INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

The Influence on Realization and Perception of Lexical Tones from Affricate's Aspiration

Chong Cao, Yanlu Xie, Qi Zhang, Jinsong Zhang

Beijing Language and Culture University, China
caochong2013@126.com, xieyanlu@blcu.edu.cn, zhangqiyts@126.com,
jinsong.zhang@blcu.edu.cn

Abstract

Consonants in /CV/ syllables usually have potential influence on onset fundamental frequency (i.e., onset f0) of succeeding vowels. Previous studies showed such effect with respect to the aspiration of stops with evidence from Mandarin, a tonal language. While few studies investigated the effect on onset f0 from the aspiration of affricates. The differences between stops and affricates in aspiration leave space for further investigations. We examined the effect of affricate's aspiration on the realization of onset f0 of following vowels in the form of isolated syllables and continuous speech by reference to a minimal pair of syllables which differ only in aspiration. Besides, we conducted tone identification tests using two sets of tone continua based on the same minimal pair of syllables. Experimental results showed that the aspirated syllables increased the onset f0 of following vowels compared with unaspirated counterparts in both kinds of contexts. While the magnitude of differences varied with tones. And the perception results showed that aspirated syllables tended to be perceived as tones that have relative lower onset f0, which in turn supported the production result. The present study may have applications for speech identification and speech synthesis.

Index Terms: affricate, aspiration, onset f0, production and perception

1. Introduction

Standard Mandarin, a tonal language, uses four tones to distinguish lexical meanings, e.g., ma1 "mother" [Tone 1], ma2 "hemp" [Tone 2], ma3 "horse" [Tone 3], ma4 "scold" [Tone 4]. These four tones are primarily signaled with different fundamental frequency (i.e., f0) patterns, i.e., Tone 1 with a high-level f0 contour, Tone 2 with a mid-rising f0 contour, Tone 3 with a low-dipping f0 contour, and Tone 4 with a high-falling f0 contour [1], as illustrated in figure 1. In addition to these four full lexical tones, there is also a neutral tone which is usually referred to as Tone 5.

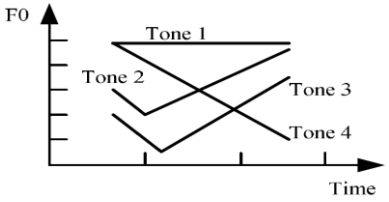

Figure 1: *Distinctive f0 patterns of Chinese four tones*.

Consonants in /CV/ syllables usually affect the onset fundamental frequency (i.e., onset f0) of following vowels, leading to local perturbations [2]. As an example, it is well documented that the onset f0 following voiceless stops is higher than following voiced stops [3-5]. In Standard Mandarin, consonants are more properly characterized as differing according to the aspirated/unaspirated contrast rather than the voiceless/voiced contrast [6]. The aspiration divides stops and affricates into two groups: voiceless aspirated and voiceless unaspirated, which are associated with different phonemes. This contrast is distinctive and phonemic in Standard Mandarin. For example, /p/ in the word 饱/pau214/ ("full") and /p'/ in the word 跑/p'au214/ ("run") are two different phonemes as these two words carry different lexical meanings.

Previous studies have shown the effect of the aspiration on the onset f0 in speech production. For instance, [7] reported that there were significant differences in the onset f0 between unaspirated consonants and corresponding aspirated consonants in continuous speech; [8] further suggested that the unaspirated consonant gave rise to a higher onset f0 of the following vowel than the aspirated cognate. This assumption was on the one hand supported by some data from various dialects in Mandarin, such as Cantonese [9, 10], Wu, Gan [11, 12]. It was on the other hand demonstrated in multiple languages (e.g., English [2], Danish [13], Korean [14], Thai [15]).

The magnitude of consonant-related perturbations of onset f0 was usually small in tonal languages (e.g., Standard Mandarin) [17]. Thus it has been argued that whether tonal language speakers use such intrinsic effect on onset f0 as a cue for tone perception? It was found that tones carried by syllables with aspirated stops were inclined to be perceived as relative lower tones (e.g., Tone 3), while tones of syllables with unaspirated counterparts tended to be perceived as higher tones (e.g., Tone 4) [16].

