Perception of Glottalization in Varying Pitch Contexts in Mandarin Chinese

Maria Paola Bissiri[1], *Margaret Zellers*[2], *Hongwei Ding*[3, 4]

[1] Institut für Akustik und Sprachkommunikation, Technische Universität Dresden, Germany
[2]Department of Speech, Music & Hearing, Kungliga Tekniska Högskolan, Stockholm, Sweden
[3]School of Foreign Languages, Tongji University, Shanghai, China
[4]School of Foreign Languages, Jiao Tong University, Shanghai, China
Maria_Paola.Bissiri@tu-dresden.de, zellers@kth.se, hongwei.ding@tongji.edu.cn

Abstract

Although glottalization has often been associated with low pitch, evidence from a number of sources supports the assertion that this association is not obligatory, and is likely to be language-specific. Following a previous study testing perception of glottalization by German, English, and Swedish listeners, the current research investigates the influence of pitch context on the perception of glottalization by native speakers of a tone language, Mandarin Chinese. Listeners heard AXB sets in which they were asked to match glottalized stimuli with pitch contours. We find that Mandarin listeners tend not to be influenced by the pitch context when judging the pitch of glottalized stretches of speech. These data lend support to the idea that the perception of glottalization varies in relation to language-specific prosodic structure.

Index Terms: prosody, voice quality, perception, glottalization

1. Introduction

Glottalization, and in particular creaky phonation, is associated across languages with lowered amplitude, positive spectral tilt, and higher first formant values than modal voice (Gordon & Ladefoged [1]), and with low fundamental frequency (F0), damping, and aperiodicity (Gerratt & Kreiman [2]). Because of its low F0, creak/glottalization has been classically described as giving the impression ... run along a ... ([3]: 32). Glottalization, despite its ... acoustic complexity, can be identified with 95% accuracy by listeners (Blomgren, Chen, Ng & Gilbert [4]), and can therefore be considered a robust tool for phonetic signaling.

Glottalization's acoustic characteristics have often led to it being associated with low pitch, especially in intonation languages. However, associating glottalization with low pitch on the basis of its having low F0 is problematic. Many tone languages associate creakiness with high tones (Gordon & Ladefoged [1]; Gussenhoven [5]), and this can be the case even in intonation languages (Pierrehumbert & Talkin [6]; Pierrehumbert [7]; Dilley, Shattuck-Hufnagel & Ostendorf [8]; Redi & Shattuck-Hufnagel [9]). Dilley et al. [8] found, for example, more frequent glottalizations of word-initial vowels if the target words were pitch-accented. Since in their corpus L* accents were rare, they claimed that glottalization was influenced by pitch-accents per se and not by low F0 ([8]: 439). Pierrehumbert [7] also reported glottalization associated with pitch accents.

In intonation languages, creaky voice and other forms of glottalization have often been associated with prosodic boundaries. Henton & Bladon [10] and Redi & Shattuck-Hufnagel [9] reported that in English glottalization is likely in utterance-final position, while Pierrehumbert & Talkin [6] and Huffman [11] found its presence at prosodic boundaries. Ogden [12, 13] noted that glottal productions characterize possible turn transition locations in Finnish. Creak and/or glottalization have also been shown to occur in other positions. For example, the glottalization of word-initial vowels is a frequent word juncture marker in Czech and in German (Bissiri et al. [14]; Kohler [15]).

Tone languages may demonstrate some of the same intonational features as intonation languages, and this appears to hold true for glottalization features. Ding et al. [16] found glottalization at the beginnings of vowel-initial syllables in Mandarin Chinese, similar to the effects reported above in German, English and Czech. In Mandarin, tone itself can also have an influence on the occurrence of glottalization. Chao [17] observed glottalization in Mandarin's falling-rising Tone 3. Ding & Helbig [18] reported glottalization most frequently in the middle (i.e. associated with the pitch valley) of Tone 3 and sometimes at the end of the falling Tone 4. This would appear to ... [19] and ... [20] ... proposals that tone and glottalization are perceived sequentially, at least in the case of Mandarin Chinese. It therefore remains the case, as Ní Chasaide & Gobl [21] pointed out, that studying pitch perception without analyzing the influence of glottalization is likely to lead to incomplete results.

