PERCEPTION OF CONSONANT LENGTH OPPOSITION IN HUNGARIAN STOP CONSONANTS

Tilda Neuberger
Dept. of Phonetics, Research Institute for Linguistics,
Hungarian Academy of Sciences
e-mail: neuberger.tilda@nytud.mta.hu

Abstract

A specific speech sound is characterized by strong variability in articulation. Nevertheless, despite great variability, human listeners are able to identify phonemes successfully. The most important acoustic cue that distinguishes geminate from singleton stop consonants in production is the closure duration. The aim of this paper is to examine the role of duration as a perceptual distinction between single and geminate stops using a binary discrimination test of stops with systematically manipulated closure duration. Results confirmed that closure duration is a sufficient and adequate perceptual cue in the distinction of single and geminate stops; however, differences between the response curve of original geminates and that of original singletons indicate that cues other than closure duration may contribute to the length contrast in Hungarian stops.

Keywords: geminate, quantity, perception, stop consonant, closure duration, manipulation

1 Introduction

One of the most interesting questions in speech perception research is how the continuous speech signal can be recognised as sequences of speech sounds. Furthermore, identification of speech sounds as discrete phonemes is also an important issue. It has been shown that a specific speech sound is characterized by strong variability in articulation, including timing of articulation (Lehiste, 1970; Lisker, 1974; Rosen, 1992). In spite of this great variability, however, human listeners are able to identify phonemes successfully, irrespective of speaker or speech condition. During the identification of consonants, listeners make decisions about place of articulation, manner of articulation, voicing contrast and length contrast.

The present paper investigates the perception of consonantal length contrast in Hungarian, which is a language that has distinctive long consonants. The aim of this study is to examine how native listeners discriminate phonological categories of quantity (singleton and geminate) along a continuous durational scale of the phonetic realisations.

There has been much research indicating that duration of singleton and geminate consonants shows considerable overlap in production (e.g., Lisker, 1958; Pickett et al., 1999; Hirata & Whiton, 2005; for Hungarian consonants see Gósy, 2004; Olaszy, 2006; Beke & Gyarmathy, 2010; Gráczi, 2012; Neuberger, 2015), which might make discrimination difficult to listeners in some cases, particularly during first or second language acquisition. In Hungarian children, for instance, the most difficult perceptual process at the phonological-phonetic level has proved to be the consonantal length contrast discrimination (Gósy, 1989; Horváth & Gyarmathy, 2010). Still, adult listeners seem to have the ability to effectively cope with this task in everyday speech situations. The effectiveness of this process may be due to the extraction of acoustic cues and phonetic information that speech sounds contain. More precisely, different phonetic categories correlate with various acoustic cues facilitating human perception of different phonemes. However, there is not a one-to-one mapping between acoustic features and perceptual categorization of speech sounds (Blumstein & Stevens, 1985).

According to the acoustic invariance theory of speech perception, invariants establish a link between acoustics and perception during the identification of phonetic segments (Stevens & Blumstein, 1978, 1981, 1985; Blumstein & Stevens, 1979; Lisker, 1985). It has been suggested that only a limited set of articulatory configurations leads to stable acoustic patterns. These patterns are the ones that define the finite set of speech sounds. Similarly, from the perspective of speech perception, the human auditory system can judge differences between sounds along a physical scale. Two items on the continuum which cannot be discriminated by the perceptual system are identified as realisations of the same phoneme, while the items that are discriminated are deemed to belong to different phonetic classes (Blumstein & Stevens, 1979). It is assumed that the acoustic signal contains invariant information that is present in all instances that correspond to the perceived linguistic unit (Wright et al., 1999).

A considerable number of studies have contributed to the elaboration of this theory. Most of these studies focused on invariant acoustic properties that can be used to classify stop consonants according to place of articulation (Fant, 1960; Stevens & Blumstein, 1978; Sussmann et al., 1991). Besides place-of-articulation categories, the acoustic and perceptual correlates of other consonantal contrasts have been examined, such as the voiced-voiceless distinction (Liberman et al., 1958; Lisker & Abramson, 1964; Stevens & Klatt, 1974; Williams, 1977; Sussmann et al., 1991) or the single-geminate distinction (Pickett & Decker, 1960; Lisker, 1974; Abramson, 1986, 1987; Hankamer & Lahiri, 1988; Hankamer et al., 1989; Schmidt & Flege, 1995). This study will focus on the single-geminate distinction.

Production and perception studies have revealed that the primary acoustic attribute that distinguishes geminates from singletons is duration (e.g., Pickett & Decker, 1960; Klatt, 1976; Hankamer et al., 1989; Ylinen et al., 2005; Ridouane, 2010). Geminate consonants occur contrastively with singleton consonants in many languages, such as Arabic, Japanese, Italian, Finnish or Hungarian. In other languages, consonant length is not contrastive, geminates only arise from morpheme concatenation (also known as fake geminates). Hence, geminates may occur across morpheme boundaries, but not morpheme-internally (e.g., English *top pick* vs. *topic*, French *Il l'aime* vs. *Il aime*). Both contrastive and non-contrastive geminates were found to be longer than matched singletons, on average. In Hungarian, investigation of the quantity contrast may be important for several reasons. In this language, all consonants can occur as both short (singleton) or long (geminate). Considerable overlap in the production of singletons and geminates can make classification of the two categories difficult. In addition, the distribution of geminates is restricted in Hungarian. If an underlying geminate occurs next to another consonant, it obligatorily degeminates and must surface as short. Therefore, the degemination process has an influence on the temporal properties of consonants. In this case, the degree of shortening is a key issue (see Siptár & Gráczi, 2014). For stop consonants, studies have commonly shown that the most important acoustic and perceptual cue for length distinction is the duration of the closure phase (Ham, 2001; Ridouane, 2010). Therefore, perception research has largely focused on the measurement of this parameter.

1.1 Previous work on perception of consonant length

As background to the present study, several perception experiments which applied an artificial manipulation technique on closure duration are described here. Listeners' responses to incrementally manipulated durations allow one to determine the point at which a singleton percept shifts to a geminate percept. Thus, durational correlates of the length distinction can be examined. Some of the following studies focused on the role of absolute closure duration as a measure of length distinction, while other studies emphasized the role of relational timing in various languages.