Previous studies concerning the effect of aspiration on the onset f0 of following vowels have mainly focused on stops. Apart from stops, affricates also have the aspirated/unaspirated contrast in Standard Mandarin. Whether the aspiration of affricates affects the realization of the onset f0, to our knowledge, was less studied. Compared with stops, affricates are an intermediate category between simple stops and a sequence of a stop and a fricative; the release of the constriction of affricates is modified as to produce a more prolonged period of frication after the release [17]. This kind of differences in articulation provides a chance for us to investigate aspiration's effects of affricates on the onset f0 of following vowels.

We conducted a production experiment and a perception experiment. In the production experiment, we calculated the

http://dx.doi.org/10.21437/Interspeech.2017-1267

onset f0 values of following vowels in the form of isolated syllables and continuous speech by reference to a minimal pair of syllables with affricates which differ only in aspiration. In the perception experiment, we conducted tone identification tests using two sets of tone continua based on the same minimal pair of syllables.

We aim to address two questions:

- How aspiration of affricates affect following vowel's onset f0 in speech production? To raise, lower or remain unchanged?
- Whether such consonant-related perturbations of onset f0 from affricate's aspiration play a role in tone perception?

2. Method

2.1. Production –isolated syllables

2.1.1. Stimuli

The minimal pair of /tʂ‘/-/tʂ/ was chosen as the target affricates of the present study. To try to limit sources of error, only syllables with single vowels were used, without nasal codas or Medials. Finally, we chose /tʂ‘ɿ/ and /tʂɿ/ as the target stimuli. Apart from the target stimuli, as supplement data, other syllables which included different voiceless Initials and Finals were also used for the purpose of calculating speakers' pitch ranges. All syllables could carry each of four Mandarin tones to construct meaningful syllables. In total, we collected seven native speakers' recordings of isolated syllables from the Chinese Interlanguage Corpus in Beijing Language and Culture University and 48 syllables for each speaker.

2.1.2. F0 annotation, measurement and normalization

The onset and offset f0 of each syllable were manually annotated by a phonetically trained listener, checked and corrected by a post PhD. Both of them have received professional phonetic training. The first period of each syllable was thrown out if it was apparently not a full cycle. Figure 2 is an annotation sample of the onset f0 of target syllable /tʂ‘ɿ/ with Tone 1.

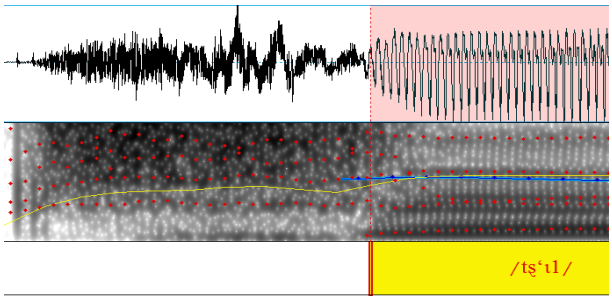

Figure 2: *A annotation sample of onset f0.*

To determine the f0 contour of each syllable, we made f0 measurements at 10 time points for each vowel including the onset f0 and offset f0. A total of 3360 (48 syllables × 10 tokens ×7 speakers) tokens were measured, giving an average of 480 tokens per speaker.

It is universally acknowledged that no two speakers share the same pitch ranges because of physiological or anatomical factors [18]. Therefore, the absolute f0 values of different speakers are not directly comparable. In this study, z-score transform was utilized to normalize f0 data. Extracted f0 values were normalized by subtracting a speaker's mean f0 value (i.e., μ) across all f0 tokens, and then dividing by the standard deviation (i.e., σ) across all f0 tokens for that speaker. The formula was as follows:

$$z = \frac{x - \mu}{\sigma} \quad (1)$$

2.2. Production –continuous speech

2.2.1. Stimuli

The same 48 syllables including target syllables /tʂ‘ɿ/ and /tʂɿ/ were incorporated into a carrier sentence, "yi2 ge4 X zi4" (one character "X") where "X" represents each of 48 syllables written in the characters systems. This carrier sentence was selected from [19]. A variety of studies suggested that word occurrence frequencies could affect tone production [20]. To avoid such influence, the characters of 48 syllables were comparable to each other in occurrence frequencies. The frequency counts were based on the JunDa Modern Chinese frequency statistics. This corpus counted the occurrences of words in newspapers, scientific materials, and various types of fiction, with a total of approximately 1.8 million words.

2.2.2. Participants

Ten female native speakers of Mandarin participated as subjects. Their ages ranged from 23 to 26 years old. None of them have language, hearing or speaking impairments.

