GLOTTALIZATION AND TIMING AT UTTERANCE FINAL POSITION IN HUNGARIAN: READING ALOUD VS. SPONTANEOUS SPEECH

Alexandra Markó & Anna Kohári

Department of Phonetics, Eötvös Loránd University, Budapest
marko.alexandra@btk.elte.hu, koharianna@gmail.com

ABSTRACT

Speakers tend to mark boundaries of larger prosodic units with glottalization and the deceleration of articulation rate. In the present study, the final parts of Hungarian read and spontaneous utterances were analyzed in the temporal domain (compared to the other parts of the utterances) and in terms of glottalization. We investigated how glottalization and deceleration are related to each other in read and spontaneous speech in Hungarian. We also analyzed if these phenomena depend on the speech mode. Our results revealed connection between glottalization and deceleration in spontaneous speech, whereas for read speech no such relation could be detected. Speech modes were also found to differ in the frequencies of the occurrence of glottalization and the magnitude of the deceleration at utterance final positions.

Keywords: utterance final lengthening, glottalization, spontaneous speech, read speech

1. INTRODUCTION

Boundaries of larger prosodic units are of a great importance in speech perception; therefore speakers tend to mark these places. Segments are typically longer in utterance-final position, and this phenomenon is considered to be universal, however the degree and extent of lengthening varies across languages (for a review see [5]). The boundary marking role of phrase/utterance final glottalization has also been confirmed for several languages [4, 8, 12, 18, 20].

Hungarian has a quantity distinction both for vowels and consonants. White and Mády [23] demonstrated the operation of utterance-final vowel lengthening in Hungarian, despite the apparent importance of duration as a cue to phonological vowel length. The phonological difference of quantity seemed to have an effect on the realization of lengthening: lengthening of short vowels was observed only in the last syllable of the utterance, while long vowels were lengthened in the penultimate syllable, too. On the other hand, Kovács [11] found lengthening of vowels also in the penultimate syllable (the methods of [23] and [11] were different). Earlier studies for Hungarian read speech showed that both vowels and consonants are longer in absolute utterance final position than utterance medially [9, 13]. Utterance final lengthening is also applied in the prosodic models of Hungarian text-to-speech systems [17, 21].

The utterances in spoken Hungarian were found to show a general trend of deceleration both in read and spontaneous speech [10, 22]. The effects that yield deceleration (e.g. utterance-final lengthening) overcome the factors that lead to acceleration (e.g. polysyllabic shortening), both in frequency and magnitude [10].

Glottalization frequently occurs utterance finally, both in read and spontaneous Hungarian speech [2, 14, 15, 16]. Glottalization extends to the last several syllables, but the voice may become regular again on the final syllable [2], for English data see e. g. [8, 20]. However the frequency of occurrence of glottalization is speaker dependent to a large extent. It has been shown that the less speakers glottalize, the more probable it is that they do it at a boundary position [16].

In the present study, the final parts of Hungarian read and spontaneous utterances were analyzed in the temporal domain (compared to the other parts of the utterances) and in terms of glottalization. According to the main hypothesis, if a speaker glottalizes more, the degree of deceleration is smaller, and vice versa. It was also supposed that the degrees of both glottalization and deceleration would be higher in reading aloud than in spontaneous speech.

2. SUBJECTS, MATERIAL AND METHOD

The material of 10 non-smoking speakers was selected from the Hungarian spoken language database BEA [6]. Female and male speakers were equal in number, their age spread between 20 and 60 years (with an average of 41.3 years).

As part of the database recordings, the subjects read aloud a shorter text (title followed by 12 sentences; number of words between 8 and 33, on average 19.2 word per sentence). 12 final segments (speech sounds) of 130 read utterances were analyzed after excluding disfluency phenomena. A total of 1,560 segments of the read subcorpus were analyzed for duration and glottalization. An additional 1,560 segments were tested for deceleration/acceleration.

In the spontaneous subcorpus, 2 minutes of speech per subject were recorded in an interview situation. This added up to 106 utterances (10-12 utterances per subject) with a total of 1,272 analyzed segments. An additional 1,272 segments were tested for deceleration/acceleration.

Segmentation was carried out on the basis of [7]. Pauses and hesitations were excluded, in line with [7].

Glottalization was annotated segment by segment as follows. It was marked if the segment could not be glottalized (voiceless consonant), or if glottalization was impossible to detect (e.g. because of noise). In the other cases, if any small part of the segment was both visually and audibly irregular, it was considered glottalized (for the details of the method see [3]). Since glottalization has a variety of functions, its motivation was being monitored: for example, it was indicated (and later ignored in the analysis) if glottalization was probably motivated by a phrase boundary within the utterance.

