Interspeech 2018
2-6 September 2018, Hyderabad, India

# Length contrast and covarying features: Whistled speech as a case study

Rachid Ridouane[1], Giuseppina Turco[2], Julien Meyer[3]
[1]LPP, CNRS-Sorbonne Nouvelle, Paris
[2]LLF, CNRS-Université Paris-Diderot, Paris
[3]Univ. Grenoble Alpes, CNRS, GIPSA-lab, Grenoble 38000, France
rachid.ridouane@univ-paris3.fr, gturco@linguist.univ-paris-diderot.fr, julien.meyer@gipsa-lab.grenoble-inp.fr

## Abstract

The status of covarying features to sound contrasts is a long-standing issue in speech: are they deliberately controlled by the speakers, or are they contingent automatic effects required by the defining features? We address this question by drawing parallels between the way gemination is implemented in spoken language and the way it is rendered in whistled speech. Audio materials were collected with five Berber whistlers in Morocco. The spoken and whistled data were composed of pairs of words contrasting singletons to geminates in different word positions.

Compared to spoken forms, whistling, while adapting to the specific constraints imposed by the medium, transposes the basic strategies used in normal speech. As in normal speech, the primary and most salient acoustic attribute differentiating whistled singletons and geminates is closure duration. But duration is not used alone. Covarying secondary attributes are conveyed which may serve to enhance the primary correlate by contributing additional properties increasing the distance between the two lexical categories. These enhancing correlates may take on distinctive function in cases where the primary correlate is not implemented. This is, for instance, the case of higher frequency values in word-initial position where duration differences cannot be acoustically implemented using whistled speech.

**Index Terms**: Gemination, whistled speech, primary feature, enhancing feature, Tashlhiyt Berber

## 1. Introduction

This paper studies how the contrast of gemination is conveyed phonetically in whistled Tashlhiyt with respect to spoken Tashlhiyt. The goal is to determine whether the primary and covarying features rendering the contrast in the spoken language are also present in its whistled form. Whistling is one of the multiple modes of expression for some languages, which has the advantage to increase the audible range of speech and to enable dialogues when speakers are far from each others (for a review, see [1]). It is mostly found in mountainous and densely vegetated landscapes. Whistlers learn to copy words and utterances of their language into simpler whistled signals carrying key phonetic cues of the original acoustic and articulatory features. These are sufficient to guarantee high levels of intelligibility of the information encoded in the whistles [2-6]. Whistled Tashlhiyt, the object of this study, was recently found among shepherds of the High-Atlas in Morocco [7]. An example of a whistled Tashlhiyt word is given in Figure 1.

Gemination is a salient property of the linguistic system of Tashlhiyt, where each single consonant has a geminate counterpart [8]. Geminates are extremely common, and can occur in various contexts, including word-initial, -medial and -final positions (e.g. [ttut] "forget him", [tiddi] "height", [imikk] "little"). At the phonetic level, the distinction between the two series of consonants in normal speech is carried not just by duration but also by a combination of other properties [9, 10, 11]. The primary property is the extra duration of geminates. This appears in every context in which the contrast occurs, even in voiceless stops following pause. In this context closure duration, measured using electropalatographic data, is extra-long even though it has no direct acoustic manifestation [12]. In addition to duration, the contrast is also implemented by further covarying attributes such as shorter preceding vowel duration and higher release amplitude [9, 11]. These correlates are secondary since they are either contextually limited (vowel shortening) or present some variability across subjects and contexts (higher release amplitude). Native listeners are however sensitive to these attributes. For example, higher release amplitude can be exploited to recover the contrast between singleton and geminate voiceless stops in utterance-initial position, a position where closure duration differences cannot be perceived [13].

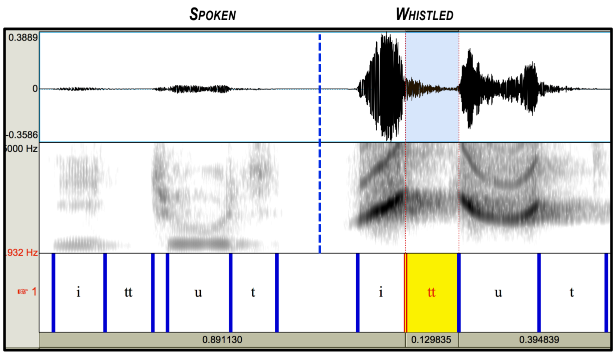


Figure 1: *Waveform and spectrogram of the spoken and whistled word [ittut] 'he forgot him'.*

The status of covarying features is a long-standing issue in speech: are they deliberately controlled by the speakers or are they contingent automatic effects required by the primary gesture? Covarying features are traditionally considered as automatic attributes which follow from universal rules of phonetic implementation of the primary features, and thus not directly planned by the speaker [14]. Recent developments in Feature Theory have challenged this "implementational" view. The fact that the secondary attributes are language-specific implies that they have to be learned and controlled by the speaker [15].