2.2.3. Procedure

The recording was conducted in a soundproof booth. Prior to recording, a set of instructions was presented to participants, directing them to read the sentences aloud at a normal speaking rate and with stress on the target words. Each sentence was presented to subjects five times and the order of repetitions was randomized. The speech was recorded with 16 kHz sampling rate and 16-bit accuracy.

2.2.4. F0 annotation, measurement and normalization

The f0 annotation, measurement and normalization work of 48 syllables were just the same as above (i.e., production-isolated syllables).

2.3. Tone Perception

2.3.1. Stimuli

In the following, four lexical tones were labeled as T1, T2, T3 and T4 respectively. Two sets of tone continua (i.e., T1_T2, T1_T4) based on /tʂ‘ɿ/ and /tʂɿ/ were synthesized. The motivation to choose these two sets of tone continua was that they differ in onset f0 values.

T1 was chosen as the initial endpoint both for T1_T2 and T1_T4 continuum. Each continuum proceeded through ten steps from one endpoint to the other, which were derived from a native female Mandarin speaker. The intermediate contours were obtained via interpolation between endpoints. Each token's vocalic duration was normalized to 250 msec. For T1_T2 continuum, we chose the point whose length is in the 25% of the f0 curve as the turning point. Figure 3 illustrates the tone continua pattern. The X-axis represents the

normalized time of each token, Y-axis represents f0 values. For T1_T2 continuum, we changed the starting point and turning point simultaneously, while kept the ending point unchanged; for T1_T4 continuum, we changed the starting point and the ending point simultaneously.

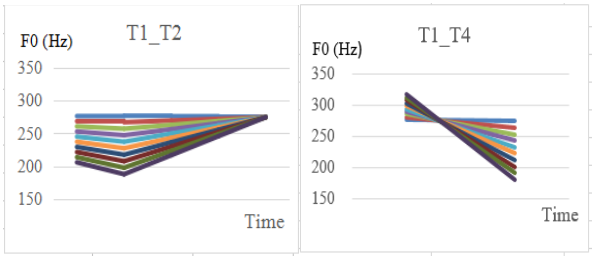

Figure 3: *Tone continua pattern.*

2.3.2. Participants

Eighteen participants from Beijing Language and Culture University participated in the perception experiment. Their ages ranged from 22 to 30 years old. None of them have language, hearing or speaking impairments.

2.3.3. Procedure

A set of instructions written in Mandarin characters was presented to participants. They were asked to identify the tone of each stimulus with a forced choice between the two endpoint tones for each continuum. Each stimulus was repeated 3 times which made a total of 120 trials (2 syllables × 10 steps × 3 repetitions × 2 tone continua) presented in quasi-random order. The formal identification task was preceded by a series of practice trials.

3. Results

3.1. Production-isolated syllables

The f0 contours of /tʂʻɹ/ and /tʂɹ/ with four lexical tones were then normalized across speakers using a z-score transform, as illustrated in figure 4. The X-axis represents the number of f0 tokens, the Y-axis represents the z-score value of f0. Numbers 1-4 in the legend stand for four lexical tones successively. The solid lines represent aspirated syllable /tʂʻɹ/, the dotted lines represent unaspirated syllable /tʂɹ/. Each curve represents an average across seven speakers.

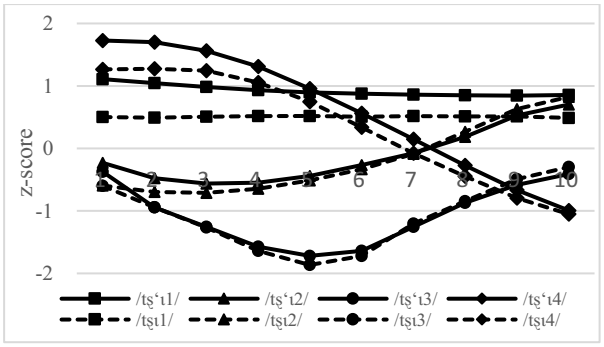

Figure 4: *F0 contours with four lexical tones.*

It can be seen from figure 4 that, f0 contours of /tʂʻɹ/ and /tʂɹ/ with four lexical tones are very similar to that illustrated in figure 1, as T1 is high-level, T2 mid-rising, T3 low-dipping and T4 high-falling. The aspirated syllable /tʂʻɹ/ seems to be produced with slightly higher onset f0 with each of four lexical tones compared with unaspirated syllable /tʂɹ/. While the magnitude of differences varies with tones. For example, the difference with T1 is much greater than that with T2/T3.

