

PERCEPTION OF PITCH IN GLOTTALIZATIONS OF VARYING DURATION BY GERMAN LISTENERS

Maria Paola Bissiri^{1,2}, Margaret Zellers³

¹ CASL Research Centre, Queen Margaret University, Edinburgh, Scotland, UK

² Institut für Akustik und Sprachkommunikation, Technische Universität Dresden, Germany

³ Department of Speech, Music & Hearing, Kungliga Tekniska Högskolan, Stockholm, Sweden
mbissiri@qmu.ac.uk, zellers@kth.se

ABSTRACT

Previous studies have shown that glottalization is not necessarily perceived as lower pitch but that pitch perception in glottalization can be influenced by the different size of prosodic domains relevant in the native language of the listener. Speakers of intonation languages were influenced by the preceding pitch context when judging the pitch of longer creaky voice stretches, while speakers of pitch-accent or tone languages were not.

The current study investigates pitch perception by German listeners in glottalized stretches of speech whose duration varied along a 10-step continuum. We found that the duration of the glottalized stretches affected the categorization of the stimuli, and that the German listeners were not influenced by the preceding pitch context, unlike in a previous study on longer stretches of glottalization of constant duration. Possibly shorter stretches of glottalization are interpreted as segmental word-boundary phenomena rather than as intonation.

Keywords: prosody, voice quality, perception, glottalization.

1. INTRODUCTION

The acoustic characteristics of glottalization, especially of creaky voice, consist in irregular and low frequency voicing, damping, lower amplitude and higher first formant values compared to modal voice [10, 11]. Because of its low frequency periods, glottalization has been most frequently associated with low pitch, but this association is not straightforward. In fact, glottalization can be associated with high tones in tone languages [11, 12] as well as in intonation languages [20, 19, 6, 21].

Studies by [1] and [2] investigated the perception of pitch in long stretches of utterance-final glottalization across different initial pitch contours by German, Swedish, English and Mandarin Chinese listeners. They found that listeners associated glottalization with final falling pitch in most cases, but in

a consistent minority of responses they associated it with final level or rising pitch. They also found that the initial pitch context influenced the German and English listeners, who gave more fall responses if initial pitch was rising, while it did not influence Swedish and Mandarin listeners' responses. The authors interpreted this to result from Swedish and Mandarin speakers being more sensitive to local pitch variation, the former because of the pitch accent contrast in Central Swedish [8, 9], the latter because of the lexical tones in Mandarin, while German and English listeners were possibly making predictions from larger (i.e. intonational) pitch domains. Moreover the Mandarin listeners in [2] associated glottalization with rising pitch more frequently than the German, Swedish and English listeners in [1]. Mandarin listeners may have interpreted creaky voice as a falling-rising Tone 3, since its pitch valley is frequently glottalized [5, 7] and Tone 3 remains perceptible if its rising portion is missing and even if it consists only of its rising portion [16]. Thus, it appears that language-specific phonological structure can influence pitch perception in glottalization.

Glottalization has also been frequently associated with prosodic boundaries in different languages. In English, glottalizations have been found at utterance ends [13, 21] and at prosodic boundaries [14, 20]. Glottalizations have also been found to characterize turn-transitions in Finnish [17, 18]. In Czech and in German glottalization of word-initial vowels is a frequent word-boundary marker [15, 3].

Given that the different sizes of pitch domains relevant in each language influence pitch perception in glottalization, the current study investigates whether glottalized stretches of different durations lead to different pitch interpretations, and if this is influenced by the preceding pitch context. Specifically, we focus on whether German listeners, who are accustomed to short stretches of glottalization as a word-juncture marker, will interpret short glottalized stretches as word boundaries instead of as intonation.

2. PERCEPTION EXPERIMENT

We employ the same AXB forced choice paradigm as in [1] and [2] to investigate whether a final syllable, with a duration of creaky voice varying along a 10-step continuum, is associated with falling, level or falling-rising pitch by German listeners.

