The Role of Word-Initial Glottal Stops in Recognizing English Words

Maria Paola Bissiri [1], *Maria Luisa Lecumberri*[2], *Martin Cooke*[3], *Jan Volín*[1]

[1] Institute of Phonetics, Charles University in Prague, Czech Republic
[2] Department of English Philology, University of the Basque Country, Vitoria, Spain
[3] Ikerbasque (Basque Science Foundation), Spain

jan.volin@ff.cuni.cz

Abstract

English word-initial vowels in natural continuous speech are optionally preceded by glottal stops or functionally equivalent glottalizations. It may be claimed that these glottal elements disturb the smooth flow of speech. However, they clearly mark word boundaries, which may potentially facilitate speech processing in the brain of the listener. The present study utilizes the word-monitoring paradigm to determine whether listeners react faster to words with or without glottalizations. Three groups of subjects were compared: Czech and Spanish learners of English and native English speakers. The results indicate that perceptual use of glottalization for word segmentation is not entirely governed by universal rules and reflects the mother tongue of the listener as well as the status (L1/L2) of the target language.

Index Terms: foreign accent, second language, glottalization, reaction times, word-initial vowels

1. Introduction

The speech signal contains a lot of information that cannot be described as phonemic, yet its contribution to effective speech communication often proves quite valuable. The presence or absence of certain inconspicuous features can in realistic conditions influence the effort exerted by the human brain to decode linguistic messages. Despite the traditional emphasis of segmental phonetics on supraglottally produced phones, some authors highlight the importance of glottal activity in the analysis of segmental chains of speech sounds. The most obvious phenomenon in this respect is the glottal stop, canonically articulated by a brief tight closure of the glottis with a subsequent abrupt release of the air trapped below it. This articulatory manoeuvre, however, requires a relatively great deal of effort and is often replaced by a short period of creaky phonation (or vocal fry) with same perceptual effect on the listener [1: 75]. For the purpose of this study and in line with common practice, we will call glottal stops and perceptually equivalent glottal gestures *glottalizations* ([2: 408], although *cf.* [3], or [4]: 455).

In the world's languages, glottalizations offer themselves for different purposes, as discussed, for example in [5]. The three most common categories of use are (1) phonemic contrast, (2) prosodic structure marking, and (3) signalling affect or emphasis. In the present paper, we concentrate on the second category, i.e., the possibility of using glottal elements as boundary markers between higher than segmental units. Specifically, it is the word boundary (or possibly autosemantic morpheme boundary) with a unit-initial vowel involved. For example, the phrase *Run out* can be pronounced without [rʌnˈaut] or with [rʌnˈʔaut] glottalization, with the latter option explicitly marking the word boundary as well as sounding more emphatic in most variants of English (cf. e.g., [6: 155; 7: 345]).

It was shown in [8] that the occurrence of glottalizations before word-initial vowels was correlated with the prosodic structure of the utterance. The deeper the prosodic break, the greater was the probability of glottalization. Similarly, metrically stronger syllables were glottalized more often than weaker ones. An extended look into the matter was presented in [2], together with other, more speaker-bound sources of variation. This variation was considered for the English language only.

However, the presence of glottalizations varies language specifically not only in terms of its usage but also of its frequency (cf. [9] and [10]). As a consequence, learners of a foreign language may encounter difficulties either because of an inappropriate number and/or use of glottal stops in their productions or in their perceptions when their expectations diverge from the actual usage in the target language. It was demonstrated in [11] that Czech speakers of English use substantially larger numbers of glottal stops before word-initial vowels than native speakers. This fact seems to contribute notably to the specific perceptual impression of the Czech accent of English.

Apart from production effects, it is reasonable to hypothesize that regular occurrence of glottalizations before word-initial vowels in one's mother tongue may lead to some sort of reliance on this boundary signal in speech perception of the second or foreign language. Therefore, we used the word monitoring paradigm [12], in which the subject is presented with a target word and then hears a carrier phrase containing that word. The subject's task is to respond as rapidly as possible when he or she detects the target word in the speech material heard. Reaction times are measured to explore the ease with which speech is processed in the brain of the listener.

The choice of three groups of subjects – Czech, Spanish, and native British speakers of English – provides an opportunity to explore the differences between native and non-native reliance on word-initial glottalization as well as the difference between the Czech, who use glottal stops frequently as a regular juncture marker, the Spanish, who use it relatively infrequently and mainly as an emphatic marker, and the native English, who use it both as a boundary and emphatic marker but not as frequently as the Czech.