For annotation, MAUS [19] (with manual correction) and Praat [1] software were used.

Speech sound duration was used for measuring the change of articulatory rate, as this parameter has been proved adequate in speech rhythm measurements and other issues as well [5, 10]. Since optimal window length cannot be defined for the analysis of accelerations/decelerations, we used arbitrary window sizes. Articulation rate was measured for the final 3, 6 and 12 segments, and was compared to the penultimate 3, 6 and 12 segments, respectively. The articulation rate of the penultimate section was divided by that of the final section, and multiplied by 100. These ratios are marked by symbols: A_3, A_6, A_{12}, where the indices mark the window size. The resulting values above 110% were labeled as deceleration, below 90% as acceleration, while values in between were considered stagnation. These categorization procedures are denoted by T_3, T_6 and T_{12}.

With regard to glottalization, our analysis was twofold. In the first procedure (marked with G), the number of all glottalizable segments was considered 100% in 12- 6-, and 3-segment long sequences. In the second procedure (marked with GV), the number of vowels was considered 100%, and the ratio of glottalized vowels was measured within the 12-, 6- and 3-segment long sequences.

Even though the analysis included arbitrary elements, the combined measurements may yield a complex picture of the utterance final articulatory phenomena under study.

Statistical analyses (Mann-Whitney-test and Spearman correlation) were performed with SPSS 20.0.

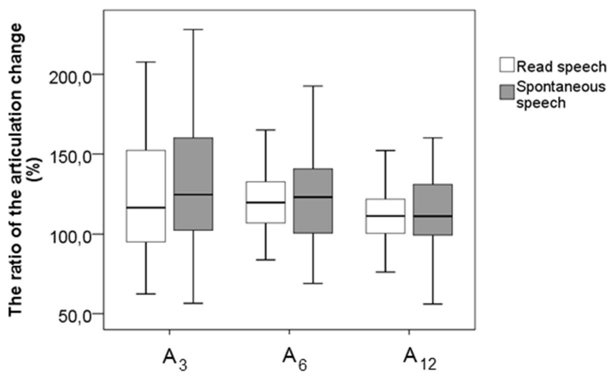
3. RESULTS

3.1. Timing at utterance final position in different speech modes

In agreement with previous studies [5] we have found that the most common timing pattern at utterance final positions is deceleration, independently of speech mode. Among the three categories, based on the temporal changes (deceleration, acceleration, or stagnation) between the last three and the preceding three segments (T_3), deceleration was found to be the most common (59.7% of the utterances), whereas acceleration and stagnation occurred in 17.2% and 23.2% of the cases, respectively. The same type of analysis based on the last 6 and 12 speech segments (T_6 and T_{12}) yielded similar distributions. The frequency of decelerating utterances was 67.0% using the T_6 method, and 52.9% with T_{12}. Acceleration was detected in 5.2% of the T_6, and 10.4% of the T_{12} results. Stagnation was present in 27.9% of the T_6, and 36.7% of the T_{12} measurements. There was not a single utterance where all three methods showed clear acceleration. Meanwhile, except for a small, 9.3% fraction of the utterances, at least one of the three methods indicated deceleration, so articulation rate decreased – in some vicinity of the utterance final position – in the vast majority of the cases.

On average, the magnitude of articulation rate change (measured by parameters A_3, A_6, and A_{12}) was higher in spontaneous speech than in read speech (Figure 1). In statistical terms, the differences were significant for A_3 (Z = 7808.0, p = 0.036), whereas those of A_6 (Z = 6507.0, p = 0.355), and A_{12} did not attain significance (Z = 6884.0, p = 0.765).