Bissiri & Zellers [22] investigated the perception of glottalization in different pitch contexts by German, English and Swedish listeners. They found that long stretches of glottalization tended to be associated with low pitch by listeners in all three of these languages, although listeners could also associate glottalization with mid or rising pitch in a substantial minority of cases. The pitch contour preceding the glottalized stretch influenced ... judgments: when the glottalization was preceded by a rising pitch contour, it was more likely to be perceived as falling pitch. Swedish listeners, however, did not show this perceptual effect. Bissiri & Zellers [22] proposed that this was because Swedish listeners' experience w... system in Central Swedish made them more sensitive to pitch in general, and therefore more likely to respond based on the actual F0 of the stimuli, than the English and German listeners, who may have relied more on their predictions about the continuing direction of the pitch contour.

If listeners from a pitch-accent language background are more sensitive to F0 than listeners from an intonation language background, it follows that listeners from a tone language should also be more sensitive to F0. The study reported below investigates Mandarin Chinese ... glottalized stretches in different pitch contexts.

2. Perception experiment

The current study uses the same AXB forced-choice paradigm as Bissiri & Zellers [22] to investigate whether Mandarin Chinese listeners would consistently associate glottalization of a final syllable with a falling, level, or rising pitch contour in that position. Following Bissiri & Zellers' findings for Swedish versus German and English, we hypothesize that Mandarin listeners, coming from a tone language background, will not be influenced by the surrounding pitch contour.

2.1. Methodology

2.1.1. *Stimuli*

The stimuli used for this experiment were the same as those used by Bissiri & Zellers [22], and a more detailed account of their creation can be found there. Stimuli consisted of the syllable sequences /da'dada/, /la'lala/, and /na'nana/. Although /da/, /la/ and /na/ are acceptable Chinese syllables, the sequences are nonsense words in Chinese. By means of a Praat (Boersma & Weenink [23]) script, twelve versions of each syllable sequence were created. Nine of them were the Pitch conditions. A flat pitch contour of 220Hz was set over the whole syllable sequence. Then, four pitch points were added: at the onset of the first consonant, the onsets of the middle and final vowels, and the end of the final vowel. The second pitch point (onset of stressed vowel) was either raised 3 semitones above the initial pitch (fall), lowered 3 semitones below the initial pitch (rise), or kept level (mid). Similarly, the fourth pitch point (end of final vowel) was raised 3st, lowered 3st, or kept level. The third pitch point (onset of the final vowel) was held constant. In this way, 9 pitch contours were created (see Figure 1).

For each initial pitch condition (mid, rising, or falling), a further version was created with the final syllable containing glottalization instead of a smooth pitch contour; these stimuli constitute the Glottalization set. The glottalization was taken from naturally-produced tokens, and was matched in length to the final syllables of the Pitch conditions. In order to make sure that the consonant preceding the glottalization remained clear, the glottal stretches were spliced into the stimuli following the formant transitions. Using a Praat script, four additional pitch points were added in the transitions to create the impression of glottalization. These pitch points, with values of 35Hz, 220Hz, 35Hz, and 220Hz respectively, were equally spaced through the duration of the formant transitions. The amplitude of the entire glottalized part of the stimuli was then reduced in order to improve naturalness. All Pitch and Glottalization stimuli were then amplitude-normalized.

2.1.2. *Participants*

Participants in the experiment were native speakers of Mandarin Chinese (N=21) between 20 and 27 years old, students at the Tongji University in Shanghai.

2.1.3. *Experiment procedure*

The experiment was presented by means of Praat's ExperimentMFC. Participants heard sequences of three stimuli. Their task was to compare the second stimulus to the first and third, and to decide which it sounded most like. The

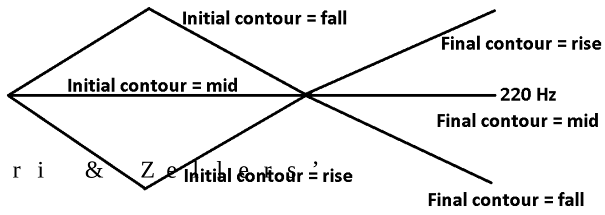

Figure 1: Schematic of pitch contours.

AXB paradigm was chosen instead of ABX because the target stimulus is adjacent to both options, thus putting fewer demands on memory. AXB sets were created from the same syllable set (/da/, /la/, or /na/), and also with the same initial pitch contour. The A and B options were always stimuli from the Pitch set, while the target item X was from the Glottalization set. Since participants in AXB tests tend to choose the more recent and thus better remembered B option, each critical pitch contour comparison was presented in both AXB and BXA format. Thus, each participant heard 126 stimuli in total: 54 test comparisons, in which the X item contained a glottalization, and 72 control comparisons, in which all the items were drawn from the Pitch set. For each participant, the order of presentation of all sequences was randomized by Praat.