Absolute duration was investigated by the manipulation of closure duration in many languages, such as Marathi, Arabic, Turkish, Bengali and English. Lisker's (1958) perceptual experiment dealt with the length distinction of Marathi stop consonants. Closure duration of [t] in the word *matə* 'mind' was artificially manipulated to increase in 20 ms steps, while the closure duration of the long counterpart in the word *mattə* 'drunk' was manipulated to decrease in 20 ms steps. The perceptual boundary value was noted between 140 and 160 ms in the former case and between 140 and 120 ms in the latter case. For Arabic, a very similar result to Lisker's (1958) results was found by Obrecht (1965). In this case, the perceptual boundary between geminate and non-geminate [bː]-[b] were between 140 and 160 ms.

The region between 120 and 160 ms of closure duration proved to be an important interval in both Turkish and Bengali length distinctions. Hankamer et al. (1989) employed two sets of stimuli with closure duration varying incrementally between that of geminate and non-geminate stops in Turkish and Bengali. Several minimal pairs were used in this experiment and the task of the listeners was to write down the word they thought they heard (e.g., Turkish [ata] 'horse (dat)' or [atːa] 'horse (loc)' and Bengali [pat̪a] 'leaf' or [pat̪ːa] 'whereabouts'). Responses to stimuli created from original geminate differed from those for non-geminates in both languages. The original geminates were identified as geminates more frequently than the original non-geminates at closure durations between 120 and 160 ms. The authors concluded that the responses were "biased by secondary features of the acoustic signal when the closure duration cue is in the ambiguous region between 120-160 ms" (Hankamer et al., 1989: 295). However, examination failed to reveal the exact secondary cue. Thus, it was assumed that the bias may be due to a combination of cues.

A perceptual distinction experiment was carried out using English test sentences with an intervocalic single consonant *(topic)* and its double counterpart *(top pick)* as stimuli (Pickett & Decker, 1960). The closure duration of [p] was altered by inserting or removing magnet tape. The effect of closure duration and rate of utterance was tested. Findings showed that closure durations shorter than 150 ms were judged as a single consonant, while closures longer than 250 ms were judged as geminate. It was also found that as the rate of utterance increased (from 2 to 8 syllables per second), the threshold closure duration decreased (from 320 to 140 ms). This raises the question of whether acoustic invariance exists in an absolute form.

Numerous researchers have assumed that acoustic invariance is **relational invariance**, and relational timing plays a crucial role in the perception of length contrast. In other words, not only, and not primarily, absolute closure duration plays a role in length distinction, but other variables, such as various durational ratios are critical. In addition, it is important to take speaking rate into consideration as well.

Pind (1995) proposed relational timing as an important acoustic property which is able to define durational categories across speaking rates. His findings showed that the ratio of vowel to rhyme duration was a stable acoustic feature in distinguishing V:C versus VC: syllable categories in Icelandic, and it remained invariant at different speech rates.

The closure duration of labial and dental stops was found to discriminate singletons and geminates in a production and a perception experiment in Italian (Pickett et al., 1999). In production, the duration of singletons and geminates showed overlap across different speaking rates, however, the ratio of consonant duration to preceding vowel duration remained stable across speaking rate. It was hypothesised that manipulation of the C/V ratio affected the perception of the quantity distinction and resulted in perceptual shifts. A significant main effect of C/V ratio was found, indicating that listeners tended to change the category of their responses as a result of the timing manipulation.

In accordance with the findings in Italian by Pickett et al. (1999), Hansen's (2004) production data for Tehrani Persian demonstrated that increased speaking rate had a greater influence on the closure duration of geminates than it did on the duration of singletons. His pilot study revealed that different threshold values were required to discriminate between singletons and geminates for isolated and for connected speech because of the overlap of isolated singleton durations and connected geminate durations.

Japanese stop quantity distinction was investigated in the theoretical framework of relational acoustic invariance (Amano & Hirata, 2010). The authors analysed the perceptual boundary between single and geminate stops across speaking rates (fast, normal, slow). They found that closure duration at the perception boundaries ranged between a relatively large interval across speaking rates (34–213 ms, mean: 110 ms). It was also found that the durational ratio of stop closure to word (CW ratio) is an invariant parameter in distinguishing the two phonemic categories.

Japanese stop quantity distinction was examined with respect to another durational ratio by Idemaru and Guion-Anderson (2010). They created test stimuli varying in previous mora duration and following vowel duration, while leaving closure duration and VOT unaltered. Listeners' 'geminate' responses increased as $C/Mora_1$ ratio increased, whereas a less important change was observed due to the C/V_2 variation.

Perception of consonant length was investigated among non-native listeners as well (Hayes, 2002; Wilson et al., 2005; Sonu et al., 2013). Results of a perception experiment of Japanese length contrast differentiation by English listeners suggested that non-native listeners did not use cues that vary by speaking rate but instead used absolute durational criteria (Wilson et al., 2005). Similar findings were observed in Korean learners of Japanese (Sonu et al., 2013). These findings suggest that non-native listeners tend to have difficulty identifying length contrasts affected by speaking rate variations.

1.2 Research questions and hypothesis of the present work

Three research questions were addressed in the present study. First, how does closure duration contribute to native listeners' discrimination of Hungarian singleton and geminate stops? Second, does place of articulation or voicing affect the differentiation of singleton and geminate stops? Third, do the response curves displayed across varied closure durations show the same pattern for the original singletons as for the original geminates?

To answer the questions above, a two-alternative forced choice test was applied for stops with systematically manipulated closure duration. The stimuli contained Hungarian stops [p, t, k, b, d, g][1] and their geminate counterparts.

Based on previous results for other languages, it was hypothesised that (i) closure duration would be a sufficient cue to quantity discrimination for native Hungarian listeners. It was also assumed that (ii) place of articulation and stop voicing might have an effect on the closure durations associated with perceptual boundaries. Finally, it was hypothesised that (iii) the original quantity of the stops would affect the listeners' responses to some extent due to probable secondary cues (e.g., preceding vowel duration, closure voicing or combination of acoustic characteristics) in the acoustic signal.