Presenting these three groups of listeners with English phrases in which the target words start with a vowel and either are or are not preceded by a glottal element addresses our hypothesis that Czech listeners should benefit from glottalizations more than the English, while Spanish listeners should find the presence of the glottalization least beneficial. Counter to this hypothesis, though, there is a hypothesis, that the English listeners, due to lower processing costs of listening to their native language, will have other cues at their disposal

compared to the non-native listeners, to whom the additional cues may not be available (see, e.g., [13]). Because of that, the native English listeners might be able to utilize the possible facilitating influence of word-boundary glotallization better than the non-native groups.

2. Method

2.1. Stimuli

The phrases that carried target words with word-initial vowels were obtained from a corpus of BBC news bulletins read by five female and three male non-professional British English speakers. Forty-eight stretches of speech were selected with the aim eliminating the predictability of the target word by means of the context. Half of the items occurred naturally with a glottalization while the other half lacked glottalization.

For each of the 48 stretches of speech a counterpart was created by either removing a glottalization from the onset of the target words or inserting it where it naturally did not occur. Thus, a set of 48 additional stimuli was created. The inserted glottalizations were natural ones taken from utterances by the same speaker in the same segmental context. The intensity trimming and F0 smoothing by means of PSOLA as implemented in Praat [14, 15] at the cutting points was necessary to obtain natural-sounding stimuli. Small abrupt F0 excursions and amplitude steps were corrected manually by linear interpolation and inspected perceptually. All the manipulated stimuli were again tested for their absence of artifacts by one native and two proficient non-native speakers of English (the authors of the present paper).

The items to which a glottalization or glottal stop was added were on average 76.96 ms longer than their natural counterparts (*s.d.* = 6.77), and those from which the glottalization was removed were 81.04 ms shorter (*s.d.* = 10.84). A Welch two-sample t-test indicated that this difference was not significant ($p = 0.126$).

Thirty-six distractors (fillers) were incorporated into the test: a) 12 stimuli with an early target, in which the target word, starting with a consonant, occurred with the first or second syllable of the phrase, b) 12 stimuli with zero target, in which the target word displayed on the screen never occurred, and c) 12 stimuli starting with a consonant placed in a later part of the phrase. In order to prevent priming, we made sure that the distractor items did not repeat the target words. Six practice items were also prepared, including all the possible stimulus types: two with a target word starting with a vowel, two with target words starting with a consonant, one with an early target, and one with a zero target.

All stimuli were set to the same mean root-mean-square energy level and flanked with a signal to indicate the beginning of each phrase (two unobtrusive beeps) and a non-speech desensitization sound of two seconds, comprising a portion of white noise followed by gliding sinewave tones and approximately 670 ms of silence. The desensitization sounds help to eliminate the influence of one test item onto the following one and refresh the ability of the listener to hear a speech item independently of what was played before.

2.2. Participants

The Czech participants of the perception experiment were 32 students of the Faculty of Philosophy at Charles University in Prague, aged 19-29. The Spanish participants were 37 students at the Department of English Philology of the University of the Basque Country in Vitoria, Spain, aged 19-30. Twenty-two of the Spanish natives were bilingual in Spanish and Basque, which should not affect the present study since the usage of glottal stops in Spanish and Basque seem to be very similar and, to our knowledge, no differences have been reported in the literature. The British English participants were 31 students and employees of the University of Bristol, aged 19-32.

The Spanish and British subjects were paid for their participation, while the Czech participants received credits for a university course.

2.3. Testing procedure

The perception experiment was implemented with the DMDX software [16] on a Lenovo ThinkPad SL510 laptop. Stimuli were randomly distributed in four blocks (A, B, A', B') so that each block contained 24 stimuli with target items, equally subdivided between manipulated and non-manipulated items and with and without glottalization, and nine distractors (fillers). Blocks A' and B' contained the counterparts of the stimuli in A and B respectively, e.g. a manipulated stimulus with glottalizaton in A had its non-manipulated non-glottalized counterpart in A'. Given the condition that a block of stimuli should not be immediately followed by the block containing the counterparts (e.g. block A should not be followed by A'), eight block orders were possible, which were alternated among test participants. The order of the stimuli was randomized in each block by the testing software.