Enhancement Theory, which was originally developed to explain variability in feature realization, posits that most of the properties that covary with the defining attributes are under



10.21437/Interspeech.2018-1060

the voluntary control of the speaker. In this view, when the acoustic difference between two sounds is insufficiently great, risking confusion, a supplementary feature is deliberately introduced to increase the acoustic difference between them [16, 17]. Although there are clear examples of enhancement that cannot be considered as biomechanical effects of the primary gesture (e.g. lip-rounding usually added to /ʃ/ in English, increasing its auditory difference from /s/), the view that most of the properties that covary with the defining attributes are actively controlled has not been unambiguously demonstrated for other contrasts (e.g. intrinsic F0 differences in vowel height and voicing [18, 19]).

The biomechanical and/or aerodynamic constraints of the speech mechanism that may result in covarying features may not be at play in whistled speech. Moreover, the technique of whistling imposes severe constraints and restrictions on speech articulation [1, 5]. During this procedure, certain phonetic details that are present in standard speech are lost. As a result whistled speech is expected to oversimplify the phonetic implementation of phonological contrasts, and thus implement only those attributes that are actively controlled by the speaker. In what follows, we examine how the difference between whistled singletons vs. geminates is acoustically implemented in different prosodic positions, and test whether whistling also gives rise to secondary attributes. Results are discussed in relation to what we know about the implementation of gemination contrast in normal speech.

# 2. Method

### 2.1. Segmenting whistled speech

Whistled speech consists in a transformation of the spoken signal into a simple melodic line made up of frequency and amplitude modulations of a whistled signal. The first harmonic (H1) constitutes the fundamental frequency (F0) of the whistled tone and determines the perceived pitch of the whistled utterance. It is generally considered to roughly correspond to the second formant (F2) of the spoken form [4] (compare the F2 contour of the spoken /u/ with the corresponding H1 contour of the whistled /u/ in figure 1 above). Tashlhiyt vowels /i u a/ are whistled in intervals of frequencies that follow the same logic as what has already been reported for Turkish or Spanish, with /i/ statistically higher in frequency than /a/ which is also higher than /u/. The intervals of /i/, /a/, and /u/ are statistically different albeit important overlap, particularly between /a/ and /u/ [7].

The transposition of stop consonants in whistling, which is the topic of this study, involves two main components, a silent interval corresponding to closure duration and consonant-vowel transitions involving adaptation motivated by the restricted frequency range of the whistled melody [4]. Looking at the transitions at the edges of the dental stop /tt/ in Figure 1, one can see a parallel between spoken F2 transition (left) and whistled H1 transition, reflecting the expected high-frequency locus at the edges of dental consonants. Velars, on the other hand, are expected to be produced within a low frequency range. Are these available acoustic attributes (duration and consonant to vowel transitions) used to implement gemination contrast? Do they vary depending on the nature of the stops and their position within a word?

### 2.2. Participants and speech material

We recorded 5 Tashlhiyt whistlers (mean age 38.2) during fieldwork organized in the High-Atlas in Morocco. All the whistlers were male native speakers of Tashlhiyt and reported using whistling since their childhood. The corpus on gemination was part of a larger list of selected sentences and isolated words, recorded in a situation of elicitation. Sound level was systematically measured with a sound level meter (Rion NL42). The data examined in this study is composed of 13 minimal or near-minimal pairs contrasting singletons /t d k g/ to their geminate counterparts /tt dd kk gg/ in three word positions: word-initial (4 pairs, e.g. [gar] 'bad' vs. [ggal] 'swear'), word-medial (4 pairs, e.g. [tagut] 'cloud' vs. [aggu] 'smoke') and word-final (5 pairs, e.g. [inig] 'search' vs. [igigg] 'palm'). A total of 276 whistled forms were analyzed: 138 in singleton and 138 in geminate condition. Word-medial pairs were produced by five whistlers, word-final pairs by four of them, and word-initial ones by three of them. Whistlers were asked to first speak a word and then whistle it. Both spoken and whistled forms were segmented and annotated based on visual inspection of the acoustic signals and spectrograms using Praat 5.034 [20]. Only data from whistled items are analyzed in this article.