2 Method

2.1 Baseline experiment

A baseline experiment was carried out to ensure high reliability in constructing the test stimuli for the main perception experiment. Therefore, a set of nonwords containing single and geminate stops was created for this pre-experiment. VCV and VCːV sequences were recorded by a 27-year-old, Hungarian native female speaker. Being a qualified phonetician, she produced the sequences maintaining fundamental

[1] Stops [c ɟ] were not analysed because there is no consens on their manner of articulation whether they are stops or affricates.

frequency and sound pressure relatively constant in each token. The recording was made with an AT 4040 side-address condenser microphone using GoldWave software in a sound-proof booth located within the Phonetic Department of the Research Institute for Linguistics of the Hungarian Academy of Sciences. Recordings were digitized with a sampling rate of 44.1 kHz (storage: 16 bits, 86 kbytes/s, mono).

The consonants of the sequences were [p, t, k, b, d, g] and their long counterparts, whereas the preceding and the following vowels were identical. Vowels were [i], [aː], or [u], since they represent high, low and back vowels, respectively. The speaker read 36 items [6 (consonants) × 2 (quantity) × 3 (adjacent vowels)], such as [ipi], [uku], [itːi], [aːdːaː] etc. She produced each item three times, resulting in 108 productions. These nonwords were prepared for the baseline experiment: 54 nonsense word realisations contained short stop consonant sequence, and 54 nonsense word realisations contained long stop consonants.

The perceptual robustness of the length distinction was evaluated in the baseline experiment. Eleven Hungarian-speaking adults (9 females, 2 males) participated in this experiment. Their mean age was 27 years (range = 22–34 years). Each listener completed a discrimination test using Praat software (Boersma & Weenink 2015). The task of the listeners was to listen to the samples and make a binary decision about whether the heard consonant was long or short, for example [itːi] or [iti]. If it was heard as long, they chose 'long' response, whereas when the presented consonant seemed to be short, they had to click the 'short' answer on the screen. Items were played in random order. Reaction time were measured by means of Praat. Identificaton accuracy was measured for each token. Identification accuracy was defined as 100% when all participants gave the same 'short' or 'long' response to a token containing short or long consonant corresponding with the speaker's intention.

Results of the baseline experiment revealed that listeners identified consonant length at an accuracy rate of 82–100%. In general, the poorest result was in the case of low vowels, while the best result was in the case of the front high vowels (identification accuracy: 98,5% on average). Therefore, listeners' responses to each token containing [i] were further analysed, while tokens containing [u] and [aː] were excluded from further investigation. Thus, listeners' responses to 36 tokens containing vowel [i] and consonants [p, t, k, b, d, g] and their long counterparts were examined with respect to identification accuracy and reaction time.

Identification accuracy of each token can be seen in Figure 1. Correct identification accuracy of 100% means unanimous agreement among the listeners. The lower the identification accuracy, the more listeners gave incorrect responses to the token. It was found that all listeners gave correct responses to at least one of the three produced token regarding place of articulation (labial, alveolar, velar) and voicing (voiceless, voiced) of stops. Given the results of the baseline identification experiment, 12 tokens were selected for manipulation in the main perception test. Selection was made based on the combination of high accuracy rate (100% in each case) and relatively short reaction time (< 1.2 ms). Overall, the mean reaction time duration was 1.5 ms.

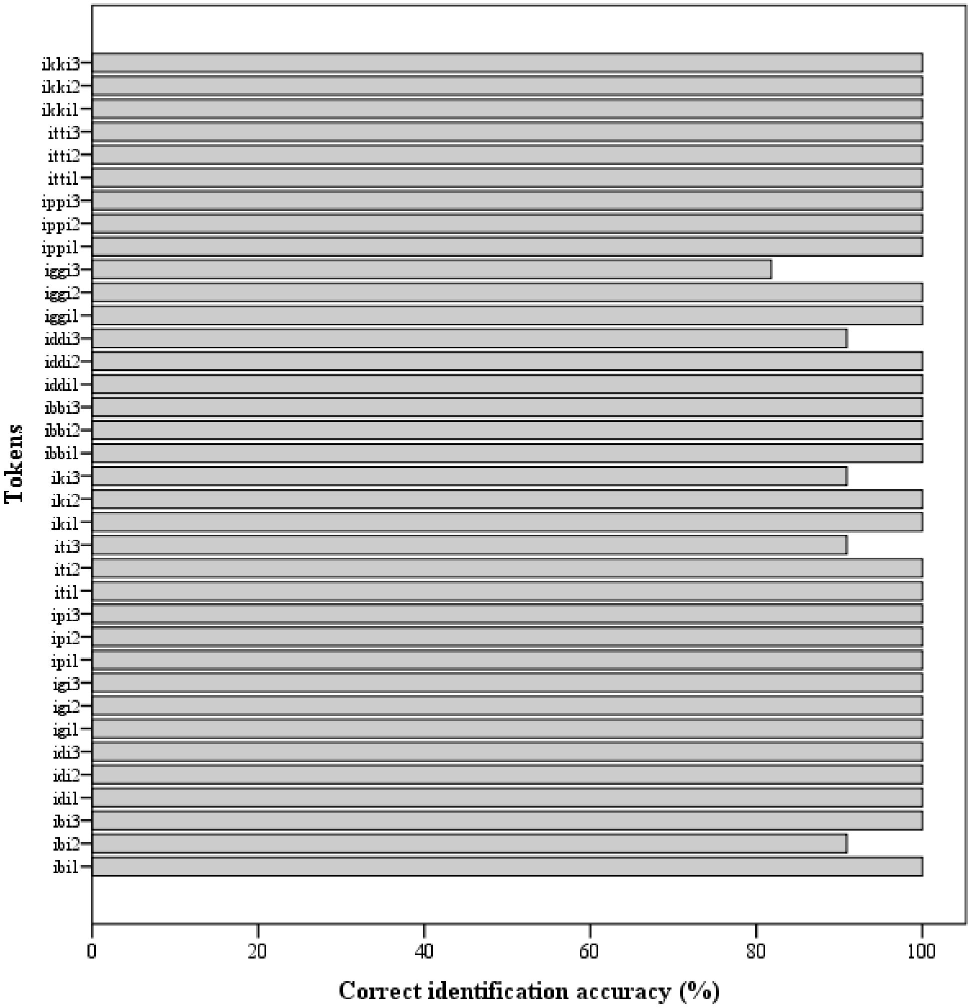

Figure 1. Correct identification accuracy for each token in the baseline experiment