We calculated the onset f0 values and took them as the dependent variable in subsequent statistical analysis of the data. Results of Paired-Sample T tests showed that there existed significant differences at a $p<0.05$ level between /tʂʻɹ/ and /tʂɹ/ in T1 condition ($t=4.445$, $p=0.004$), T2 condition ($t=4.433$, $p=0.004$) and T3 condition ($t=1.824$, $p=0.048$). However, no significant differences were found in T4 condition ($t=1.824$, $p=0.118$).

3.2. Production- continuous speech

Figure 5 depicts the normalized f0 contours of /tʂʻɹ/ and /tʂɹ/ with each of four lexical tones, just as in figure 4.

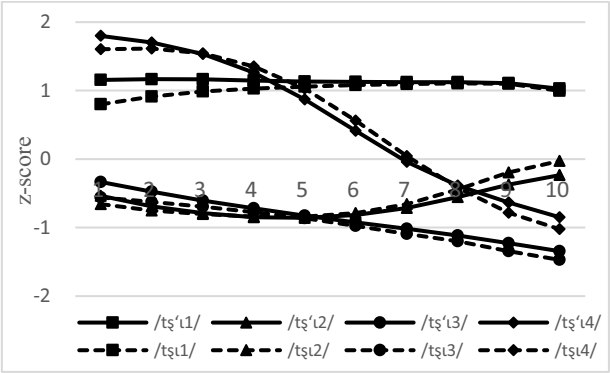

Figure5: *F0 contours with four lexical tones.*

It can be seen from figure 5 that, f0 contours of T1, T2 and T4 are very similar to that illustrated in figure 1. While T3 is a little different, it just fall from the moderate starting level without the final rise which was called as 'half third tone' [1]. Evidently, the onset f0 value is higher for aspirated syllable /tʂʻɹ/ than for unaspirated syllable /tʂɹ/ with each of four lexical tones. The difference in T1 is greater than that with T2 and T3 too. Results of Paired-Sample T tests showed that there existed significant differences at a $p<0.01$ level between /tʂʻɹ/ and /tʂɹ/ across all four lexical tones.

3.3. Tone Perception

Identification accuracy and category boundary locations were assessed by means of probit analyses of individual identification curves. Figure 6 a-b illustrate the identification curves of target syllables /tʂʻɹ/ and /tʂɹ/ across two sets of tone continua respectively.

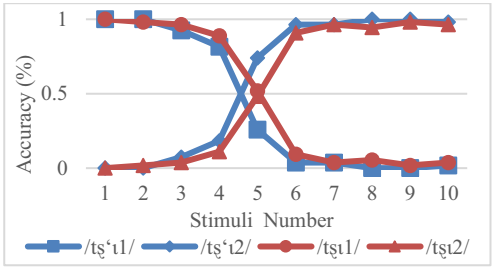

Figure 6a: *Identification curves of T1_T2 continuum.*

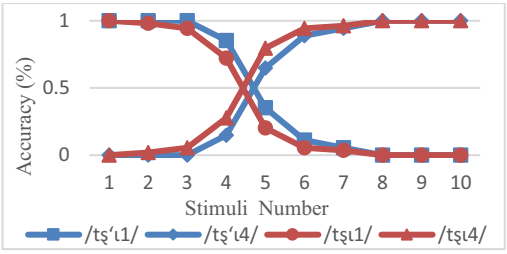

Figure 6b: *Identification curves of T1_T4 continuum.*

In figure 6 a-b, the X-axis represents stimuli number, the Y-axis represents the accuracy rate of T1. As we can see, for T1_T2 continuum, the category boundary of /tʂʻɿ/ is between stimulus 4 and 5, while the category boundary of /tʂɿ/ falls at a later point, about between stimulus 5 and 6. For T1_T4 continuum, the category boundary of /tʂʻɿ/ is very close to the stimulus 5, while the category boundary of /tʂɿ/ is close to the previous stimulus (i.e., stimulus 4). Table 1 lists the category boundaries of /tʂʻɿ/ and /tʂɿ/ in two sets of tone continua. The values in the parentheses denote standard errors.

