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