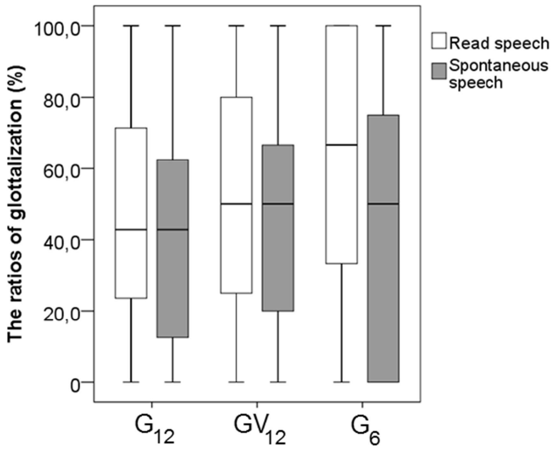
Figure 1: Articulation change ratios as a function of speech mode



3.2. Glottalization at utterance final position in different speech modes

The next parameters to be discussed are the glottalization ratios G_3, G_6, and G_{12}, which quantify the glottalization of last 3, 6 and 12 segments. Averaging over the whole corpus, the mean of G_3 was found to be 52.0%, i.e. glottalization has been realized in about half of the possibly glottalized segments (within the utterance final three segments). For the last 6 segments this ratio (G_6) was 53.3%, whereas for the last 12 speech segments (G_{12}) only 44.1%. Considering the vowels, the means of the glottalization ratios were somewhat higher (GV_3: 59.0%, GV_6: 60.0%, GV_{12}: 50.1%). Due to the limited number of possible outcomes, some of the glottalization ratio parameters were not suitable for statistical testing. Therefore only the ratios G_{12}, GV_{12}, and G_6 were investigated for different speech modes (Figure 2). The glottalization ratios were generally higher in read speech than in spontaneous speech, but the Mann-Whitney-test showed difference only for G_6 ($Z = 4882{,}0$, $p < 0.001$), while the other two measures (G_{12}, GV_{12}) did not attain significance ($Z \geq 6884.0$, $p \geq 0.175$).

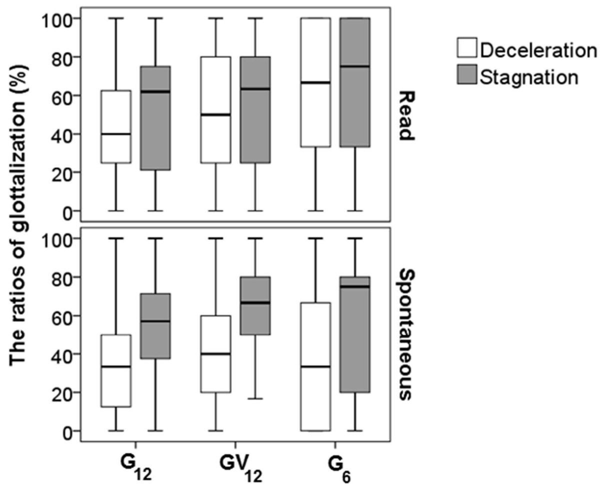
Figure 2: Glottalization ratios as a function of speech modes



3.3. Relation between glottalization and timing at utterance final position

The relation between glottalization and timing at the utterance final position was analyzed separately for read and spontaneous speech, due to the differences discussed in the previous section. Concerning the temporal changes based on the three categories (deceleration, acceleration, or stagnation), glottalization appeared more frequently if stagnation was detected at the end of the utterance than for decelerating endings. The analysis of temporal development based on the last 6 segments (T_6, see Figure 3) showed significant effect on the ratios of glottalization in spontaneous speech (G_{12}: $Z = 1363.0$, $p = 0.002$; GV_{12}: $Z = 1378.5$, $p = 0.001$; G_6: $Z = 1262.5$, $p = 0.018$). The timing analyses based on the different window sizes show the same trends but their significance could not be verified (T_3: $p \geq 0.369$; T_{12}: $p \geq 0.695$).

Figure 3: The frequency of glottalization (G_{12}, $GV_{12,}$ GV_6) at decelerating and stagnating utterance final positions based on T_6

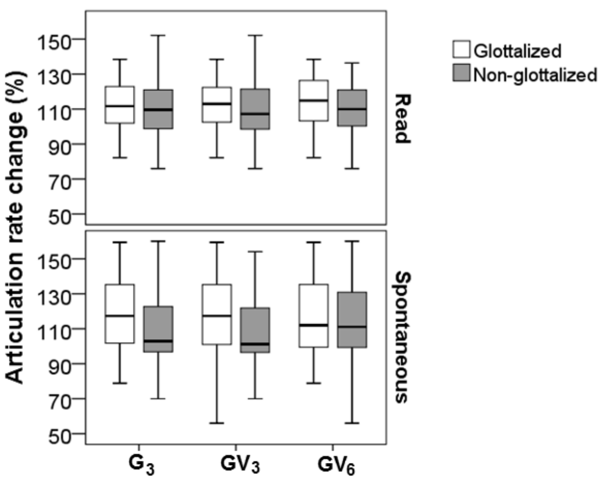


Although timing patterns for read speech show similar tendencies to spontaneous speech, the glottalization appeared more frequently at stagnating endings than at decelerating units. Yet, again, the statistical tests did not confirm this relation (T_3: $p \geq 0.475$; T_6: $p \geq 0.439$; T_{12}: $p \geq 0.162$). The investigation of the accelerating utterance boundaries was omitted because of the small number of such endings, as discussed earlier.