The experiment was preceded by a practice session to get participants accustomed to the task. During practice, participants heard 3 control items with the syllables /ba'baba/. According to feedback from a pilot participant, optional breaks after every 32 items were included in the experiment. The time taken by participants to complete the task was between 17-20 minutes.

2.1.4. *Data cleanup*

Prior to the data analysis, participants' responses to control items were checked to make sure they had accurately matched pitch patterns. The control items always contained two identical contours and one that was different; any participants who failed to identify the identical contours in more than 10% of the control items were considered to have not performed the task correctly. One Chinese listener, who systematically gave the incorrect response in 93% of the control stimuli, was excluded on this basis. Of the 20 Chinese participants whose results were kept, 16 had an accuracy above 95%.

2.2. Results

The cases in which participants had to choose between a) fall and mid, b) fall and rise, and c) mid and rise pitch contours were analyzed separately in three subsets of 360 observations each. We found no statistically significant difference in any of the three data subsets on the basis of the three syllable sequences with /da/, /la/ and /na/; therefore, in the following analysis, the three syllable sets will be collapsed.

Glottalization was most frequently associated with falling pitch, but with a remarkable minority of non-fall responses (see Figure 2). When participants had to choose between fall and mid, fall responses were 65.8%, lower than the percentage of fall responses found by Bissiri & Zellers [22] for German (69.4%), English (73%) and Swedish (66.7%) listeners. Most Chinese listeners identified glottalization with falling pitch in over 50% of the cases; however, two of them gave a majority of mid responses and three gave 50% mid responses. When

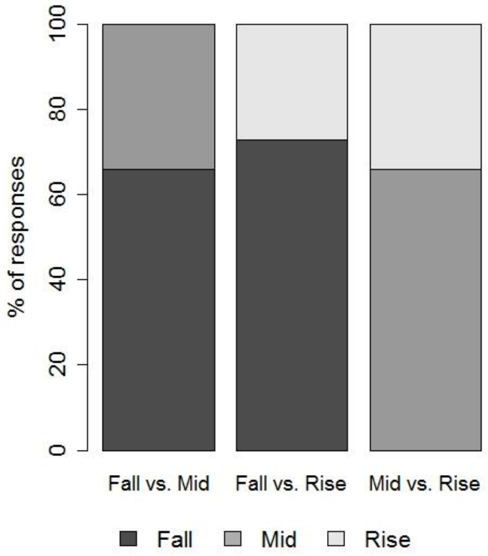

Figure 2: Responses by Chinese listeners.

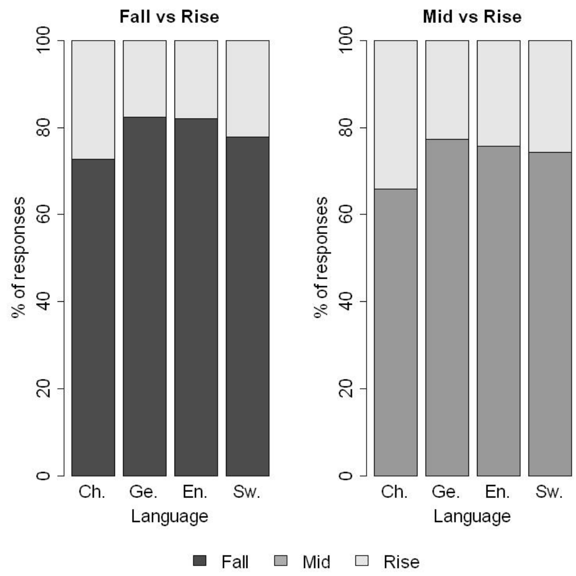

Figure 3: Responses by Chinese listeners and by Bissiri & Zellers’ [22] German, English and Swedish listeners when rise was one of the options.

rise was one of the options, there was a substantial minority of rise responses: 27.2% in the fall and rise subset, and 34.2% in the mid and rise subset. The Chinese listeners gave significantly more rise responses than Bissiri & Zellers’ [22] German, English and Swedish listeners: subset fall and rise χ^2(1, N = 1368) = 10.9798, $p < 0.001$, subset mid and rise χ^2(1, N = 1368) = 14.0703, $p < 0.001$ (see Figure 3). In the fall and rise subset, three Chinese participants had a majority of rise responses. Two of them had 50% rise responses in the mid and rise subset, while three other Chinese participants had a majority of rises.