2.2 Perceptual test stimuli

In order to construct the perceptual test stimuli with different closure durations, as mentioned above, 12 tokens were selected for manipulation: 6 tokens with singleton and 6 tokens with germinate stops. Using these tokens, two sets of stimuli were created. For one of them, tokens with a short consonant were used (e.g., [ipi] or [idi]). For the other one, tokens with long consonants were used (e.g., [ipːi] or [idːi]). In the first case, stimuli were made by artificially lengthening the closure duration of original singletons [p, t, k, b, d, g] in 10 ms steps up to the closure duration of the matched geminates [pː, tː, kː, bː, dː, gː]. For instance, closure duration of the originally singleton [t] was manipulated from 104 ms to 224 ms in twelve 10 ms steps in order to create a continuum from [ipi] to [ipːi]. For the other set of stimuli, the closure duration of the original geminates was shortened likewise in 10 ms steps, resulting in nonsense words with artificially shortened geminate. Minimum and maximum values of closure durations and the number of steps of manipulation are listed in Table 1 for each consonant. The reason for having two sets of stimuli, i.e., shortened geminates and lengthened singletons, was to observe any possible differences in listeners' 'short' or 'long' responses for original singletons and original geminates with equal closure duration. It is assumed that if response curves for the originally geminate stimuli are located somewhere else than those of the originally singleton stimuli, acoustic cues other than closure duration may play a role in singleton vs. geminate stop distinction.

Incremental manipulation of closure duration was conducted using a Praat script. In order to do this, sequences had to be first segmented into speech sounds. Consonant closure boundaries (start time and end time) were marked based on visual observation of the oscillogram and spectrogram, as well as auditory feedback. Closure duration was defined as the time interval between the termination of the preceding vowel and the stop burst. Start time of the closure duration was measured at the offset of the vertical striations of the preceding vowel's formants, while the end time of closure duration was measured right before the release burst. The content of closure duration, namely the silent interval in voiceless stops and voicing in voiced stops, was preserved during the manipulation. Temporal and spectral properties of the adjacent vowels, the VOT of voiceless stops, and the burst releases remained unaltered. Altogether, 138 tokens were created.

Table 1. Details of stimuli derived from closure duration manipulation

Consonant	**Minimum-maximum closure duration (ms)**	**Number of steps of manipulation**
[p]-[pː]	103–204	10
[t]-[tː]	104–224	12
[k]-[kː]	104–204	10
[b]-[bː]	102–222	12
[d]-[dː]	104–224	12
[g]-[gː]	105–175	7

2.3 Perceptual test subjects

Forty-four native speakers (33 females, 11 males) of standard Hungarian participated in this experiment. They were recruited from a university in Budapest. All of them were undergraduate students of linguistics, having little experience in phonetic studies. Their age ranged from 18 to 27 years, and the mean age was 21 years. No participant reported being diagnosed with a speech or hearing disorder.

2.4 Perceptual test conditions

The experiment was run in a quiet room at the university in Budapest. Each participant listened to the recorded samples through Sennheiser HD 419 headphones. The nonwords were played one by one in random order. The listeners' task was to make a binary decision, choosing between 'short' or 'long' responses, by clicking the appropriate button in Praat software (i.e., the same procedure as in the baseline experiment described above).

2.5 Analyses

For evaluating the binary data, listeners' responses to each item were summarized. Percentage of geminate responses at each duration was measured in the case of all stop consonants. Then the two sets, i.e. originally singleton and originally geminate stimuli, were analysed separately. Fitting a logistic function to the sets and plotting response curves was carried out using MATLAB 2015b.

Binary logistic regression was used for statistical analysis in SPSS 20.0 software. A generalized linear mixed model (GLMM) was constructed by 'responses' as the target (or dependent) variable, 'closure duration' was the fixed effect, and 'speaker' was the random effect. The interaction of closure duration and place of articulation, as well as closure duration and voicing, were analysed (as fixed effects). To compare responses to stimuli created from originally singleton and originally geminate stops, paired samples *t*-test in SPSS were used.

3 Results

Since closure duration is indeed a major cue to the length distinction in production, it is expected that listeners also use this parameter in perception. Figures 2 and 3 show listeners' responses by means of response curves. Response curves represent the percentage of 'long' responses (y-axis) at different closure durations (x-axis) for each stop. A typical sigmoid function was plotted in accordance with the closure duration of the consonants. As Figures 2 and 3 show for voiceless and voiced stops (respectively), listeners judged consonants with relatively long closure durations as 'long' and they hardly judged consonants with relatively short closure durations as 'long'. It is also worth noting that approximately 100 ms closure duration induced a total agreement of 'short' response among the 44 participants (in this case, proportion of 'long' responses was close to zero, see Figures 2 and 3). In contrast, closure durations approaching 200 ms triggered unanimous 'long' decisions (in this case, the proportion of 'long' responses was close to the maximum, see Figures 2 and 3). Binary logistic regression showed that closure duration had a main effect on listeners' 0-1 (short-long) responses: $F(1, 6067) = 1317.391$; $p < 0.001$.

A comparison of the response curves for the voiceless and voiced stops revealed similar patterns for [p, t, k, b, d, g]. As can be seen in Figures 2 and 3, when moving more posterior in place of articulation, response curves were arranged at shorter closure durations along the time axis. Statistical analysis revealed a significant interaction between closure duration and place of articulation for the stop: $F(2, 6067) = 29.427$; $p < 0.001$ and between closure duration and voicing of the stop: $F(1, 6067) = 50.082$; $p < 0.001$. This means that not only closure duration but other variables have effects on listeners' responses as well.