We formulate the following hypotheses: A) syllables with longer creaky voice portions will more frequently be associated with falling pitch, syllables with shorter creaky voice portions with level pitch, and syllables with almost equal creaky and modal voice portions with falling-rising pitch; B) German listeners will tend to interpret word-initial glottalization as a segmental word-boundary phenomenon, rather than an intonation phenomenon to be associated to falling pitch, and this will lead to a bias towards level pitch responses; C) assuming that Hypothesis B is correct, listeners' responses will also be less influenced by initial pitch than in [1].

2.1. Methodology

2.1.1. Stimuli

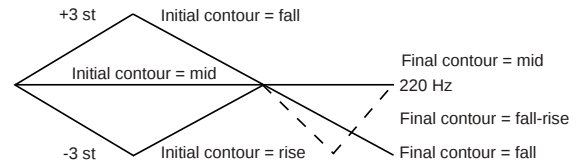
The stimuli employed in the current study were /da'da?a/ sequences, whose first two syllables were identical to those in the /da'dada/ stimuli by [1]; more details on their creation can be found there.

In order to create the vowel for the final /?a/ syllable, an /a/ starting with a creaky voice portion and continuing with modal voice was produced by the first author. The duration of the modal and of the creaky portion of the vowel were then increased by copy-editing single pitch-periods using Praat [4], resulting in a natural-sounding vowel of longer duration.

To simulate a glottal stop closure the /d/ closure at the onset of the final syllable in the /da'dada/ sequence by [1] was replaced with silence. After the silence, the final vowel in the sequence was replaced with a modal portion of 390 ms duration from the /a/ vowel recorded for the current study, starting from the beginning of modal voice. Although alveolar transitions were still present at the offset of the previous vowel, the final syllable in the resulting /da'da?a/ sequence had no audible alveolar artifacts. All stimuli were created by manipulating this base sequence.

Using a Praat script, nine pitch contours were created, beginning from a level pitch of 220Hz throughout the stimulus (see Fig. 1). Four pitch points were set 1) at the onset of the initial consonant, 2) at the onset of the second vowel, 3) at the onset and 4) at the end of the third vowel. Three initial pitch

Figure 1: Schematic of the nine pitch contours. The final fall-rise contour is dashed.



contours, fall, mid and rise, were created by either raising the second pitch point by 3 semitones (fall), lowering it by 3 semitones (rise) or keeping it level (mid). Three final pitch contours, fall, mid and fall-rise, were created by either lowering the fourth pitch point by 3 semitones (fall), keeping it level (mid) or adding a fifth pitch point at equal distance between the third and the fourth and lowering it by 3 semitones (fall-rise). These stimuli constituted the pitch set.

For each initial pitch condition (fall, mid and rise) a 10-step glottalization continuum was created. The final mid pitch condition constituted the first continuum step. In order to create the further 9 steps, a different 390 ms portion of the creaky-modal /a/ was spliced at the end of the stimuli for each step by shifting its starting point to the left of $1/9^{th}$ of the final vowel duration. Therefore the final vowel included larger portions of creaky voice at each step, being creaky throughout in step 10 (see Fig. 2).