Unlike the paradigm employed in [12], the target word presented in orthography on the screen was displayed not only before, but also during the interval when the phrase was played. This way subjects did not need to memorize it, which we considered easier for non-native speakers of English. Two seconds after the appearance of the target word, the signal indicating the beginning of the phrase was played, followed by 300 ms silence and the actual test item. The desensitization sound closed the sequence and after a 1.15 seconds delay the following target word was displayed on the screen.

The stimuli were presented through Sennheiser HD201 headphones in sound-proof booths in the above-mentioned academic settings in Bristol, Prague and Vitoria. Before the test, subjects listened to the six practice items and adjusted the volume to a comfortable level. The instructions informed the subject about the testing procedure. The non-native subjects were also informed they should not worry if they did not know any given target word, as their competence in English was not tested, and they would not be asked any comprehension questions later.

The duration of the pauses between blocks of the test was at the liberty of the subjects, but a couple of minutes were suggested. The test lasted about 40 minutes, including the instruction, the practice and the pauses between blocks.

3. Results

3.1. Invalid responses

In line with common practice, three types of responses were labelled as *invalid*: a) no response (miss), when the subject failed to press the key, b) false alarms, when the subject pressed the key too early (reaction time < 150 ms), and c) hesitations, when the subject pressed the key too late (reaction time > 1000 ms).

Out of the total of 9600 responses (100 listeners × 96 target items per test), 12.1% were found to be invalid. These invalid responses were not shared equally by the three language

groups. Figure 1 hints that the native speakers of English contributed with only about 15 %, while the Spanish speakers produced more than a half of all the invalid responses.

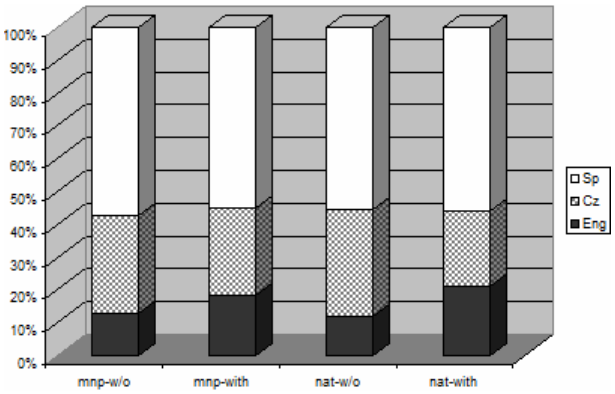

Figure 1. *Relative contribution of the three language groups to the total number of invalid responses by manipulated (mnp) and natural (nat) items and non-glottalized (w/o) vs. glottalized (with) items.*

A different aspect of these results is summarized in Tables 1 and 2 where the numbers of invalid responses are expressed relative to the total of responses in the category. There is a tendency towards greater numbers of invalid responses in manipulated items (glottalizations added or removed). We discuss this finding together with other results below in Section 4.

Table 1. *Percentages of invalid responses for manipulated vs. natural stimuli across Czech, Spanish, and English subjects (counts in brackets).*

Stimulus type	Czech	Spanish	English
manipulated	11.7 % (179)	20.0 % (356)	7.0 % (104)
natural	10.2 % (157)	16.6 % (294)	5.1 % (76)

Table 2. *Percentages of invalid responses for non-glottalized vs. glottalized stimuli across Czech, Spanish, and English subjects (counts in brackets).*

Stimulus type	Czech	Spanish	English
non-glottalized	12.8 % (196)	19.7 % (350)	5.1 % (76)
glottalized	9.1 % (140)	16.9 % (300)	7.0 % (104)

The numbers of invalid responses were greater for non-glottalized items in the case of Spanish and Czech listeners, while the native speakers of English produced the opposite result. Although the differences do not appear large, they were all highly significant (McNemar's chi-squared test for matched items). One has to be cautious since chi-square tests produce significance even for small effects if the sample is extensive.

3.2. Reaction times

Word-monitoring reaction times for individual language groups of the listeners are displayed in Table 3. It is clear that the native speakers of English were the fastest and that glottalization before the word-initial vowel led to shorter reaction times in all three groups of listeners.