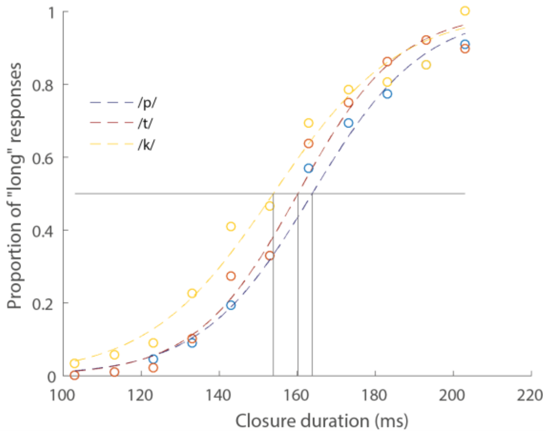

Figure 2. Response curves for manipulated voiceless stops (horizontal line at 50% is also marked, vertical lines indicate closure duration values at 50% boundary points)

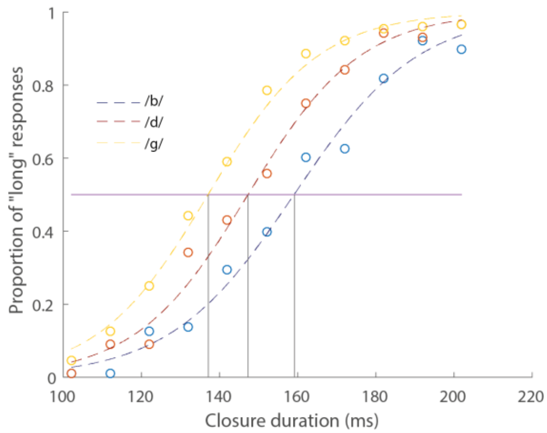

Figure 3. Response curves for manipulated voiced stops (horizontal line at 50% is also marked, vertical lines indicate closure duration values at 50% boundary points).

These differences in the response curves for different stops are related to the fact that stops of different places of articulation exhibit closure duration differences. Previous production studies provided objective data on closure duration differences among labial, dental and velar stops in Hungarian. For instance, Gósy and Ringen (2009) measured that the mean closure duration was 69 (±14) ms for labial, 59 (±13) ms for dental, and 53 (±14) ms for velar voiced stops in the intervocalic position in isolated words. In Gráczi's (2013) data using nonsense words, mean closure duration of intervocalic voiceless stops seemed to be shorter by moving more posterior in place of articulation: 90 (±13) ms in labial, 72 (±15) ms in dental, 70 (±13) ms in velar stops on average. The mean closure duration of intervocalic voiced stops was shorter than that of voiceless stops: 70 (±10) ms for labial, 52 (±10) ms for dental, and 60 (±12) for velar stops (Gráczi, 2013). For spontaneous speech material, the mean closure duration was 79 (±11) ms for labial, 71 (±18) ms for dental, 63 (±18) ms for velar stops (Neuberger, 2015). The former values represent singleton stop consonants. Spontaneous speech data for voiceless geminate stops showed an average closure duration of 115 (±20) ms for labial, 122 (±31) for dental, 106 (±27) ms for velar stops (Neuberger, 2015). Considering these mean values, it was expected that the perceptual shift in listeners' responses from singleton to geminate would follow the same tendency described by production data.

In order to observe the perceptual shift from singleton to geminate, in the next step of analysis, closure duration at the 50% perceptual boundary point was measured. This represents the 50% rate between the identification of a singleton versus a geminate. The values of 50% boundary closure duration are shown in Table 2 for each stop. The more posterior the place of articulation, the shorter the 50% boundary closure duration. This result shows consistency with the above mentioned production data. More specifically, since stops proved to be produced with shorter closure durations by moving more posterior in place of articulation, it could be expected that 50% boundary closure duration would show this tendency as well. The 50% boundary durations of stops with respect to place of articulation showed this tendency, however, statistical analysis did not confirm that these differences were significant ($p > 0.05$). In terms of place of articulation, the shortest duration boundary involved velar stops and the longest duration was attributed to labial stops. Listeners judged voiced consonants as 'long' at a shorter closure duration than voiceless ones. The shortest closure duration perceptually belonged to the voiced velar stop, while the longest one belonged to the voiceless labial stop. Nevertheless, voicing did not show a significant main effect on 50% boundary values ($p > 0.05$). The average boundary value in the entire data set was 153 ms. Listeners tended to judge consonants as 'short' below this duration, and as 'long' above this duration.

Table 2. 50% boundary closure duration (ms) of stop one by one and in total, as well as across place of articulation and voicing

[p]	**[t]**	**[k]**	**Voiceless**
163	160	153	159
[b]	**[d]**	**[g]**	**Voiced**
159	147	137	147
Labial	**Alveolar**	**Velar**	**Total**
161	153	145	153

A comparison of the response curves of originally short and originally long consonants was made in order to reveal the possible presence of any secondary cues that may help listeners in discriminating the two phonological categories. It was assumed that in case of differences between response curves of the two conditions (stimuli made by originally singletons vs. originally geminates), closure duration may not be the only cue to length discrimination. The results showed that response curves of stimuli created from different original quantity were not identical (Figures 4 and 5 for voiceless and voiced stops respectively). In each graph, blue curves show listeners' responses to stimuli created from originally singleton stops, while red curves shown listeners' responses to stimuli created from originally geminate stops (x-axis: manipulated closure duration; y-axis: percentage of listeners' 'long' responses, horizontal yellow line: 50% boundary). A paired samples *t*-test showed the difference was significant in [k]: $t(10) = -3.798$, $p = 0.003$; in [b] $t(13) = 3.879$; $p = 0.002$; and in [g] $t(7) = -3.815$; $p = 0.007$.

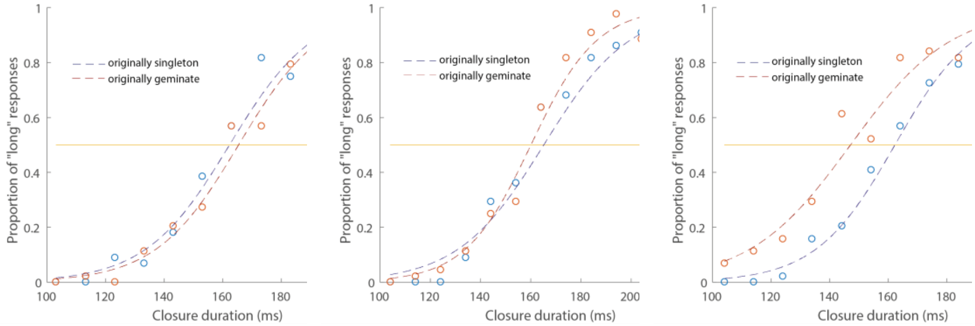

Figure 4. Response curves for voiceless stops with respect to the original length of stimuli ([p] – left; [t] – middle, [k] – right)

