Perception of Glottalization in Varying Pitch Contexts Across Languages

Maria Paola Bissiri[1], *Margaret Zellers*[2]

[1] Institut für Akustik und Sprachkommunikation, Technische Universität Dresden, Germany
[2] Institutionen för Tal, Music & Hörsel, Kungliga Tekniska Högskolan, Stockholm, Sweden
Maria_Paola.Bissiri@tu-dresden.de, zellers@kth.se

Abstract

Glottalization is often associated with low pitch in intonation languages, but evidence from many languages indicates that this is not an obligatory association. We asked speakers of German, English and Swedish to compare glottalized stimuli with several pitch contour alternatives in an AXB listening test. Although the low F0 in the glottalized stimuli tended to be perceived as most similar to falling pitch contours, this was not always the case, indicating that pitch perception in glottalization cannot be predicted by F0 alone. We also found evidence for cross-linguistic differences in the degree of flexibility of pitch judgments in glottalized stretches of speech.

Index Terms: prosody, voice quality, perception, glottalization

1. Introduction

In intonation languages, creaky voice and other forms of glottalization have often been associated with prosodic boundaries. [1] and [2] find that in English glottalization is likely at utterance ends, while [3],[4] on English and [5] on Swedish show its presence at prosodic boundaries, and [6] notes that glottal productions are characteristic of turn transition locations in Finnish. Creak or glottalization have also been identified in other intonational locations. [7] found glottalization at the onset of phrase-initial vowels, while [8] found glottalization associated with pitch-accented words in English. These latter findings may be regarded as strengthening effects. In German and Czech the glottalization of word-initial vowels is a frequent word juncture marker ([9], [10]). In German, glottalization can also appear as a reduction phenomenon of supraglottal stop articulation ([9]).

The acoustic characteristics of glottalization have often led to it being associated, especially in intonation languages, with low pitch. Creakiness is associated across languages with lowered amplitude, positive spectral tilt, and higher first formant values than modal voice ([11]), and crucially, with low fundamental frequency (F0), damping, and aperiodicity ([12]). This low F0 has led to the classic characterization of creak/glottalization giving the impression of “a stick being run along a railing” [13: 32]. Although glottalization is acoustically a complex phenomenon (or set of phenomena), it can be identified with 95% accuracy by listeners ([14]) and can therefore be considered a robust tool for phonetic signaling.

As even a cursory review of the literature demonstrates, however, associating glottalization with low pitch simply because it normally involves low F0 is problematic. Many tone languages have creakiness associated with high tones ([11], [15]), and this may occur even in intonation languages ([2], [3], [7], [8]). E.g. [7] found that word-initial vowels were more frequently glottalized if the target words were pitch-accented. Since in their corpus L* accents were rare, they claim that glottalization was influenced by pitch-accents per se and not by low F0 [7: 439]. It has been rightly pointed out by [16] that studying pitch perception without investigating the influence of glottalization is an incomplete endeavor. [17] has suggested that tone and glottalization may be perceived sequentially rather than in combination, and [18] proposes that this may be because glottalization does not play a role in identifying tone at all.

Despite the common pairing of glottalization with pitch variation across languages, we do not know of any studies explicitly testing whether or how pitch is perceived in glottalization. The current research investigates whether there is a perceptual association between glottalization and pitch features for listeners from three native languages: German, English and Swedish.

2. Perception experiment

We used an AXB forced-choice paradigm to investigate whether listeners would consistently associate glottalization of a final syllable with a falling, level, or rising pitch contour in that position. The literature discussed previously leads us to two alternative hypotheses:

- Hypothesis A: glottalization, with low F0, will be associated with low/falling pitch in all cases
- Hypothesis B: glottalization will not be associated with any pitch percept

A production experiment by [19] on Southern British English found that longer stretches of glottalization were found with pitch contours that were uniform in their direction. Therefore, a third hypothesis is possible:

- Hypothesis C: glottalization will be associated with pitch contours moving in the same direction as the preceding pitch

2.1. Methodology

2.1.1. Stimulus creation

Stimuli were based on sequences of syllables: /da'dada/, /la'lala/, and /na'nana/. These sequences were originally produced naturally by the first author, a native speaker of Italian, in both modal and glottalized versions. For each sequence, a modal version was chosen, and the vowel durations were manipulated in Praat [20] so that all of the stimuli had the same vowel lengths (i.e. all first syllables were the same length, all second syllables were the same length, etc.). The overall durations of the stimuli were also compressed to 85% of the original length, after impressionistic listening suggested that the original versions were too slow to sound natural.