Table 3. *Mean reaction times in milliseconds across the mother tongue of the respondents with respect to the presence or absence of glottalization.*

Lang.	Glott.	Mean RT (ms)	St. err.	No. of resp.
English	yes	371,0	4.6	1488
	no	417,2	5.0	1488
Czech	yes	413,2	4.5	1536
	no	453,8	4.9	1536
Spanish	yes	430,7	4.2	1776
	no	457,8	4.6	1776

Mixed-design ANOVA was run on the collected data with glottalization as the within-group variable and language as the between-group variable. The glottalization factor had two levels: the presence or absence of word-initial glottal element. The language factor refers to the mother tongue of a respondent in the perception test, therefore it has three levels. Both the main effect of the language and the main effect of glottalization were significant: $F(2, 4797) = 50.99$, $p < 0.001$, and $F(1, 4797) = 134.7$, $p < 0.001$. Post-hoc Tukey LSD tests showed that in all the three language group the difference in reaction times to glottalized and non-glottalized items was significant at the level of $\alpha = 0.001$.

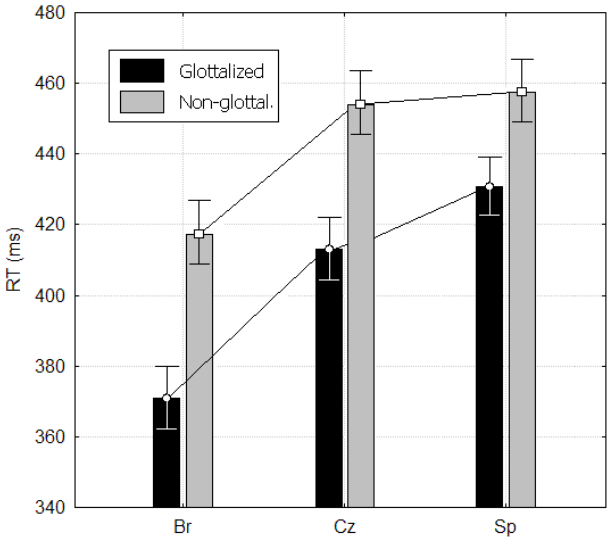

Figure 2. *Mean reaction times with 95-percent confidence intervals to words with and without glottalized word-initial vowels across British, Czech and Spanish groups of listeners*

The interaction between the two factors also proved to be significant, although not as highly as the two main effects: $F(2, 4797) = 3.11$, $p = 0.045$. Lines in Figure 2 depict the interaction. They indicate that the Spanish listeners were main contributors in that to them the presence or the absence of the glottalization before the word initial vowel made smaller difference than in the Czech and the British group, as predicted by our hypotheses in Section 1. (Relative falls in reaction times for non-glottalized items were approximately 11% for the British, 9% for the Czech and 6% for the Spanish.)

Two-way ANOVA was also performed on reaction times with respect to the manipulation factor. Reactions to manipulated items were significantly slower than to natural items: $F(1, 9308) = 39.3$, $p < 0.001$. Interaction between

manipulation and glottalization factor was also significant: $F(1, 9308) = 11.7$, $p < 0.001$. Post-hoc tests revealed that manipulation actually did not matter when glottalization was added, but when it was eliminated from the signal, the reaction times increased significantly. The average increase was 36 ms.

4. Discussion and Conclusions

The results of the experiment agree with our original idea that (i) native processing would be superior to non-native perception and (ii) that the presence or absence of glottalization in the L1 of the listeners would interact with non-nativeness to reduce or increase the difference between native and non-native processing times.

It seems that the 40-ms advantage of Czech learners could be related to the fact that the glottal stop is frequently and systematically used in Czech (also due to its phonotactics), and the foreign accent peculiar to Czech English is, among other things, manifested through frequent glottalizations of word-initial vowels [11]. Spanish speakers, who use glottal elements much less often and generally for different purposes, benefited from their presence in the speech signal to a significantly smaller extent.

The question remains why native speakers of English (the British group) displayed such a considerable effect of glottalization if its use is described as optional in their mother tongue. We speculate that processing one's own native language enables one to draw on additional resources which are not available to non-native speakers and which may have obscured the picture of glottal stop usage by this group of listeners.

The native speakers of English also differed from the two 'foreign' groups in that they produced more invalid responses on the glottalized items. This result also warns against premature conclusions about the positive effect of glottalized word initial vowels.