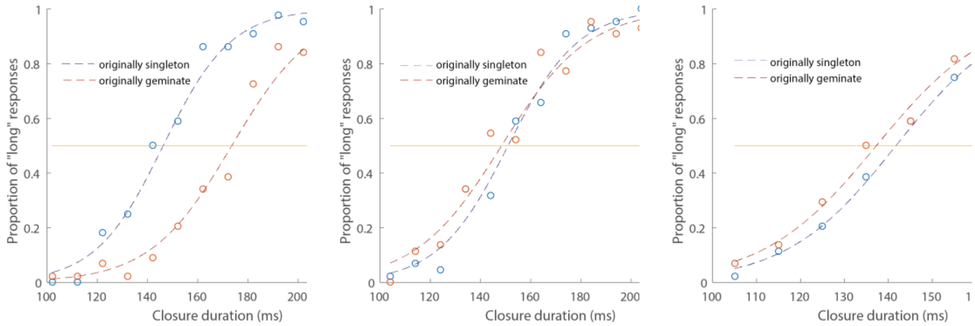

Figure 5. Response curves for voiced stops with respect to the original length of stimuli ([b] – left; [d] – middle, [g] – right)

The difference between the 50% boundary closure duration of the originally singleton and originally geminate stops indicated that listeners' reactions differed according to the original quantity of some of the stop consonants (see Table 3). The largest shift in the boundary location occurred in [b] with 27 ms, while the smallest shift occurred in case of [d] with 2 ms. In the latter case, listeners' responses to original geminates and original singletons were basically the same. The difference between the two stimuli conditions proved to be minor in case of [p] and [t] as well (minor differences in response curves of [p], [t] and [d] can also be observed in Figures 4 and 5).

For the alveolar and velar stops, the perceptual boundary of originally geminate consonants appeared earlier than in the case of responses to stimuli created from original singletons. The displacement of 50% boundary closure durations was the largest between [kː] and [k] (15 ms). At the 50% perception boundary closure, duration was 147 ms in original geminates and 162 ms in original singletons. Interestingly, labial stops showed a different behaviour. In the case of [p] and even more in the [b], response curves of original singletons were positioned at shorter closure durations. The shift in the boundary location was larger in the case of voiced labial stops than in the case of voiceless labial stops. The possible reasons for this displacement are discussed later in this article.

Table 3. 50% boundary closure duration (ms) with respect to the original length of stimuli

	[p]	[t]	[k]	[b]	[d]	[g]
Originally singleton	162	165	162	146	150	141
Originally geminate	165	160	147	173	148	137

4 Discussion and conclusion

Production of speech sounds has been shown to be highly variable, which may pose difficulties for listeners, especially concerning the perception of those temporal cues which are linguistically significant, i.e. define relevant phonemic contrasts, like distinctive consonant length. There has been notable research on how listeners deal with this problem in many languages but there is only a limited number of studies about the perception of consonant length contrast in Hungarian. The present study reported data on the role of closure duration in the perceptual distinction between Hungarian single and geminate stops in non-word items. Examination of native listeners' quantity discrimination was conducted with the intention of obtaining detailed information about the relationship between acoustic and perceptual cues of consonant length in Hungarian.

The first hypothesis – closure duration would be a sufficient cue to quantity discrimination for native Hungarian listeners – was established by the present data. Results of the present study confirmed the statement, which has been formulated for many languages that closure duration is the main perceptual cue to the stop consonant quantity distinction in the Hungarian language as well. This was substantiated by statistical analysis. The stimuli scale of closure duration between 100 ms and 200 ms provided a good medium to observe how listeners' responses shifted from 'singleton' to 'geminate'. Response curves created from the present data formed S shapes on the dimension of closure duration, which implies that listeners' perception operates more or less categorically during the process.

As mentioned above in the introduction, studies describing production data on Hungarian singleton and geminate consonants reported largely overlapping durations. This overlap is due to varying speaking rates, individual articulatory properties, stress and phonetic positions of the given consonant etc., and mostly occurs in the interval of 80–120 ms of duration (e.g., production data of Gráczi, 2012; Neuberger, 2015). The question arises whether this overlap in duration causes difficulties for listeners in identifying consonant length. Based on the responses in the 80–120 ms interval in the present perception data, it would be expected for listeners to judge these consonants as singleton, even if they were geminates. But this was not the case in everyday speech situations. This is the reason why it is supposed that, in continuous speech, the perception of quantity extends to larger units than speech sounds (syllable or word-sized units) and listeners identify consonant length considering relational properties and not absolute duration of consonants.

The second hypothesis was that place of articulation and voicing of the stops might have an effect on the closure durations related to perceptual boundaries. The present data supported this hypothesis. Values of the 50% boundary closure duration showed correspondence with the position of the tongue in the mouth during articulation of stop: The more posterior the place of articulation, the shorter the 50% boundary closure duration. We can conclude that this parameter is associated with the different durations of intervocalic stops across places of articulation in production (Gósy & Ringen, 2009; Gráczi, 2012). In the case of shorter closure duration (which is more common in posterior stops), the perceptual boundary between singletons and geminates is situated at lower values.

The third hypothesis was that the original quantity of the manipulated stops would affect the listeners' responses. Results of the present investigation, along with results from earlier studies, suggested that acoustic cues other than closure duration may play a role in the identification of quantity contrast. The plausible presence of other acoustic cues was revealed by differences in response curves comparing originally geminate to originally singleton stimuli. The displacement of the two response curves along the time axis varied in size across consonant quality. The greatest displacement was manifested in velar stops, on average. This means that possible secondary cues are more salient in these consonants than in labial or alveolar ones. In most of the analysed stops consonants, the 50% boundary closure durations were lower in the case of originally geminate stimuli than in the case of originally singletons. This finding indicates that listeners tended to identify originally geminate stimuli as 'long' even at shorter closure durations than they did in case of originally singleton stimuli. This suggests that the [+long] feature might be coded in the speech signal by other acoustic properties besides closure duration (e.g., preceding vowel duration, closure voicing), and that the listeners' auditory system might be sensitive to these articulatory/acoustic characteristics. At this stage in the research process, many questions remained unanswered. What acoustic cues cause the differences in listeners' responses between the two stimuli conditions? Further acoustic and perceptual analysis is required to determine the relevant secondary cues. The issue of cues other than closure duration is more significant in languages in which word-initial voiceless stop geminates exist (for instance Pattani Malay, see Abramson, 1987; or Cypriot Greek, see Muller, 2003; or Swiss German, see Kraehenmann & Lahiri, 2008). Despite the fact that the silent interval in the closure of singleton and geminate voiceless stops in utterance-initial position is not audible, listeners are capable of differentiating these two categories. Kraehenmann and Lahiri (2008) revealed by using electropalatography that even if audible closure duration is missing, speakers articulate initial geminate stops with longer oral closures than singletons (based on duration of contact of the tongue and hard palate). Still, there must be other perceptual cues besides closure duration in determining the length distinction of initial stops, such as intensity of stop burst, rate of formant transitions, fundamental frequency perturbations (Abramson, 1987) or voice onset time (Muller, 2003) etc. These acoustic features may function as secondary cues in quantity distinction of non-initial stop geminates. The task of investigating the role of other cues in Hungarian length distinction will be addressed in a future study.