### 2.3. Acoustic measurements and statistical analysis

The whistled forms were annotated on the word level and, for each target item, both temporal and non-temporal measurements were taken. The temporal parameters include: (i) duration of pre-consonantal vowels in medial and final positions, and (ii) duration of stop closure. Non-temporal parameters include: (i) the frequency values of H1 at consonant-vowel transitions for word-initial and word-medial positions and (ii) the frequency values of H1 at vowel-consonant transition for word-final position. Figure 1 illustrates the measurements taken in medial position: duration of preceding vowel /i/, duration of closure for /tt/ and duration of following /u/ (results for following vowels are not reported here as they are not affected by gemination), in addition to H1 frequency values at /i/-/tt/ transition and /tt/-/u/ transition. The average of H1 frequency values (in Hz) was taken at the C-V and V-C transition. Given that whistled closure corresponds to a silent interval, this measurement could not be taken for word-initial position, since no direct signal of the relative durations of stop closures is visible. For this specific subset, only non-temporal measurements are presented. Note also that for word-final position, closure duration was measurable only for forms produced with a whistled stop release.

A series of regression models (one for each acoustic parameter, in each position) was performed [21] using the statistical R software [22]. For the temporal cues, the models contained duration (in ms) of the (whistled) stop closure and of the preceding vowel as function of condition (singleton vs. geminate). For the non-temporal cues, the frequency values of the H1 at the onset/offset of the preceding/following vowel were modeled as function of condition (singleton vs. geminate) and place of articulation (velar vs. dental). For each model, random effects were implemented (subject and item), and random-slopes were modelled for the effect of condition. Main effects of each predictor and of interactions were tested by using the Likelihood ratio test as implemented in the *anova*()-function.

## 3. Results

In this section we provide the results on how gemination contrast is implemented in whistling, starting with data in medial position.

### 3.1. Medial position

#### *3.1.1. Duration*

The mean duration of the (whistled) stop closure and of the preceding vowel in singleton and geminate conditions are illustrated in Figures 2 and 3, respectively. Singletons and geminates display significant durational differences both for consonants and for preceding vowels.

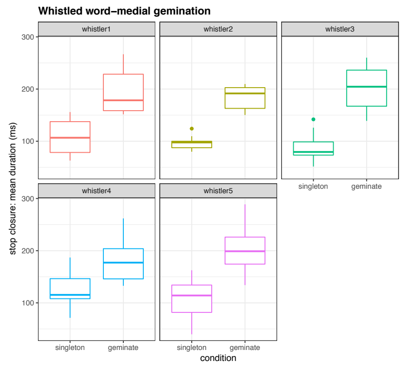


Figure 2: *Mean duration (in ms) of stop closure as function of condition, split by whistler.*

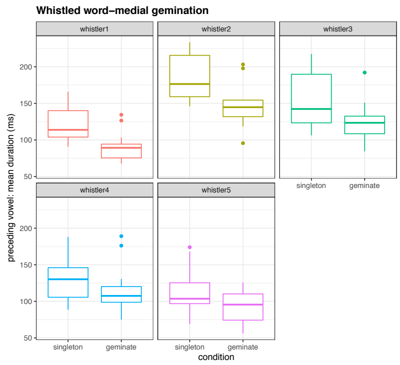


Figure 3: *Mean duration (in ms) of the preceding vowel as function of condition, split by whistler.*

The Likelihood ratio tests reveal an effect of condition for both models (consonant stop closure: $\chi2(1)=17.5$, $p< .0001$; preceding vowel: $\chi2(1)=6.15$, $p< .05$). The mean duration of the stop closure in word-medial position is longer in geminates (on average 88.9 ms difference) than in singletons; the duration of the preceding vowel is shorter in geminate than in singleton condition (on average 27.5 ms difference).

#### *3.1.2. H1 on C-V and V-C transition*

Figure 4 shows the mean values of H1 (in Hz) in C-V transition as function of condition (singleton vs. geminate) and place of articulation (velar vs. dental).