Table 1: *Category boundaries of two continua.*

continua	/tʂʻɿ/	/tʂɿ/
T1_T2	4.58(0.68)	5.09(0.61)
T1_T4	4.85(0.84)	4.45(0.78)

It can be seen from table 1 that, for T1_T2 continuum, the category boundary of /tʂʻɿ/ is smaller than /tʂɿ/, which means that /tʂʻɿ/ tended to be perceived as T2 that has a relative lower onset f0 compared with T1. For T1_T4 continuum, the category boundary of /tʂʻɿ/ is larger than /tʂɿ/, which means that /tʂʻɿ/ tended to be perceived as T1 that has a relative lower onset f0 compared with T4. Results of Paired-Sample T tests taking the category boundary as the dependent variable showed that there were significant differences at a $p<0.01$ level between /tʂʻɿ/ and /tʂɿ/ both in T1_T2 continuum ($t=-3.517$, $p=0.003$), and in T1_T4 continuum ($t=3.709$, $p=0.002$).

4. Discussion

The production results of isolated syllables and continuous speech suggested that the onset f0 following aspirated syllables with affricates was higher than that following unaspirated counterparts. Conversely, the perception result suggested that tones carried by aspirated syllables tended to be perceived as those tones which have relative lower onset f0 values (i.e., T2 in T1_T2 continuum, T1 in T1_T4 continuum), which in turn supported the production result. While the magnitude of differences was related to the tone itself. When the tone was high-level (i.e., T1), the difference between aspirated and unaspirated syllables with affricates was greater than when the tone was mid-rising (i.e., T2) or low-dipping (i.e., T3).

There are differences in opinions about the aspiration's effect of stops on onset f0 values. For example, [7] indicated the higher onset f0 following aspirated stops than following unaspirated stops, which was consistent with our production result. But [8] reported a contrary result which suggested that the onset f0 was higher following unaspirated stops than following aspirated stops. The contexts of target syllables in [8] were more complex: the target syllables were combined into disyllabic words which were incorporated into two carrier sentences. It is feasible that the tonal co-articulation gave rise to the disagreement. Our perception result was in agreement with [16] which found that tones carried by aspirated syllables with stops tended to be perceived as those tones which have relative lower onset f0 values.

It was argued that the difference between aspirated and unaspirated consonants was that, the state of vocal folds for aspirated consonants is the same as that of voiceless consonants at the time of articulatory release or immediately after, while the state of vocal folds for unaspirated consonants is the same as voiced consonants at the same time or immediately after [21]. In addition, in Swedish, when aspiration occurs, it serves as one of cues for the distinction between voiced and voiceless stops, since voiced consonants are invariably unaspirated and voiceless consonants are invariably aspirated [22]. According to the above two points: we may tentatively speculate that the effect of the aspirated/unaspirated contrast on onset f0 would be the same as the voiceless/voiced contrast. Voiceless stops often have a higher onset f0 than voiced stops [3-5], thus, aspirated syllables with affricate have a higher onset f0 than unaspirated counterparts too.

Pressure drop across the glottis is a very crucial indicator of voicing. Before the release of consonants, the air pressure of unaspirated consonants in the oral cavity is generally higher than that of aspirated consonants, because unaspirated consonants have longer closure duration [23]. At the time of articulatory release, for unaspirated consonants, the articulation explosion takes place at the instant when the glottis becomes completely closed, or immediately before, the vibration of the vocal folds in the succeeding vowel portion starts immediately after the explosion, thus there is little air flowing out of the subglottal area leading to little decrease of subglottal pressure. However, for aspirated consonants, the longer voice onset time and the larger peak glottal opening give rise to a considerable period of aspiration which results in much decrease in subglottal pressure [24]. Some studies which suggested that the onset f0 following aspirated consonants was lower observed the differences in subglottal pressure between aspirated and unaspirated consonants at the release, while may neglect the differences in oral pressure before the release.

5. Conclusions

By conducting production and perception experiments, we found that the onset f0 following aspirated affricates was higher than that following unaspirated affricates. The present study only discussed aspiration's effects in isolated syllables and carrier sentences. Further examination of the effects of tonal co-articulation and more complex contexts on onset f0 is needed.

6. Acknowledgements

This work is supported by funds of Advanced Innovation Center for Language Resources and Intelligence, the Special Program for Key Basic Research fund of Beijing Language and Culture University (the Fundamental Research Funds for the Central Universities)(16ZDJ03) and the Research Funds of Beijing Language and Culture University（16YCX220）. The asterisked author is the corresponding author.