For Chinese listeners, the initial pitch contour had no influence on the association of glottalization with a final pitch (subset fall and mid: χ^2 (2, N= 360) = 1.556, n.s.; fall and rise: χ^2 (2, N= 360) = 1.374, n.s.; mid and rise: χ^2 (2, N= 360) = 0.9633, n.s.). Figure 4 represents responses by Chinese across initial pitch when they had to choose between fall and mid, in comparison with Bissiri & Zellers’ [22] Swedish, German and English participants. German and English listeners had significantly more fall responses when initial pitch was rise, while Swedish listeners, similarly to the Chinese listeners in the present study, did not.

3. Discussion

As with the English, German and Swedish listeners in Bissiri & Zellers’ [22] analysis, the Mandarin Chinese listeners in the current study were most likely to associate glottalization with falling pitch rather than mid or rising pitch. However, to an even greater extent than Bissiri & Zellers’ listeners, the Mandarin listeners were flexible in their association of glottalization with different pitch contours, with around 30% of responses preferring a mid or rising contour even when a falling contour was also presented. Furthermore, there was no statistically significant effect of the initial pitch contour on the choice of final pitch contour to match the glottalization.

The lack of effect of initial pitch contour on Mandarin listeners’ judgments seems most likely to be the result of the Mandarin tone system, in which every syllable is assigned a lexical tone. In the case of the current stimuli, there was no reason for the Mandarin listeners to make predictions about the pitch of the final syllable on the basis of the first two unless they adopted a strategy of assigning lexical meanings that could constrain the last syllable. Six of the Chinese participants reported after the experiment that the stimuli sounded similar to their native language. However, the stimuli did not constitute existing words in Chinese, and we can therefore reasonably exclude the possibility that a mapping of some stimuli to real Chinese words introduced a bias in the responses.

It was possible that an influence of tone sandhi from the preceding pitch contour could constrain perception of the pitch of the glottalized syllable, but this does not appear to have been the case in the current data. There are several possible reasons for this. The first is of course that the pitch percept in the glottalized stretches was too strong to allow for alternative interpretations. However, the substantial minority of non-falling responses suggests that this is not the case, since if the pitch percept was extremely strong, it seems unlikely that those non-falling responses would be possible. Therefore, we must look elsewhere for an explanation.

One possibility is that glottalization may be so firmly associated with Tones 3 and 4 that listeners constrained by the glottalization even if the pitch contours did not match the canonical form. The higher percentage of rise responses by the Chinese compared to Bissiri & Z German, English and Swedish listeners might be explained by an association of glottalization with the falling-rising Tone 3. Liu & Samuel [24] report that Tone 3 is perceived by Mandarin Chinese natives with the same accuracy even if its rising portion is missing. They also found that listeners could still perceive Tone 3 most of the time even if it consisted only of its rising portion, which is surprising given its redundancy for Tone 3 perception. These results indicate a) a possible flexibility in tone perception by Mandarin Chinese listeners allowing them to associate glottalization with rising pitch, if they relate both acoustic cues to Tone 3, and b) a possible relevance of other secondary cues beside F0 for tone perception. Liu & Samuel [24] propose duration and amplitude