One of the interesting outcomes of the study is the occurrence of a greater number of invalid responses in manipulated items compared with the natural items and longer reaction times to manipulations where the glottalization is excised from the signal. We knew from the outset that our manipulations must not leave any artifacts in term of clicks, formant discontinuities or F0 excursions. Yet is has been also known for some time that any disturbances in the natural flow of speech (minor shifts in rhythmic patterns) lead to longer reaction times in word-monitoring experiments (see, e.g., [17]). This is in agreement with contemporary accounts of the role of speech rhythm in cerebral processing of the speech signal [18]. According to these accounts human brain predicts the timing of the incoming acoustic events from the context and the generalized knowledge of the rhythmic patterns of the language. When we added into or excised glottalizations from our natural phrases, the rhythmic contexts were changed even if no consciously perceivable artifacts were apparent.

In conclusion, the presence or absence of the glottal stop before word-initial vowels clearly matters. The results of our study suggest that a direct comparison of native and non-native speech processing may not be as straightforward as it was previously believed. While the behaviour of the Czech compared with the Spanish group is in line with linguistic descriptions and our expectations, the advantage of 'nativeness' may act against simple extrapolations of speech processing trends.

Of the many remaining questions we would also like to analyze the variation in individual responses of the subjects in order to address the question whether the variation in production among individuals (as described, e.g., in [2]) is paralleled in perception or whether their perception is governed by more general rules (i.e., the rules that generalize the production of a large body of speakers encountered by the listener throughout his or her life experience).

Moreover, the semantic status of the target word, its size (in terms of the number of syllables), or the phonological density of its competitors (the number of words that are phonemically similar) will have to be considered. Although the repeated measures design which we used does not require addressing all these issues at once, we would like to focus on them in the nearest future.

5. Acknowledgements

The authors would like to acknowledge the support of the European Union Grant MRTN-CT-2006-035561 "Sound to Sense" and express gratitude to all participants and assisting colleagues in Bristol, Vitoria and Prague.

6. References

[1] Ladefoged, P. and Maddieson, I., The sounds of the world's languages, Cambridge, MA, Blackwells, 1996.

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

[3] Batliner, A., Burger, S., Johne, B. and Kießling, A., "MÜSLI: a classification scheme for laryngealizations", in Proceedings of ESCA Workshop on Prosody. Lund, pp. 176-179, 1993.

[4] Hanson, H.M., Stevens, K.N., Hong-Kwuang, J.K., Chen, M.Y. & Slifka, J., "Towards models of phonation", Journal of Phonetics 29: 451-480, 2001.

[5] Gordon, M. and Ladefoged, P., "Phonation types: a cross-linguistic overview", Journal of Phonetics 29: 383-406, 2001.

[6] Cruttenden, A., Gimson's Pronunciation of English, London, Edward Arnold, 1994.

[7] Wells, J.C., Longman Pronunciation Dictionary, 3rd edition (1st edition 1990), Harlow, Pearson Education, 2008.

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

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

[10] Umeda, N., "Occurrence of glottal stops in fluent speech", Journal of the Acoustical Society of America, 64(1): 88-94, 1978.

[11] Bissiri, M.P. and Volín, J., "Prosodic structure as a predictor of glottal stops before word-initial vowels in Czech English", in R. Vích [Ed], 20th Czech-German Workshop – Speech Processing, Prague, 23-28, 2010.

[12] Kilborn, K. and Moss, H., "Word monitoring", Language and Cognitive Processes 2(6): 689-694, 1996.

[13] Garcia Lecumberri, M.L., Cooke, M., and Cutler, A., "Non-native speech perception in adverse conditions: a review", Speech Communication 52: 864-886, 2010.

[14] Charpentier F., Moulines E. Pitch-Synchronous Waveform Processing Techniques for Text-to-Speech Synthesis Using Diphones. Proceedings of Eurospeech 89 (2): 13-19. 1989

[15] Boersma, P. & Weenink, D. "Praat: doing phonetics by computer" [Computer program]. Version 5.1.38, retrieved in November 2010 from http://www.praat.org/

[16] Forster, K. I. and Forster, J. C. "DMDX: A windows display program with millisecond accuracy". Behavior Research Methods, Instruments, & Computers, 35: 116-124, 2003.

[17] Buxton, H., "Temporal predictability in the perception of English speech", in A. Cutler and D. R. Ladd [Eds], Prosody: Models and Measurements, 111-121. Berlin, Springer-Verlag, 1983.

[18] Grossberg, S., "Resonant neural dynamics of speech perception", Journal of Phonetics 31: 423-445, 2003.