7. References

[1] Y. R. Chao, *A grammar of spoken Chinese*. Berkeley: University of California Press, 1968.

[2] J. M. Hombert, *Consonant types, vowel quality, and tone*, *Tone: A linguistic survey*. Academic Press, 1978, pp. 77-111.

[3] M. Haggard, S. Ambler, and M. Callow, "Pitch as a voicing cue," *The Journal of the Acoustical Society of America*, vol. 47, no. 2B, pp. 613-617, 1970.

[4] A. S. House and G. Fairbanks, "The influence of consonant environment upon the secondary acoustical characteristics of vowels," *The Journal of the Acoustical Society of America*, vol. 25, no. 1, pp. 105-113, 1953.

[5] D. H. Whalen, "Coarticulation is largely planned 7/3," *Journal of Phonetics*, vol. 18, pp. 3-35, 1990.

[6] B. R. Huang and X. D. Liao, *Modern Chinese*. Beijing: Higher Education Press, 2010.

[7] S. Chilin. "Generation and Normalization of Tonal Variations," *Journal of Chinese Linguistics Monograph Series*, 2001, pp. 32-52

[8] C. X. Xu and Y. Xu, "Effects of consonant aspiration on Mandarin tones," *Journal of the International Phonetic Association*, vol. 33, no. 2, pp. 165-181, 2003.

[9] E. Zee, "The effect of aspiration on the F0 of the following vowel in Cantonese," *UCLA Working Papers in Phonetics*, vol. 49, pp. 90-97, 1980.

[10] A. L. Francis, V. Ciocca, and V. K. M. Wong, "Is fundamental frequency a cue to aspiration in initial stops?" *The Journal of the Acoustical Society of America*, vol. 120, no. 5, pp. 2884-2895, 2006.

[11] F. Shi, "The influence of aspiration on tones," *Journal of Chinese Linguistics,* vol. 26, no. 1, pp. 126-145, 1998.

[12] L. King, H. Ramming, and L. Schiefer, "Initial F0-contours in Shanghai CV-syllables: an interactive function of tone, vowel height, and place and manner of stop articulation," in *ICPhS 1987- 11th International Congress of Phonetic Sciences*, *August 1-7, Tallinn, Estonia, Proceedings* 1987, pp. 154-157.

[13] V. Jeel, "An investigation of the fundamental frequency of vowels after various Danish consonants, in particular stop consonants," *Annual Report of the Institute of Phonetics, University of Copenhagen*, vol. 9, pp. 191-211, 1975.

[14] R. S. Weitzman and M. S. Han, "Acoustic Features in the Manner Differentiation of Korean Stop Consonants," *The Journal of the Acoustical Society of America*, vol. 40, no. 5, pp.1272-1273, 1966.

[15] J. Gandour, "Consonant types and tone in Siamese," *Journal of phonetics*, vol. 2, pp. 337-350, 1974.

[16] Y. F. Yang and L. J. Jin, "Stop consonants and tone perception," *Acta Psychologica Sinica*, no. 3, pp.236-242, 1988.

[17] P. Ladefoged and I. Maddieson, "The sounds of the world's languages," *Language*, vol. 74, no.2, pp. 374-376, 1998.

[18] H.Q. Bao and M. C. Lin, *Essentials of experimental phonetics*. Beijing University Press, 1994.

[19] P. A. Hallé, Y. C. Chang, and C. T. Best, "Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners," *Journal of Phonetics*, vol. 32, no.3, pp. 395-421, 2004.

[20] Y. Zhao and D. Jurafsky, "The effect of lexical frequency and Lombard reflex on tone hyperarticulation," *Journal of Phonetics*, vol. 37, no. 2, pp. 231-247, 2009.

[21] P. Ladefoged, *Preliminaries to linguistic phonetics*. Chicago: University of Chicago Press, 1971.

[22] A. Löfqvist, "Interarticulator programming in stop production," *Journal of Phonetics*, vol. 8, pp. 475-490, 1980.

[23] R. Kagaya and H. Hirose, "Fiberoptic electromyographic and acoustic analyses of Hindi stop consonants," *Annual Bulletin, Research Institute of Logopedics and Phoniatrics*, vol. 9, pp. 27-46, 1975.

[24] P. Ladefoged, "Some physiological parameters in speech," *Language and speech*, vol. 6, no. 3, pp. 109-119, 1963.