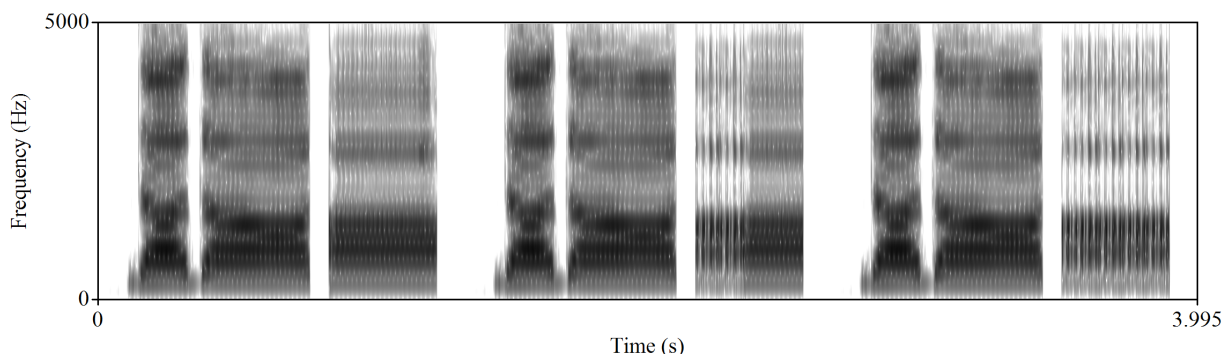
2.1.2. Experiment procedure

The participants were 42 native speakers of German who carried out the task at TU Dresden. They were offered 5 Euro for their participation.

An AXB perceptual experiment was carried out using Praat ExperimentMFC. The participants heard sequences of three stimuli and had to indicate if the middle stimulus was more similar to the first or the third. There were 126 stimulus sequences; of these, 36 were control comparisons in which all three stimuli belonged to the pitch set. In the 90 test comparisons the first and third stimuli belonged to the pitch set while the target stimulus was one of the continuum steps. Given the large number of stimuli due to the continuum, it was not possible to present all sequences both as AXB and BXA to each participant. For this reason two experiment versions with opposite order were prepared by means of a Perl script and alternated among the participants. All sequences were presented in random order by Praat.

Before the experiment the participants carried out a practice session with five /ba'ba?a/ control se-

Figure 2: Spectrogram of three /da'daʔa/ sequences constituting step 1, 5 and 10 of the glottalization continuum.



quences. Breaks were inserted after every 32 items. The participants took about 17-20 minutes to complete the task.

The responses to the control comparisons were checked to see if participants could correctly identify pitch patterns since the control sequences always presented two identical stimuli and a different one. Twelve participants who had not reached at least 90% correct responses in the control items were excluded from further analysis. If a participant had not reached the threshold, the same experiment version was employed for the next participant in order to keep the experimental order balanced.

2.2. Results

Participant responses when they had to choose between fall vs. mid, fall vs. fall-rise, and mid vs. fall-rise were analyzed in three separate subsets of 900 observations each by means of logistic regression (see Fig. 3). No significant differences were found between the three initial pitch conditions, therefore the data presented in the following analysis were modeled without initial pitch as a predictor.

For the case in which participants had to choose between fall and mid responses, a logistic regression model was used to fit the percentage of fall responses with the percentage of creaky voice in the target vowel as predictor. The number of fall responses significantly increased with larger creaky voice stretches in the target vowel ($p < 0.001$). Step 10 of the continuum, in which the whole final vowel was creaky, was associated to falling pitch in 54% of the cases.

In order to analyze the subset of responses in which participants had to associate the continuum stimuli to a falling or a falling-rising pitch contour, a quadratic logistic regression model was employed. The linear and the quadratic coefficients

were both highly significant ($p < 0.001$) showing that the choice of the quadratic model was appropriate. The number of fall-rise responses increases in the first half of the continuum, where the target vowel has almost equal portions of creaky and modal voice, and decreases in the second half of the continuum, where the vowel becomes predominantly creaky. Step 10 of the continuum with the fully creaky vowel was associated to final falling pitch in 63% of the cases.