Since it was intended to analyze the relation between glottalization in the last 3 segments and temporal change, the percentages of glottalization were transformed into two categories based on whether it contains glottalized segment(s) or not. The percentages of glottalized vowels in the last 6 segments have been categorized the same way. The occurrence or the absence of glottalization was not found to be consistently connected to the magnitude of the articulation rate change. However, in spontaneous speech the deceleration based on the last 12 speech segments (A_{12}) was larger where the glottalization was absent in the last 3 segments (G_3, GV_3), and smaller where the glottalization appeared (Figure 4). These differences were supported by statistical tests (G_3: $Z = 1001.0$, $p = 0.048$; GV_3: $Z = 1001.0$, $p = 0.022$). The timing analyses based on the other window sizes (A_3, A_6) did not show marked connection with the presence of glottalization ($p \geq$

0.453). The appearance of glottalized vowels in the last 6 segments (GV_6) did not affect the articulation changes ($p \geq 0.482$). In the read speech no relationship between the lack or the presence of glottalization (G_3, GV_3, G_6) and the temporal changes (A_3, A_6, A_{12}) could be verified by the statistical tests ($p \geq 0.085$).

Figure 4: The articulation change rate (A_{12}) at utterance final positions as a function of glottalization (G_3, GV_3, GV_6)



3. DISCUSSION

Glottalization and deceleration frequently appear near the ends of the utterances. Both phenomena play important roles in the indication and perception of the boundaries in speech. In this study we have tried to investigate whether the glottalization and the deceleration showed some inherent connection or they marked the boundaries independently of each other. Before examining this correlation the main characteristics of these phenomena were analyzed.

In Hungarian we have found that deceleration appears consistently at the end of the utterance both in read and spontaneous speech. Although it is hard to define the accurate place of deceleration, based on our data it could be measured in the last 3 or, 6 or 12 segments. Only a few utterance final sequences showed stagnating or accelerating patterns. It seems that not only the last two syllables are involved in deceleration but more. Glottalization, similarly to deceleration, is also not confined to the last two syllables only, but appears frequently in the last 4 or 5 syllables, too, in agreement with previous studies [16]. Note, that the ratios of glottalization were the highest when the last 6 segments were considered.

However, these two phenomena also depend on speech mode. The magnitude of deceleration was found to be higher in spontaneous speech than in read speech, although this connection could be proven only in the case of the last three segments. Glottalization ratios were found to have an opposing trend: glottalization tend to appear more frequently in read speech than in spontaneous speech. One may assume, that these relations could be explained by the differences in the length of the utterances or in articulation rate, but in the studied corpus neither the lengths (as measured in terms of the number of speech segments), nor the average articulation rates did depend on the speech modes ($p \geq 0.557$).

The relationship between glottalization and temporal changes show further differences according to speech modes. In read speech, we did not find clear correlation between the two phenomena. However, in spontaneous speech glottalization appeared less frequently if the ending was decelerating than for stagnating endings. Also, the occurrences of glottalization tended to coincide with smaller magnitudes of deceleration than those observed if glottalization was absent. It is important to note, that these results were not consistent for all analyzed window sizes. Glottalization in three studied lengths appeared together with some type of articulation rate change, but not with all of them. This may be caused by the fact that articulation rate does not change uniformly at the end of the utterances, rather it shows a certain variability on different time scales.

It seems that glottalization and deceleration are not inherently connected to each other. For a better understanding of their complex relationship a more detailed analysis of inter- and intra-speaker differences could be performed in the future.

7. REFERENCES

[1] Boersma, P., Weenink, D. 2013. *Praat: doing phonetics by computer [Computer program]*. Version 5.3. http://www.praat.org.

[2] Bőhm, T., Ujváry, I. 2008. Az irreguláris fonáció mint egyéni hangjellemző a magyar beszédben [Irregular phonation as an individual speaker's characteristic in Hungarian speech]. *Beszédkutatás 2008,* 108–120.

[3] 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.

