal of Phonetics,* 30: 437-464, 2002.

[6] R Development Core Team "R: A Language and Environment for Statistical Computing", R Foundation for Statistical Computing, Vienna, http://www.R-project.org/, version 3.0.2.

[7] G. Demenko & A. Wagner, "Prosody annotation for unit selection text-to-speech". Archives of acoustics 32(1):25-40.

[8] A. Wagner "A comprehensive model of intonation for application in speech synthesis", Ph.D. thesis, Poznań. 2008.

[9] W. Heeren and V.J. Van Heuven, "Perception and Production of Boundary Tones in whispered Dutch", Proceedings of Interspeech, 2011-2414, 2009.

[10] W. Heeren and V.J. Van Heuven, "Acoustics of whispered boundary tones: Effects of vowel type and tonal crowding", *Proceedings of the ICPhS XVII,* 851-854, 2011.

[11] X.-L., Li and B-L. Xu, "Formant comparison between whispered and voiced vowels in Mandarin", *Acta Acustica* 91: 1079-1085, 2005.

[12] W. Meyer-Eppler, "Realization of prosodic features in whispered speech", *Journal of the Acoustical Society of America,* 29: 180-182, 1957.

[13] I.B. Thomas, "Perceived pitch of whispered vowels" Journal of the Acoustical Society of America (46): 468-470, 1969.

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

[15] O. Niebuhr, C.Lill and J.Neuschulz, "At the segment-prosody divide: The interplay of intonation, sibilant pitch and sibilant assimilation," *Proceedings of the ICPhS XVII*, 1478-1481, 2011.