Using a Praat script, twelve versions of each syllable sequence were created. The first nine were the Pitch conditions. The initial pitch was set at 220Hz, and four pitch points were set: at the onsets of the first consonant, of the

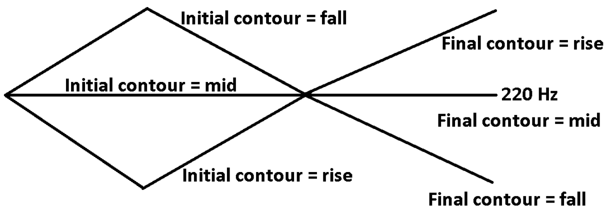

Figure 1: Schematic of possible pitch contours.

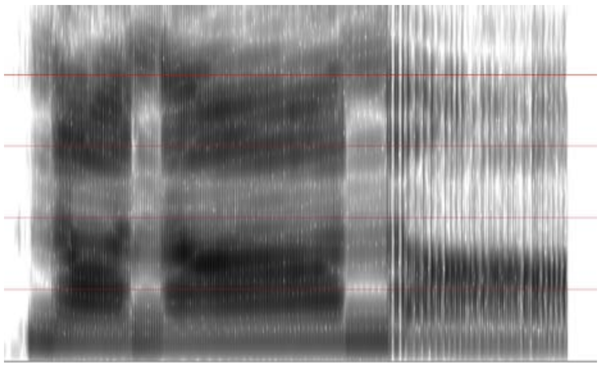

Figure 2: Spectrogram of /na/ sequence with glottalized final syllable.

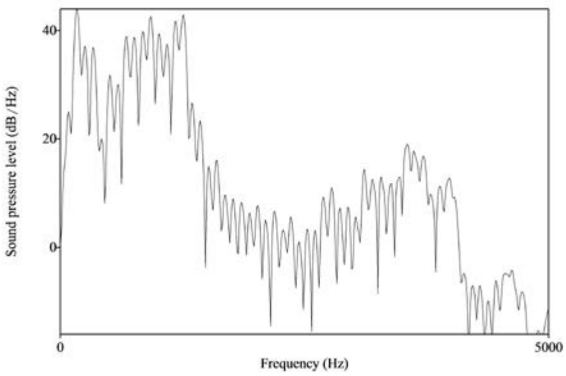

Figure 3: Spectral slice taken from glottalized stretch.

middle and final vowel, and at the end of the final vowel. The second pitch point (onset of stressed vowel) was either kept level (mid), raised 3 semitones above the beginning pitch (fall), or lowered 3 semitones below the beginning pitch (rise). Similarly, the fourth pitch point (end of final vowel) was either kept level, raised 3st, or lowered 3st. Since the third pitch point (onset of the final vowel) was held constant, this created 9 pitch contours, as illustrated in Figure 1.

For each initial condition (mid, rising, or falling), a further version was created with the final syllable containing glottalization rather than a smooth pitch contour; these constitute the Glottalization set. The glottalization was taken from tokens produced by the first author, and was matched in length to the final syllables of the pitch cases. The /da/ and /la/ sets used the same glottal stretch, taken from a /la/ sequence; the /na/ set had its glottalization taken from a /na/ sequence as otherwise the mismatch between nasalization in the vowels was apparent.

The glottal stretches were spliced into the stimuli following the formant transitions out of the final consonant in order to make sure the consonant remained clear. To create the impression of glottalization in the transitions, four additional pitch points were added using a Praat script. These were equally spaced through the duration of the formant transitions and had values of 35Hz, 220Hz, 35Hz, and 220Hz respectively. This procedure was suggested by the procedure used by [18] to create synthetic creak in her stimuli. An amplitude reduction was then applied to the entire glottalized portion of the stimuli in order to improve naturalness. A spectrogram of a /na/ sequence with glottalization at the end is shown in Figure 2, with the spectrum of the glottal stretch shown in Figure 3. Finally, all stimuli (Pitch and Glottalization) were amplitude-normalized.

2.1.2. Participants

Participants in the experiment were native speakers of German (N=35), English (N=13), and Swedish (N=14). Participants who completed the task at TU Dresden were offered €5 for their time; participants at KTH were not compensated. In addition to testing at these two locations, 7 people also completed the experiment remotely, with instructions to sit in a quiet place and to use headphones.