The analyses reveal an effect of condition ($\chi2(1)=13.1$, $p< .001$), an effect of articulation ($\chi2(1)=4.53$, $p< .05$), and no interaction ($p= .1$). The mean H1 frequency values at C-V transition are higher (on average 130.2 Hz difference) in geminates than in singletons and higher (568.1 Hz) for dentals /d/-/t/ than for velars /g/-/k/. For V-C transition, only place of articulation has an effect on H1 frequency ($\chi2(1)=6$, $p< .05$), going in the same direction as for consonants in C-V transition (641.2 Hz).

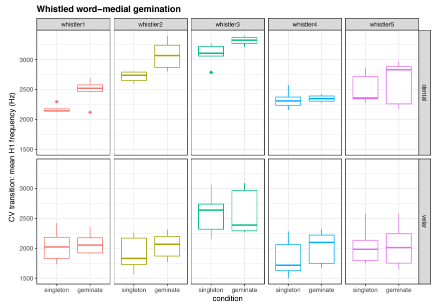


Figure 4: *Mean H1 frequency values (in Hz) at the C-V transition as function of condition and place of articulation, split by whistler.*

### 3.2. Final position

#### *3.2.1. Duration*

The durations of the stop closure and of the preceding vowel are displayed in Figures 5 and 6, respectively.

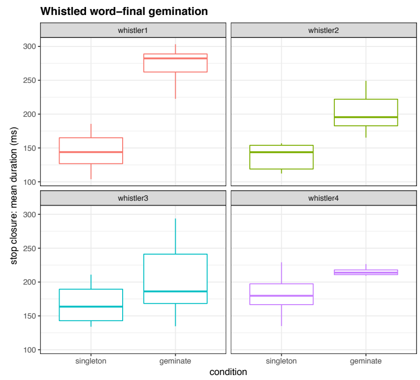


Figure 5: *Mean duration (in ms) of the stop closure as function of condition, split by whistler.*

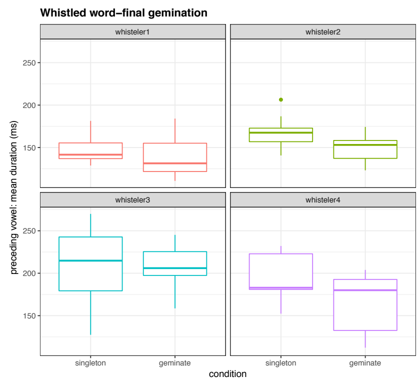


Figure 6: *Mean duration (in ms) of the preceding vowel as function of condition, split by whistler.*

The analyses reveal an effect of condition (stop closure: $\chi2(1)=34.6$, $p< .0001$; preceding vowel: $\chi2(1)=7.4$, $p< .01$). The mean duration of the stop closure in word-final position is longer in geminates than in singletons (71 ms); the duration of the preceding vowel is also shorter in geminate than in singleton condition (18.5 ms). Note, however, that the parameter on closure duration in this position is measured only for items produced with a visible release at the signal (12 unreleased items were excluded from this analysis). Similar to medial position, the analyses show no significant effect of gemination on the H1 frequency values at V-C transition.

### 3.3. Initial position

As already stated above, closure duration for whistled stops in word-initial position cannot be measured since no direct signal of the relative duration of these segments is visible in the acoustic waveform and spectrogram. Hence, only results on the H1 frequency values at C-V transition are presented.

#### *3.3.1. First harmonics on C-V transition*

The mean values of H1 (in Hz) in C-V transition as function of condition (singleton vs. geminate) and place of articulation (velar vs. dental) are shown in Figure 7.