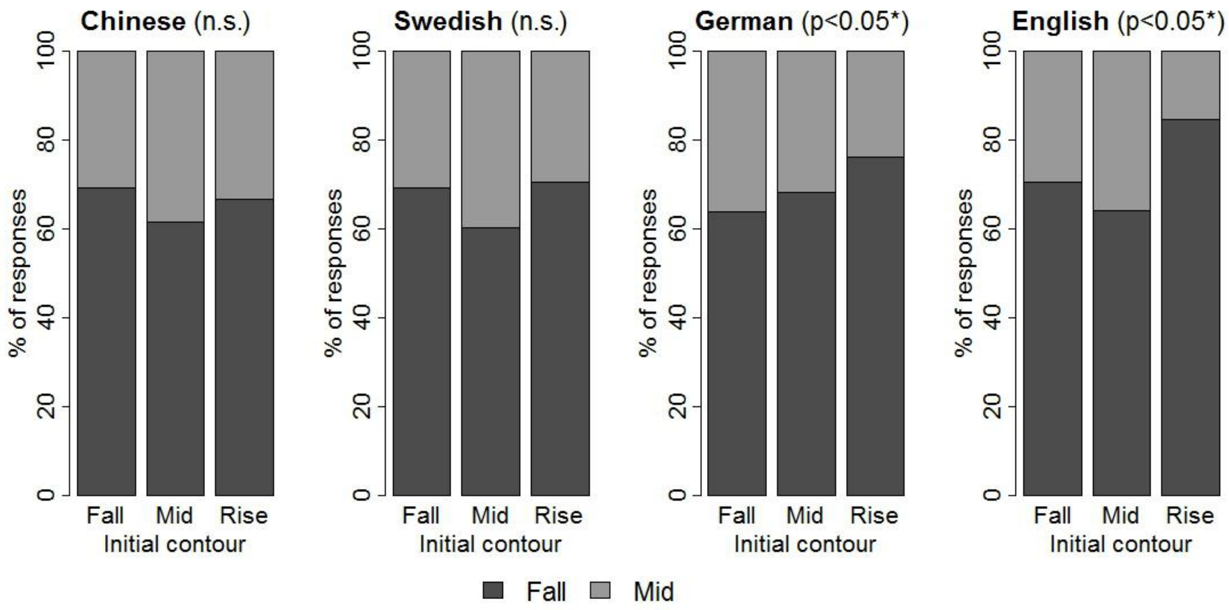

Figure 4: Responses by Chinese listeners and by Bissiri & Zellers' [22] Swedish, German and English listeners across initial pitch when choosing between fall and mid contours (Pearson's χ^2 test).

as possible candidates, and Lee et al. [25] also suggest voice quality or glottalization as a relevant perceptual cue for Mandarin tones.

Another alternative for the interpretation of our results is that listeners did not treat the syllables as Chinese lexical items at all. Since, as mentioned above, some participants treated the items as coming from a nonsense language, they could have potentially interpreted the pitch according to a tone-based system without being completely influenced by the form of the Mandarin tone system.

Combined with Bissiri & Zellers' [22] results, a consistent picture of the perception of glottalization begins to emerge from the data. Specifically, the different functional load of prosodic features across languages leads to different perceptual patterns. In Mandarin Chinese, where pitch is essential to lexical meaning, and every syllable has its own tone assigned, listeners were apparently not influenced by the pitch of preceding syllables but only made judgments about the pitch of the glottalized syllable itself. In Swedish, where pitch also contributes to lexical meaning but glottalization is not so firmly associated with either tone, listeners apparently were strongly inclined to seek a pitch interpretation of the stimuli, with most responses matching glottalization to the lowest available pitch contour in the pair. In German and English, where pitch is assigned on the intonational rather than a pitch-accent or tonal level, larger-scale contours appear to have a stronger influence on listeners' expectations, with listeners apparently making many of their pitch judgments on the basis of prediction of how a pitch contour would continue, rather than on calculation of the pitch of the glottalized stretch alone. It is important to note, however, that the current task intentionally influences listeners towards interpreting glottalization as having pitch, or being related to pitch variation. In other words, the task may have biased listeners towards a categorization of glottalization within the tonal/intonational system of their native language. A task which depended on a function of glottalization not (necessarily) tied to pitch variation –e.g. turn transition, as in Ogden [12, 13] –might lead to a different result.

4. Conclusion

Our data, combined with Bissiri & Zellers' findings, give growing evidence for a language-specific link between perception of pitch and perception of glottalization in long stretches of creaky voice. It appears that with tone or intonation languages can lead them to focus on different prosodic information when judging the pitch of glottalized stretches. In particular, the influence of the surrounding pitch context is different across different languages. Mandarin Chinese listeners, similarly to Bissiri & Zellers' Swedish listeners, were not influenced by the previous pitch contour while associating glottalized syllables to the possible pitch alternatives. Their performance contrasted with Bissiri & Zellers' German and English listeners, speakers of intonation languages, who were influenced by the previous pitch contour in their responses. Creak may lead Mandarin listeners to a falling-rising Tone 3 interpretation as well as a falling Tone 4 one, while the one-syllable size of the tone domain apparently influences them to ignore the preceding pitch context when making their judgments. Thus the different sizes of pitch domains relevant in each language, and even the internal phonological structure itself, appear to influence listeners' judgments about the for glottalization.

5. Acknowledgements

The authors are grateful to Xinping Xu for conducting the experiment, and to all of the Chinese participants. This research was supported by the European Union grant MC-IEF GeCzEnGlott, by the postdoctoral research of prosody in context (VR 435-2011-6871) from the Swedish Research Council (Vetenskapsrådet) and by the National Social Science Foundation of China (13BYY009).

6. References

[1] Gordon, M. & P. Ladefoged (2001). Phonation types: a cross-linguistic overview. Journalof Phonetics 29: 383-406.