2.1.3. Experiment procedure

The experiment was presented using Praat's ExperimentMFC. Participants heard sequences of three stimuli, and were asked to compare the middle stimulus to the first and third, and to decide which it sounded most like. An AXB paradigm was preferable to ABX since both options are adjacent to the target stimulus, putting fewer demands on memory. AXB sets were drawn from the same syllable set (/da/, /la/, or /na/), and also all had the same initial pitch contour. A and B were always stimuli from the Pitch set, while X in the target items was from the Glottalization set. In order to address the tendency in AXB tests for participants to choose the B option (since it is more recent and thus better remembered), participants heard each critical pitch contour comparison in both AXB and BXA format. This resulted in a set of 54 test comparisons. To these were added 72 control comparisons, in which all three items came from the Pitch set, so each participant heard 126 stimuli in total. The order of presentation was automatically randomized for each participant by Praat.

Before beginning the task, participants heard 3 practice items, which were all control items with the syllables /ba'baba/. Following feedback from a pilot participant, the test was set up so that participants could take a break every 32 items if they wanted. Participants took between 17-20 minutes to complete the task.

2.1.4. Data cleanup

Before the data was analyzed, the participants' responses to the control items were checked for correctness in matching patterns. Since the control items always involved two identical items and a third that was different, participants who gave incorrect responses in more than 10% of the control items were excluded from the analysis. Of the participants, 5 German listeners and one Swedish listener were excluded on this basis. All participants whose results were kept had an accuracy rate of 95% or higher.

2.2. Results

No statistically significant difference was found on the basis of the different syllable sets; in what follows, stimuli from the three sets will thus be treated together.

For all listeners, there was a strong preference for glottalization to be identified with a falling pitch contour. This is unsurprising given the fact that the stretch of glottalization had low F0 and was easily long enough for listeners to be able to make a pitch analysis. However, although low pitch was strongly preferred by listeners, it was not always chosen. Figure 4 illustrates the pitch contours chosen by listeners (in cases when a falling contour was one of the options; cases in which the only options were a mid and a rising contour are excluded from the current analysis).

When a final falling contour was an option, listeners chose it as most similar to the glottalized stimulus in 75.4% of cases, with no statistically significant differences across languages (X-squared (4, N= 2016) = 4.5253, n.s.). Final mid contours were chosen in 15.2% of cases, while rising contours were chosen in 9.4%. When no falling contour was given as an option, participants chose the final mid contour in 76.2% of cases, and the final rising contour in the remaining 23.8%.

The initial pitch contour had an influence on listeners' decisions about which final contour the glottalization sounded most like. When a rising initial contour was heard, "fall" responses increased and "mid" responses decreased. However, this effect varied across languages. In German (X-squared (2, N = 540) = 6.4931, $p < 0.05$) and English (X-squared (2, N = 234) = 8.7318, $p < 0.05$), the proportion of falls chosen compared to mid contours increases when the initial contour is a rise. However, in Swedish (X-squared (2, N= 234) = 2.1923, n.s.), there is no change in the proportion of fall to mid contours chosen. The different distributions are shown in Table 1 and Figure 5 (overleaf).

English and Swedish natives show more fall responses in the case of falling initial contour than in the case of mid initial

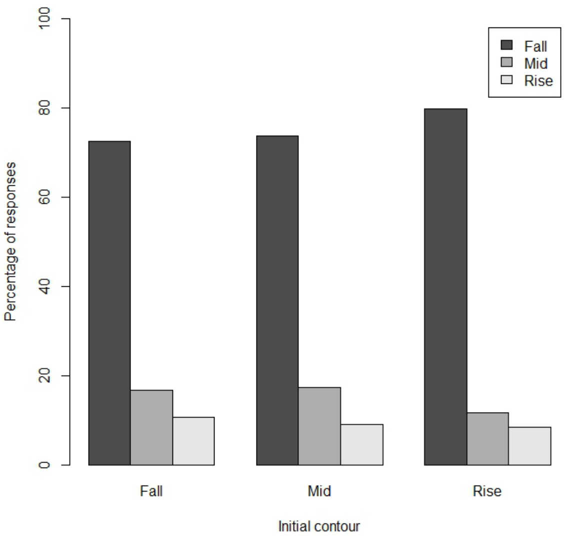

Figure 4: Responses by all listeners when a falling pitch contour was one of the options presented.

contour, while it is the opposite for German natives (cf. Figure 5). However, this trend was not statistically significant.

Table 1: Percentage of fall and mid responses across languages in different initial pitch contour conditions.