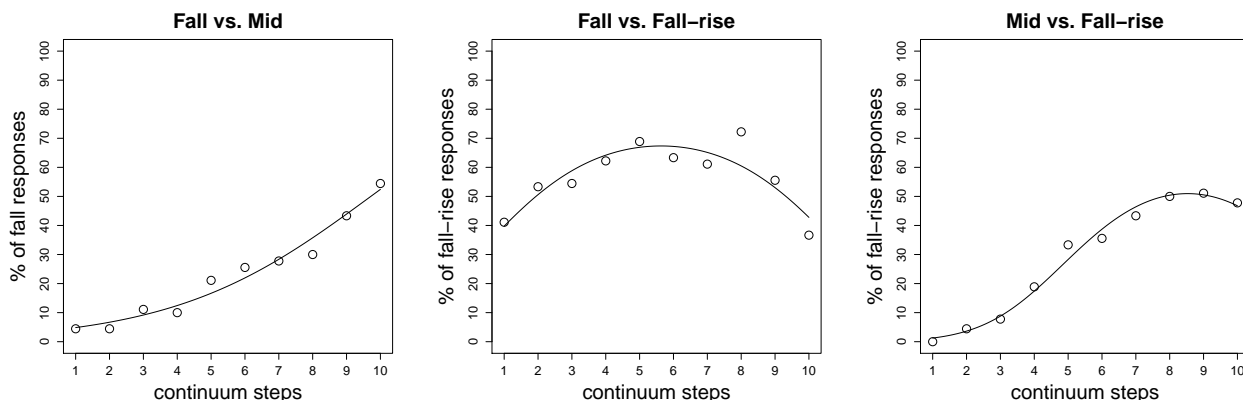
When the participants had to choose between a mid and a fall-rise pitch contour, their responses were also fitted by a quadratic logistic regression model. Again the linear and the quadratic coefficients were both highly significant ($p < 0.001$). The number of fall-rise responses increases gradually as the creaky portion of the vowel increases in the continuum, reaching a maximum at step 9 and slightly decreasing at step 10 (see Fig. 3).

3. DISCUSSION

The results of our study give support to the three hypotheses discussed in Section 2. As formulated in hypothesis A, an increasing duration of the creaky portion of the target vowel was associated with an increasing number of fall responses when participants had to compare falling and level pitch contours. Stimuli with shorter creaky portions were more frequently associated to level pitch in both the fall vs. mid and the mid vs. fall-rise subsets. In the fall vs. fall-rise subset most stimuli in the middle region of the continuum were associated to a fall-rise contour. The fitted logistic regression lines indicate a more gradient – rather than categorical – shift of the perception of pitch in the glottalized stretches.

Hypothesis B is also confirmed, given the preference for mid responses by the German listeners. When they had to choose between fall and mid they

Figure 3: Responses by German listeners when they had to choose between fall and mid, fall and fall-rise, and mid and fall-rise pitch with fitted logistic regression lines, where the predictor is the percentage of creaky voice in the target vowel varying along a 10-step continuum. Each circle represents 90 observations



gave a majority of mid responses for steps 1 to 9, and only 54% fall responses for the throughout creaky vowel in step 10, while in [1] they associated a final creaky syllable – /da/, /la/ or /na/ – to falling pitch in a larger number of cases, 69%. Possibly the larger number of mid responses in the current study was due to the interpretation of the glottalization as a word-boundary marker rather than as intonation. This would also explain why in this experiment, unlike in [1], initial pitch did not influence German listeners’ responses, thus confirming our hypothesis C.

The finding that there is a gradient shift between perceiving glottalization as a word boundary cue versus as a pitch cue is somewhat surprising, given that these two types of signaling would not appear at first glance to belong to the same prosodic system. This result is a good example of the complex interaction between different prosodic features which contribute to meaning. Listeners must constantly be on the lookout for possible information about many levels of linguistic structure simultaneously. Furthermore, speakers vary in the exact phonetic features of their productions, requiring listeners to adapt rapidly in order to stay on top of these multiple structures. Thus the gradient shift between the word-boundary interpretation and the intonation interpretation may represent either a struggle or a reluctance on the part of our listeners to commit to a single interpretation of middle cases. This is likely to be the best strategy for prosodic listening in conversation, since outside of the experimental setting other cues will usually be available to further guide or constrain possible interpretations.

We interpret our current results in light of two well-known functions of glottalization in German. However, as discussed in [3] not all languages use

glottalization for the same purposes. German in particular is known to use glottalization very reliably for word-boundary marking in the context of phonologically initial vowels [15]. Listeners from other languages may find short stretches of glottalization in such contexts merely disruptive, or they may be more inclined to interpret even short glottal stretches as low/falling pitch. We are currently conducting further studies across a number of other languages in order to determine the extent to which perception of these features varies between them.