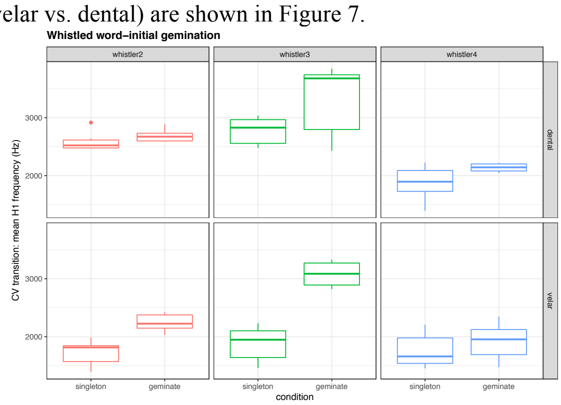


Figure 7: *Mean H1 frequency values (in Hz) at the C-V sequence as function of condition and place of articulation, split by whistler.*

The analyses reveal a main effect of each predictor (condition: $\chi2(1)$=4.3, $p< .05$; place of articulation: $\chi2(1)$=6.4, $p< .05$) and an interaction ($\chi2(1)$=5.1, $p< .05$). The mean H1 frequency values at the C-V transition are higher in geminates than in singletons (490.8 Hz) and for the dental consonants /d/-/t/ than for the velar consonants /g/-/k/ (530.7 Hz). Furthermore, the difference in H1 frequency values across conditions is higher for velars (575.4 Hz) than for dentals (412.7 Hz).

## 4. Discussion and conclusion

Because of constraints inherent to the whistled production, whistled speech is expected to oversimplify the phonetics of spoken speech, and thus implement only those attributes that are actively controlled by the speaker. Our results show that compared to spoken forms, whistling transposes the basic strategies used in normal speech to convey gemination contrast. As for normal speech, duration is used as the basic correlate that distinguishes singletons from geminates, the latter being significantly longer than the former. The longer duration for geminates was observed both in medial and final positions. Given that whistled stops translate acoustically into silent intervals, no durational differences could be observed in word-initial position. This mirrors the absence of durational differences between singletons and geminates in absolute phrase-initial position for voiceless stops in normal speech [12]. Here too, closure duration for spoken voiceless stops cannot be measured, as no direct signal of the duration of these segments is visible in the acoustic waveform and spectrogram.

Interestingly, supplementary secondary cues are also conveyed in whistling. Vowels are thus shorter before geminates both in intervocalic and final positions. Similarly, whistled geminates display higher H1 frequency values at C-V transition in initial and intervocalic positions. The fact that these supplementary correlates are also present in whistled speech raises the question of whether they are actively controlled by the speakers/whistlers and thus exploited as additional cues to gemination.

The higher H1 frequency at C-V transition in prevocalic position presumably transposes the higher amplitude of the release in geminate stops in normal speech. As already stated, this attribute can be the sole distinctive cue to gemination for spoken utterance-initial voiceless stops, since closure duration cannot be acoustically implemented. The higher amplitude of the release in spoken forms is generally considered to be an automatic outcome of the longer duration of the stops. That is, the higher air pressure rise behind the closure after long consonants results automatically in higher or stronger amplitude of the release [23]. This aerodynamic account cannot apply for our whistling data, suggesting that this attribute is probably actively controlled and does not solely result from implementational effects of the primary duration attribute. Higher H1 frequency values for whistled geminates may be viewed as an enhancing correlate, which increases the perceptual distance between singletons and geminates. This correlate is probably computed online as opposed to the defining attribute. It is introduced precisely because word-initial context puts the defining attribute in jeopardy, as duration is not perceptually recoverable in this position. This also implies that speakers/whistlers have a tacit knowledge of the physical pressures that shape lexical forms.

The shorter vowel duration before geminates has been observed in various unrelated languages, such as Bengali, Italian, or Norwegian (see [12] for a review). Different hypotheses have been provided to account for this shortening, including structural accounts related to the phenomenon of closed syllable shortening [24]. This shortening could also be accounted for on perceptual grounds. The fact that it is also transposed in whistling suggests that it may have a functional load. In line with [25], this shortening may be produced in order to enhance the length contrast on a geminate segment. This shortening can also take on a distinctive feature in case duration cannot be recovered from the acoustic signal, as for example for non-released stops in word-final position.

Enhancement Theory offers a basis for accounting for the variable acoustic attributes defining the singleton/geminate contrast in both normal and whistled speech. Starting from the observation that languages tend to preserve useful contrasts, the account adopted proposes that supplementary features may be marshaled to reinforce existing contrasts between two sounds along an acoustic dimension that distinguishes them. Once deliberately introduced, these features tend to survive, and may eventually supplant the feature which they originally served to enhance. This is the case for the higher H1 frequency values at C-V transition in initial position and for the shorter vowel duration preceding non-released stops in word-final position.

## 5. References

[1] J. Meyer, "Whistled Languages. A Worldwide Inquiry about Whistled Speech", Berlin: Springer, 2015.