There was an unexpected result which ought to be interpreted. That is, the 50% boundary closure duration of originally singleton [b] and geminate [bː] also showed significant difference, but the response curves of the originally singleton stimuli were positioned at shorter closure durations than that of the geminate. This result may lead to the conclusion that the originally geminate [bː] stimuli did not contain secondary cues which seemed to be evident in the velar and alveolar stops. Or, which is more likely, there was some aspect of the stimuli that made identification complicated. After reanalysis, it can be seen that the percentage of voicing in closure phase (which was not 100% in the original geminate [bː] nor in the original singleton [b]) remained stable along with manipulation of closure duration. Using aerodynamic modelling, Westbury and Keating (1986) stated that closure of a relatively long intervocalic stop is likely to be initially voiced and then voiceless. Moreover, Ohala (1983) concluded that geminates not voiced through release are generally categorized as voiceless. By shortening the closure duration of [bː], the voiced part also decreased, and by lengthening closure duration of [b], the voiceless part also increased, which might have caused uncertainty in the identification of stop quality in [p] or [b]. (It must be noted that all [d] and [g] tokens in the material were fully voiced, therefore, the above mentioned problem was not present in cases of these consonants.) Previous research has confirmed that voiceless stops can be distinguished from voiced counterparts based on voicing in closure phase and are (significantly) longer that voiced counterparts, thus, consonant duration may function as secondary cue to voicing contrast (Maddieson, 1997; Olaszy, 2006; Gósy & Ringen, 2009; Gráczi, 2011). It is supposed in this study that an inverse phenomenon also exists, that is, voicing properties (partially-voiced closure) can be a secondary cue to length distinction, by influencing the percentage of voicing in closure that distinguishes singleton and geminate stop consonants. Nevertheless, further examination is needed to provide evidence for what effects partial voicing have on the percept of geminate stops.

Another important issue to investigate is the Hungarian length contrast in the face of variation in speaking rate and on the basis of larger units than speech sounds. Findings from an investigation of acoustic and perceptual correlates of various phonetic distinctions may be useful in understanding many issues, such as the development of speech perception, second language learning or processing models of speech perception.

Acknowledgments

The paper was supported by the János Bolyai Research Scholarship and by the Postdoctoral Fellowship Program of the Hungarian Academy of Sciences. This paper has benefited considerably from contributions of Mária Gósy, András Beke and Anita Auszmann. The author thanks reviewers for their suggestions and constructive feedback.

References

Abramson, A. S. (1986). The perception of word-initial consonant length: Pattani Malay. *Journal of the International Phonetic Association*, 16(01), 8-16.

Abramson, A. S. (1987). Word-initial consonant length in Pattani Malay. *Haskins Laboratories Status Report on Speech Research*, 143-147.

Amano, S. & Hirata, Y. (2010). Perception and production boundaries between single and geminate stops in Japanese). *The Journal of the Acoustical Society of America*, 128(4), 2049-2058.

Beke, A. & Gyarmathy, D. (2010). Zöngétlen résmássalhangzók akusztikai szerkezete. [The acoustic structure of voiceless fricatives]. *Beszédkutatás* 2010. 57-76.

Blumstein, S. E. & Stevens, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *The Journal of the Acoustical Society of America*, 66(4), 1001-1017.

Blumstein, S. E. & Stevens, K. N. (1985). On some issues in the pursuit of acoustic invariance in speech: A reply to Lisker. *The Journal of the Acoustical Society of America*, 77(3), 1203-1204.

Boersma, P. & Weenink, D. (2015). *Praat: doing phonetics by computer [Computer program]*. Version 5.3. Retrieved from www.praat.org on 2 Jan 2015.

Fant, G. (1960). *Acoustic Theory of Speech Production.* The Hague: Mouton.

Gósy M. (2004). *Fonetika, a beszéd tudománya.* Budapest: Osiris.

Gósy M. (1989). *Beszédészlelés.* Budapest: MTA Nyelvtudományi Intézet.

Gósy, M. & Ringen, C. O. (2009). Everything you always wanted to know about VOT in Hungarian. In: *IXth International Conference on the Structure of Hungarian,* Debrecen, Hungary. http://icsh9.unideb.hu/pph/handout/Ringen_Gosy_handout.pdf

Gráczi, T. E. (2011). Voicing contrast of intervocalic plosives in Hungarian. In: *Proceedings of the 17th International Congress of Phonetic Sciences*, 759-762.

Gráczi T. E. (2012). *Zörejhangok akusztikai fonetikai vizsgálata a zöngésségi oppozíció függvényében*. Doktori disszertáció. Budapest: ELTE BTK.

Gráczi T. E. (2013). Explozívák és affrikáták zöngésségének időviszonyai. *Beszédkutatás 2013*, 94–120.

Ham, W. (2001). *Phonetic and phonological aspects of geminate timing*. New York: Routledge.

Hankamer, J. & Lahiri, (1988). A. The timing of geminate consonants. *Journal of Phonetics*, (16). 327-338.

Hankamer, J., Lahiri, A. & Koreman, J. (1989). Perception of consonant length: Voiceless stops in Turkish and Bengali. *Journal of Phonetics*, 17(4), 283-298.

Hansen, B. B. (2004). Production of Persian geminate stops: Effects of varying speaking rate.