4. CONCLUSION

The current results shed light on different modes of perception of glottalization which may be employed at different times by listeners. The German listeners in our study appeared to perceive glottalization as a word-boundary cue when it was of relatively short duration, but as contributing to pitch when it was of a relatively longer duration. As we predicted, the possibility of using glottalization as a word-boundary cue appeared to block the influence of preceding pitch on the perception of glottalization, indicating that although there is overlap between glottalization’s two functions, they remain separate. Future research will shed light on how different functions of glottalization may interact in other languages.

5. ACKNOWLEDGEMENTS

We thank Pier Giovanni Bissiri for his valuable comments on the statistical analysis. This research was supported by a Marie Curie IEF Fellowship within the 7th European Community Framework Programme, and by a Postdoctoral Grant from Vetenskapsrådet (the Swedish Research Council).

6. REFERENCES

- [1] Bissiri, M., Zellers, M. 2013. Perception of glottalization in varying pitch contexts across languages. *Proc. INTERSPEECH* Lyon. 253–257.
- [2] Bissiri, M., Zellers, M., Ding, H. 2014. Perception of glottalization in varying pitch contexts in Mandarin Chinese. *Proc. of Speech Prosody* Dublin. 633–637.
- [3] Bissiri, M. P., Lecumberri, M., Cooke, M., Volín, J. 2011. The role of word-initial glottal stops in recognizing English words. *Proc. INTERSPEECH* Florence. 165–168.
- [4] Boersma, P., Weenink, D. 2015. Praat: doing phonetics by computer [Computer program]. Available <http://www.praat.org/>.
- [5] Chao, Y. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.
- [6] Dilley, L., Shattuck-Hufnagel, S., Ostendorf, M. 1996. Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics* 24, 423–444.
- [7] Ding, H., Helbig, J. 1996. Sprecher- und kontextbedingte Varianz des dritten Vokaltones in chinesischen Silben: eine akustische Untersuchung. *Proc. of DAGA* Bonn. 514–515.
- [8] Fant, G., Kruckenberg, A. 1989. Preliminaries to the study of Swedish prose reading and reading style. *STL-QPSR* 30(2), 1–80.
- [9] Gårding, E. 1989. Intonation in Swedish. *Lund University Working Papers in Linguistics* 35, 63–88.
- [10] Gerratt, B., Kreiman, J. 2001. Toward a taxonomy of nonmodal phonation. *Journal of Phonetics* 29, 365–381.
- [11] Gordon, M., Ladefoged, P. 2001. Phonation types: a cross-linguistic overview. *Journal of Phonetics* 29, 383–406.
- [12] Gussenhoven, C. 2004. *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
- [13] Henton, C., Bladon, A. 1988. Creak as a sociophonetic marker. In: Hyman, L., Li, C., (eds), *Language, Speech and Mind: studies in honor of Victoria A. Fromkin*. London: Routledge 3–29.
- [14] Huffman, M. 2005. Segmental and prosodic effects on coda glottalization. *Journal of Phonetics* 33, 335–362.
- [15] Kohler, K. J. 1994. Glottal stops and glottalization in German. *Phonetica* 51, 38–51.
- [16] Liu, S., Samuel, A. 2004. Perception of Mandarin lexical tones when f0 information is neutralized. *Language and Speech* 47(2), 109–138.
- [17] Ogden, R. 2001. Turn transition, creak and glottal stop in Finnish talk-in-interaction. *Journal of the International Phonetic Association* 31, 139–152.
- [18] 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 29–62.
- [19] Pierrehumbert, J. 1995. Prosodic effects on glottal allophones. In: Fujimura, O., Hirano, M., (eds), *Vocal fold physiology: voice quality control*. San Diego: Singular Publishing Group 39–60.
- [20] Pierrehumbert, J., Talkin, D. 1992. Lenition of /h/ and glottal stop. In: *Papers in Laboratory Phonology II*. Cambridge: Cambridge University Press 90–117.
- [21] Redi, L., Shattuck-Hufnagel, S. 2001. Variation in the realization of glottalization in normal speakers. *Journal of Phonetics* 29, 407–429.