[2] Gerratt, B.R. & J. Kreiman (2001). Toward a taxonomy of nonmodal phonation. Journal of Phonetics 29: 365-381.

[3] Catford, J.C. (1964). Phonation types: the classification of some laryngeal components of speech production. In: Abercrombie, D. et al. (eds.) In honour of Daniel Jones, London: Longmans, pp. 26-37.

[4] Blomgren, M., Y. Chen, M.L. Ng, & H.R. Gilbert (1998). Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers. Journal of the Acoustical Society of America 103(5): 2649-2658.

[5] Gussenhoven, C. (2004). The phonology of tone and intonation. Cambridge: Cambridge University Press.

[6] Pierrehumbert, J. & D. Talkin (1991). Lenition of /h/ and glottal stop. In Papers in Laboratory Phonology II. Cambridge: Cambridge University Press, 90-117.

[7] Pierrehumbert, J. (1995). Prosodic effects on glottal allophones. In: Fujimura, O., Hirano, M. (eds.), Vocal fold physiology: voice quality control. Singular Publishing Group, San Diego, pp. 39-60.

[8] Dilley, L., S. Shattuck-Hufnagel, & M. Ostendorf (1996). Glottalization of word-initial vowels as a function of prosodic structure. Journal of Phonetics 24: 423-444.

[9] Redi, L. & S. Shattuck-Hufnagel (2001). Variation in the realization of glottalization in normal speakers. Journal of Phonetics 29: 407-429.

[10] Henton, C. & A. Bladon (1988). Creak as a socio-phonetic marker. In Hyman, L.M. & C.N. Li (eds.) Language, Speech and Mind: studies in honor of Victoria A. Fromkin. London, pp. 3-29.

[11] Huffman, M.K. (2005). Segmental and prosodic effects on coda glottalization. Journal of Phonetics 33: 335-362.

[12] Ogden, R. (2001). Turn transition, creak and glottal stop in Finnish talk-in-interaction. Journal of the International Phonetic Association 31: 139-152.

[13] Ogden, R. (2004). Non-modal voice quality and turn-taking in Finnish. In Couper-Kuhlen, E & Ford, C. (eds.) Sound patterns in interaction: cross-linguistic studies from conversation. Amsterdam: John Benjamins, pp. 29-62.

[14] Bissiri, M. P., M.L. Lecumberri, M. Cooke & J. Volín, (2011). The role of word-initial glottal stops in recognizing English words. Proceedings of Interspeech 2011, Florence, Italy, pp. 165-168.

[15] Kohler, K. J. (1994). Glottal stops and glottalization in German. Phonetica 51: 38-51.

[16] Ding, H., O. Jokisch & R. Hoffmann (2004). Glottalization in inventory construction: a cross-language study. Proceedings of ISCSLP 2004, Hong Kong, pp. 37-40.

[17] Chao, Y.R. (1968). A Grammar of Spoken Chinese. Berkeley, University of California Press.

[18] Ding, H. & J. Helbig (1996). Sprecher- und kontextbedingte Varianz des dritten Vokaltones in chinesischen Silben – eine akustische Untersuchung. Proceedings of DAGA 1996, Bonn, Germany, pp. 514-515.

[19] Silverman, D. (1997). Laryngeal Complexity in Otomanguean Vowels. Phonology 14: 235-261.

[20] Frazier, M. (2008). The interaction of pitch and creaky voice: data from Yucatec Maya and cross-linguistic implications. UBC Working Papers in Linguistics: Proceedings of Workshop on Structure and Constituency in the Languages of the Americas (WSCLA), pp. 112-125.

[21] Ní Chasaide, A. & C. Gobl (2004). Voice quality and f0 in prosody: towards a holistic account. Proceedings of the 2nd International Conference on Speech Prosody, Nara, Japan, pp. 189-196.

[22] Bissiri, M.P. & M. Zellers (2013). Perception of glottalization in varying pitch contexts across languages. Proceedings of Interspeech 2013, Lyon, France, pp. 253-257.

[23] Boersma, P. & D. Weenink (2013). Praat: doing phonetics by computer [Computer program]. Available http://www.praat.org/.

[24] Liu, S. & A.G. Samuel (2004). Perception of Mandarin lexical tones when F0 information is neutralized. Language and Speech 47(2): 109-138.

[25] Lee, C.-Y., L. Tao & Z.S. Bond (2008). Identification of acoustically modified Mandarin tones by native listeners. Journal of Phonetics 36: 537-563.