[2] A. Classe, “Phonetics of the Silbo Gomero”, *Archivum Ling*. 9, pp. 44-61, 1956.
[3] R-G. Busnel, and A. Classe, “Whistled Languages”, Berlin: Springer, 1976.
[4] A. Rialland, “Phonological and phonetic aspects of whistled languages”, *Phonology* 22, pp. 237–271, 2005.
[5] J. Meyer, “Acoustic strategy and typology of whistled languages; phonetic comparison and perceptual cues of whistled vowels”. *Journal of the International Phonetic Association,* 38, pp. 69-94, 2008.
[6] C.H. Shadle, “Experiments on the acoustics of whistling”, *The Physics Teacher* 21(3), pp. 148-154, 1983.
[7] J. Meyer, B. Gautheron, and R. Ridouane, “Whistled Moroccan Tamazight: phonetics and phonology”, *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, 2015.
[8] R. Ridouane, “Leading issues in Tashlhiyt Berber phonology”, *Language and Linguistics Compass* 10, pp. 644-660. DOI: 10.1111/lnc3.12211, 2016.
[9] N. Louali, and G. Puech. “Les consonnes ‘fortes’ du berbère: indices perceptuels et corrélats phonétiques”, In Actes des 20e Journées d’Etude sur la Parole, pp. 459–464, 1994.
[10] O. Ouakrim, “Un paramètre acoustique distinguant la gémination de la tension consonantique”, *Etudes et Documents Berbères* 11, pp. 197–203, 1994.
[11] R. Ridouane, “Suites de consonnes en berbère: Phonétique et phonologie”, Thèse de Doctorat Unifié, Université Paris 3, 2003.
[12] R. Ridouane, “Geminates at the junction of phonetics and phonology”, *Papers in Laboratory Phonology* 10, pp. 61–90, 2010.
[13] R. Ridouane, and P. Hallé, “Word-initial geminates: from production to perception”, In H. Kubozono (ed.), *Phonetics and Phonology of Geminate Consonants*, Oxford Studies in Phonology and Phonetics - Oxford University Press, pp. 66-84. 2017.
[14] N. Chomsky, and M. Halle, “The Sound Pattern of English”, NewYork: Harper and Row, 1968.
[15] M. J. Solé, and J.J. Ohala, “What is and what is not under the control of the speaker: intrinsic vowel duration”, *Papers in laboratory phonology*, *10*, pp. 607-655, 2010.
[16] K.N. Stevens, and S.J. Keyser, “Primary features and their enhancement in consonants”, *Language* 65.1, pp. 81-106, 1989.
[17] S. J. Keyser, and K.N. Stevens, “Enhancement and overlap in the speech chain”, *Language* 82.1, pp. 33-63, 2006.
[18] D.H. Whalen, B. Gick, M. Kumada and K. Honda, “Cricothyroid activity in high and low vowels: exploring the automaticity of intrinsic F0”, *Journal of Phonetics* 27(2), pp. 125-142, 1999.
[19] P. Hoole, and K. Honda, “Automaticity vs. feature-enhancement in the control of segmental F0.” *Where do phonological features come from*, pp. 131-171, 2011.
[20] P. Boersma, and D. Weenink, “Praat: doing phonetics by computer” [Computer program], Version 6.0.37, retrieved 3 February 2018 from http://www.praat.org/, 2018.
[21] R. Baayen, and R. Harald, “Analyzing Linguistic Data. A Practical Introduction to Statistics using R”, Cambridge University Press, Cambridge, UK, 2008.
[22] R Development Core Team: A language and environment for statistical computing, R Foundation for Statistical Computing: Vienna (www.r-project.org), 2008.
[23] J.J. Ohala, “The origin of sound patterns in vocal tract constraints”, In MacNeilage, P. (ed.), The Production of Speech, New York: Springer, pp. 189–216, 1983.
[24] I. Maddieson, Phonetic cues to syllabification. In V. Fromkin (ed.). Phonetic Linguistics. Essays in Honor of Peter Ladefoged. New York: Academic Press, pp. 203–221, 1985.
[25] R. Kluender, K. Randy, L. Diehl, and R. Wright, “Vowel-length differences between voiced and voiceless consonants: an auditory explanation”, *Journal of phonetics* 16, pp. 153-169, 1988.