German	**Initial fall**	**Initial mid**	**Initial rise**
Glottalization = fall	63.9%	68.3%	76.1%
Glottalization = mid	36.1%	31.7%	23.9%
English			
Glottalization = fall	70.5%	64.1%	84.6%
Glottalization = mid	29.5%	35.9%	15.4%
Swedish			
Glottalization = fall	69.2%	60.3%	70.5%
Glottalization = mid	30.8%	39.7%	29.5%

3. Discussion

The findings reported above primarily support our Hypothesis A, that glottalization, at least in stretches long enough for the listener to perform pitch analysis, is associated with falling pitch by native speakers of the three languages tested. Given the low F0 and long stretch of glottalization, this finding is not very surprising. However, we also find an influence of the initial pitch contour on listeners' interpretation of the final glottalization, with initial rising pitch contours more likely to lead to an interpretation of the final glottalization as falling pitch. When the three languages are taken separately, this effect attains statistical significance for English and German, but not for Swedish listeners.

Research on pitch perception across languages suggests that listeners whose native language has lexical tone contrasts are more sensitive to pitch variation in speech stimuli than listeners from non-tonal languages (e.g. [21], [22]). Amongst other findings, [21] report a difference in perception between speakers of English and speakers of Swedish. The current data add another dimension to these findings: different language backgrounds appear to lead to different perception of glottalization in varying pitch contexts.

One reason for the cross-linguistic differences reported here may be that Swedish listeners, who are reportedly more sensitive to pitch variation in perception tasks [21], may have been using the available information differently, or to a different degree, than the German and English listeners. The pitch context did not have an effect on their decisions. Since pitch variation is more meaningful to Swedish listeners due to the Swedish pitch accent system, which is used for lexical contrasts (cf. [5]), Swedish listeners may be more inclined to gather as much pitch information as possible rather than making strong predictions about what will follow.

Conversely, for German and English listeners, the pitch context had much more of an effect on their decision-making.

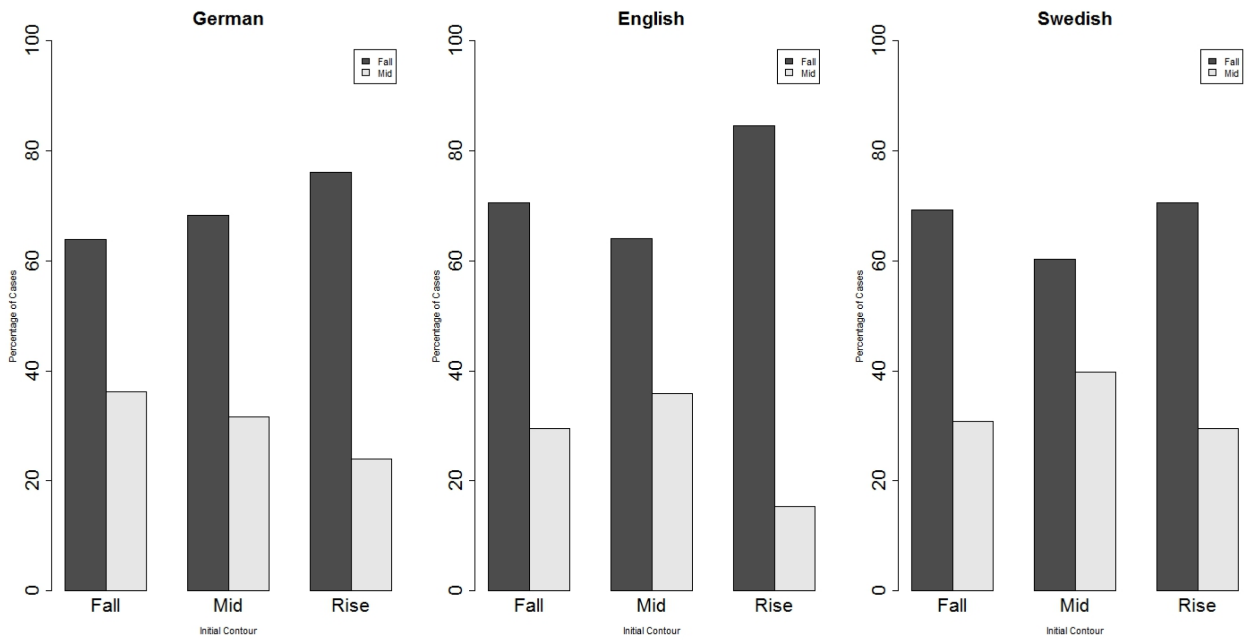

Figure 4: Distribution of fall and mid responses following different initial pitch contours, in German, English and Swedish.