In: *Proceedings of the 2003 Texas Linguistics Society Conference*. Somerville: Cascadilla Press, 86-95.
Hayes, R. L. (2002). *The Perception of Novel Phoneme Contrasts in a Second Language: A Developmental Study of Native Speakers of English Learning Japanese Singleton and Geminate Consonant Contrasts*. Coyote Papers, 12, 28-41. Retrieved February 7, 2011, http://coyotepapers.sbs.arizona.edu/CPXII.htm.
Hirata, Y. & Whiton, J. (2005). Effects of speaking rate on the single/geminate stop distinction in Japanesea). *The Journal of the Acoustical Society of America*, 118(3), 1647-1660.
Horváth V. & Gyarmathy D. (2010). A beszédhallás szerepe a beszédhang-differenciálásban. [The role of hearing in speech sound differentiation]. *Gyógypedagógiai Szemle* 2010/2. 126-135.
Idemaru, K. & Guion-Anderson, S. (2010). Relational timing in the production and perception of Japanese singleton and geminate stops. *Phonetica*, 67(1-2), 25-46.
Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America*, 59(5), 1208-1221.
Kraehenmann, A. & Lahiri, A. (2008). Duration differences in the articulation and acoustics of Swiss German word-initial geminate and singleton stops). *The Journal of the Acoustical Society of America*, 123(6), 4446-4455.
Lehiste, I. (1970). Suprasegmentals, Cambridge, Massachusetts: M. I. T. Press.
Liberman, A. M., Delatire, P. C. & Cooper, F. S. (1958). Some cues for the distinction between voiced and voiceless stops in initial position, *Language and Speech*, 1(3), 153-167.
Lisker, L. (1958). The Tamil occlusives: short vs. long or voiced vs. voiceless. *Indian Linguistics, Turner Jubilee*, 1, 294-301.
Lisker, L. (1985). The pursuit of invariance in speech signals. *The Journal of the Acoustical Society of America*, 77(3), 1199-1202.
Lisker, L. & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20(3), 384-422.
Lisker, L. (1974). On time and timing in speech. In T. A. Sebeok et al. (Eds.), *Current trends in linguistics* (Vol. 12.). The Hague: Mouton. 2387-2418.
Maddieson, I. 1997. Phonetic Universals. In J. Laver & W. J. Hardcastle (Eds.), *The handbook of phonetic sciences*. Oxford: Blackwells. 619-639.
Muller, J. S. (2003). The production and perception of word initial geminates in Cypriot Greek. In: *Proceedings of the 15th International Congress of Phonetic Sciences*. 1867-1870.
Neuberger, T. (2015). Durational correlates of singleton-geminate constrast in Hungarian voiceless stops. In *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow. Paper number 0422.1-5 retrieved from http://www.icphs2015.info/pdfs/Papers/ICPHS0422.pdf
Obrecht, D. H. (1965). Three experiments in the perception of geminate consonants in Arabic. *Language and Speech*, 8(1), 31-41.
Ohala, J. (I983). The origin of sound patterns in vocal tract constraints. In MacNeilage, P. F. (ed.), *The production of speech*. New York: Springer Verlag. 189-216.
Olaszy, G. (2006). *Hangidőtartamok és időszerkezeti elemek a magyar beszédben*. Nyelvtudományi Értekezések 155. Budapest: Akadémiai Kiadó.
Pickett, J. M. & Decker, L. R. (1960). Time factors in perception of a double consonant. *Language and Speech*, 3(1), 11-17.
Pickett, E. R., Blumstein, S. E. & Burton, M. W. (1999). Effects of speaking rate on the singleton/geminate consonant contrast in Italian. *Phonetica*, 56(3-4), 135-157.
Pind, J. (1995) Speaking rate, voice-onset time, and quantity: The search for higher-order invariants for two Icelandic speech cues. *Perception & Psychophysics*, 57, 291-304
Ridouane, R. (2010). Geminates at the junction of phonetics and phonology. *Papers in Laboratory Phonology*, 10, 61-90.
Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 336. 1278, 367-373.

Schmidt, A. M. & Flege, J. E. (1995). Effects of speaking rate changes on native and nonnative speech production. *Phonetica*, 52(1), 41-54.
Siptár P. & Gráczi T. E. (2014). Degemination in Hungarian: Phonology or phonetics? *Acta Linguistica Hungarica* 61, 443-471.
Sonu, M., Kato, H., Tajima, K., Akahane-Yamada, R. & Sagisaka, Y. (2013). Non-native perception and learning of the phonemic length contrast in spoken Japanese: training Korean listeners using words with geminate and singleton phonemes. *Journal of East Asian Linguistics*, 22(4), 373-398.
Stevens, K. N. & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *The Journal of the Acoustical Society of America*, 64(5), 1358-1368.
Stevens, K. N. & Blumstein, S. E. (1981). The search for invariant acoustic correlates of phonetic features. In Eimas, Peter D. and Miller, Joanne L. *Perspectives on the study of speech*. Hillsdale, NJ: Lawrence Erlbaum. 1-38
Stevens, K. N. & Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *The Journal of the Acoustical Society of America*, 55(3), 653-659.
Sussman, H. M., Mccaffrey, H. A. & Matthews, S. A. (1991). An investigation of locus equations as a source of relational invariance for stop place categorization. *The Journal of the Acoustical Society of America*, 90(3), 1309-1325.
Westbury, J. R. & Keating, P. A. (1986). On the naturalness of stop consonant voicing. *Journal of linguistics*, 22(1), 145-166.
Williams, L. (1977). The Voicing Contrast in Spanish. *Journal of Phonetics*, 5(2), 169-184.
Wilson, A., Kato, H. & Tajima, K. (2005). Native and non-native perception of phonemic length contrasts in Japanese: Effects of speaking rate and presentation context. *The Journal of the Acoustical Society of America*, 117(4), 2425-2425.
Wright, R., Frisch, S. & Pisoni, D. B. (1999). *Speech perception*. Wiley Encyclopedia of Electrical and Electronics Engineering, Research On Spoken Language Processing. Progress Report No. 21. Indiana University. http://www.iu.edu/~srlweb/old-site/publication/manuscript211.pdf
Ylinen, S., Shestakova, A., Alku, P. & Huotilainen, M. (2005). The Perception of Phonological Quantity based on Durational Cues by Native Speakers, Second-language Users and Non-speakers of Finnish. *Language and Speech*, 48(3), 313-338.