[4] Fant, G., Kruckenberg, A. 1989. Preliminaries to the study of Swedish prose reading and reading style. *Speech Transmission Laboratory Quarterly Progress and Status Report* 30/2. Stockholm: Royal Institute of Technology. 1–80. http://www.speech.kth.se/prod/publications/files/qpsr/1989/1989_30_2_001-080.pdf

[5] Fletcher, J. 2010. The prosody of speech: Timing and rhythm. In: Hardcastle, W. J., Laver, J., Gibbon, F. E. (eds), *The Handbook of Phonetic Sciences.* 2nd edition. Oxford: Wiley-Blackwell, 521–602.

[6] Gósy, M. 2012. BEA: A multifunctional Hungarian spoken language database. *Phonetician* 105–106, 51–62.
[7] Grabe, E., Low, E. L. 2002. Durational variability in speech and the rhythm class hypothesis. *Papers in Laboratory Phonology* 7. 515–546.
[8] Henton, C., Bladon, A. 1988. Creak as a sociophonetic marker. In: Hyman, L. M., Li, C. N. (eds), *Language, Speech and Mind. Studies in Honour of Victoria A. Fromkin.* London–New York: Routledge, 3–29.
[9] Kasssai, I. 1979. *Időtartam és kvantitás a magyar nyelvben. [Duration and quantity in Hungarian]* Budapest: Akadémiai Kiadó.
[10] Kohári, A. 2013. Temporal patterns of segments and intervals in Hungarian language. In: Mertens, P., Simon, A. C. (eds), *Proceedings of the Prosody-Discourse Interface Conference* (IDP). Leuven. 51–57. http://www.ling.arts.kuleuven.be/franitalco/idp2013/papers/Mertens_Simon_2013_Proceedings_IDP2013.pdf
[11] Kovács, M. 2002. *Tendenciák és szabályszerűségek a magánhangzó-időtartamok produkciójában és percepciójában. [Tendencies and regularities in production and perception of vowel duration]* Debrecen: Debreceni Egyetem Kossuth Egyetemi Kiadója,
[12] Lehiste, I. 1965. Juncture. In *Proc. 5th ICPhS*. New York: S. Karger, 172–200.
[13] Magdics, K. 1966. A magyar beszédhangok időtartama [Duration of Hungarian speech sounds]. *Nyelvtudományi Közlemények* 68, 125–139.
[14] Markó, A. 2009. Stigmatizált hanglejtésforma a spontán beszédben [A stigmatized intonation contour in spontaneous speech]. *Beszédkutatás 2009,* 88–106.
[15] Markó, A. 2010. A prozódia szerepe a spontán beszéd tagolásában [The role of prosody in the organization of spontaneous speech]. *Beszédkutatás 2010,* 82–99.
[16] Markó, A. 2011. A glottalizáció határjelző szerepe a felolvasásban [Boundary marking by glottalization in reading aloud]. *Beszédkutatás 2011,* 31–45.
[17] Olaszy, G. 2002. Model to predict Hungarian sound durations for continuous speech. *Acta Linguistica Hungarica* 49, 321–345.
[18] Redi, L., Shattuck-Hufnagel, S. 2001. Variation in the realization of glottalization in normal speakers. *Journal of Phonetics* 29, 407–429.
[19] Schiel, F. 1999. Automatic phonetic transcription of non-prompted speech. In Ohala, J. J., Hasegawa, Y., Ohala, M., Granville, D., Bailey, A. C. (eds), *Proc. 14th ICPhS*. San Francisco: University of California, 607–610.
[20] Slifka, J. 2006. Some physiological correlates to regular and irregular phonation at the end of an utterance. *Journal of Voice* 20, 171–186.
[21] Tóth, B. P. 2013. *Rejtett Markov-modell alapú gépi beszédkeltés. [TTS based on Hidden Markov Model]* Doktori értekezés. [PhD thesis] Budapest: BME. ttp://www.omikk.bme.hu/collections/phd/Villamosmernoki_es_Informatikai_Kar/2013/Toth_Balint_Pal/ertekezes.pdf.
[22] Váradi, V., Beke, A. 2013. Az artikulációs tempó variabilitása felolvasásban. [The variability of articulation rate in read speech]. *Beszédkutatás 2013,* 26–41.
[23] White, L., Mády, K. 2008. The long and the short and the final: Phonological vowel length and prosodic timing in Hungarian. In: Barbosa, P. A., Madureira, S., Reis, C. (eds), *Proc. of Speech Prosody 2008.* Campinas, Brasil, 363–366.