We propose that this is because German and English listeners made stronger predictions about the upcoming pitch. Falling contours would lead to an expectation for a continued pitch fall, and taking declination into account, so could level mid contours. However, a rising contour would lead to an expectation of continued rising pitch. This would make the disruption of the pitch contour with glottalization much more salient than in the cases when it followed level or low pitch. This greater degree of saliency could then lead to the increased judgments that glottalization sounded like falling pitch following a rising contour. Some German participants reported informally that they tried to assign meanings such as “question” or “statement” to the stimuli as they listened. This could be taken as support for the idea that the German listeners, at least, were making predictions about what they were hearing in the stimuli, although participants were not systematically asked about their strategies so it is not clear how many may have done so.

The fact that listeners were able to choose mid or rising pitch as a match for glottalization even in cases when low pitch was available makes it clear that for German, English and Swedish listeners, there is more to the perception of glottalization than the simple interpretation of pitch. These cases cannot be simply treated as errors, since the listeners retained in the analysis were extremely accurate in matching the pitch contours they heard in the control items. Although the F0 information from the glottalization was clearly available to listeners, leading them to match glottalization with low/falling F0 in the majority of cases, the minority in which this did not happen was a substantial one. This lends support for the idea that glottalization need not be perceived as pitch information, even if it can be. Additional research is necessary to identify what the specific alternatives may be. One possibility is that glottalization is motivated directly by different aspects of the linguistic structure than pitch. [23] and [6] (among others) find extended creaky stretches associated with turn transition, raising the possibility that this phonetic aspect may be tied to interactional structure rather than the pragmatic structures typically associated with intonation. However, this is only one of many possibilities which should be investigated in the future.

4. Conclusions

The data reported here serve as evidence that although F0 within glottalization can be perceived as pitch, it need not be, and that linguistic experience can influence the degree to which listeners listen for pitch information or respond to glottalization in some other way. Specifically, Swedish listeners, whose experience with lexical pitch accents may make them more sensitive to pitch in general, were most likely to respond in a way consistent with the actual F0 of the glottalized stimuli, while German and English listeners appeared to be more flexible in their interpretations of glottalization, possibly giving more weight to predictions about the continuation of the pitch contour. Further research is necessary to give a clearer characterization of how listeners of different languages use the available acoustic information to come to a complete perceptual impression of glottalized stretches.

5. Acknowledgements

The authors are grateful to Oliver Niebuhr for early discussions on the development of this study, to Ivan Kraljevski for his support on participant recruiting, and to all participants. This research was supported by the European Union grant MC-IEF GeCzEnGlott and by the postdoctoral research grant ‘Perception of prosody in linguistic contexts’ from the Swedish Science Foundation (Vetenskapsrådet).

6. References

[1] 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. 3-29.

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

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

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

[5] Fant, G., & A. Kruckenberg (1989). Preliminaries to the study of Swedish prose reading and reading style. STL-QPSR, 30(2): 1-80.

[6] Ogden, R. (2003). Voice quality as a resource for the management of turn-taking in Finnish talk-in-interaction. In Proceedings of 15th International Congress of Phonetic Sciences, Barcelona, Spain.

[7] 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.

[8] 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.

[9] Kohler, K. J., “Glottal stops and glottalization in German”, Phonetica 51: 38-51, 1994.

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

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

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

[13] 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.

[14] 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.

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

[16] Ní Chasaide, A. & Gobl, C. (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.

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

[18] 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), 112-125.

[19] Zellers, M. & B. Post (2010). Aperiodicity at topic structure boundaries. In Proceedings of Speech Prosody 2010, Chicago, USA.

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

[21] Burnham, D., Francis, E., Webster, D., Luksaneeyanawin, S., Attapaiboon, C., Lacerda, F. & Keller, P. (1996) Perception of lexical tone across languages: evidence for a linguistic mode of processing. Proceedings of ICSLP 1996, 2514-2517.

[22] Burnham, D. & Francis, E. (1995) The role of linguistic experience in the perception of Thai tones. In Thongkum, T. (Ed.) South East Asian linguistic studies in honour of Vichin Panupong (Science of Language Vol. 8). Bangkok, Chulalongkorn UP.

[23] Local, J. & Kelly, J. (1986) Projections and silences: notes on phonetic and conversational structure. Human Studies 9: 185-204.