UNRELEASED PLOSIVE CONSONANTS IN THAI: AN ACOUSTIC STUDY

Paula Alejandra CANO CÓRDOBA[1], Thi Thuy Hien TRAN[1], Nathalie VALLÉE[1], Christophe SAVARIAUX[1], Silvain GERBER[1], Nicha YAMLAMAI[2], Morgane ROSSIGNOL[1]

[1] Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France
[2] Faculty of Arts, Univ. Silpakorn, Bangkok, Thailand
{paula.cano-cordoba, thi-thuy-hien.tran, nathalie.vallee, christophe.savariaux}@gipsa-lab.grenoble-inp.fr; yamlamai_n@su.ac.th; morgane.rossignol@etu.univ-grenoble-alpes.fr

ABSTRACT

Stops are consonants usually produced with an audible noise generated by the sudden release of the complete closure of the vocal tract. However, published studies have reported that in Thai, syllable-final unvoiced stops are frequently produced without an audible release. Our aim is to provide acoustic details of such consonantal productions. By analysing acoustic data from 10 native Thai speakers and by focusing on vowel-consonant transitions, we propose a set of acoustic cues on which listeners might rely on to discriminate and categorize the unreleased stop allophones. Our study also investigates whether acoustic properties of non-released stops differ depending on whether they are at the end of monosyllables or at the end of the penult syllable of disyllabic compounds (end-of-word coda *vs* word-internal syllable coda). Our findings show consistent differences in stop durations, formant transitions, and intensity curves according to place of articulation, within-word position and speaker gender.

Keywords: unreleased final stop, Thai, syllable boundary, VC transition, acoustic cues

1. INTRODUCTION

Plosive consonants occur in all languages [1], [2]. They belong to the category of stops which are produced with a total blockage of the airstream in the vocal tract. Because of the complete closure of the vocal tract and the increase in pressure behind the closure, the sudden release of the airstream generates an abrupt and salient noise [3] which is called burst [4]. The burst contains acoustic information that participates to the consonant identification [5], [6]. Nevertheless, in some languages such as Korean [7], Vietnamese [8] and Thai [9] [10], there is no audible release when plosives are syllable ending consonants. The syllable-final position is particularly subject to articulatory weakening and acoustic change or lenition [11]. Some articulatory weakening processes result in phonological neutralizations leading to a restriction of syllable-final consonantal inventories. This is the case in some East or Southeast Asian languages, as for example in Korean or Vietnamese. In Korean, the three-way manner contrast /t/ (lenis) ~ /t*/ (fortis) ~ /tʰ/ (aspirated) is neutralized in coda position and has the same surface phonetic representation [t̚], an unvoiced and unreleased stop [12]. In Vietnamese, [t̚] is an allophone of /t/ and /d/ whose contrast is no longer maintained at syllable end [13]. Thai is phonologically classified as a monosyllabic language, but it is also partially polysyllabic lexically since 35% of Thai words are two-syllable semantic compounds, and 25% are words longer than 2 syllables [14]. Thai has 21 phonemic consonants. Among them there is a set of 8 plosives that can all appear in syllable-initial position /p, pʰ, b, t, tʰ, d, k, kʰ/ contrasting 3 places of articulation (bilabial, coronal, velar) and 3 laryngeal modes (voicing (except for the velar place), non-voicing and aspiration) [15]. However, due to a neutralization process which gives rise to onset-coda asymmetries in Thai [16], only 3 surface phonetic plosives [p̚, t̚, k̚] described as unvoiced and unreleased are found in syllable final position [10].

Several studies have shown interest in unreleased stops mainly because native speakers of these languages are able to easily identify places of articulation despite the absence of burst [8], [17], [18]. This suggests that other acoustic cues than audible release provide robust information for successful perception and categorization of syllable-final plosives. For example in Vietnamese, [19] show that spectral features of the second half of vowel duration contain information on coda's place of articulation, which can guide listeners' processes of identification of unreleased stops. [20] observed in Thai, like [19] in Vietnamese, that vowel duration is influenced by the place of articulation of the syllable-final plosive. In a cross-language perception study, [18] showed that Thai speakers perform better than native English speakers in the identification of Thai and Korean unreleased plosives. Furthermore, in Vietnamese, acoustic features of the syllable's rhyme shows differences depending on whether the unreleased plosive is an end-of-word coda or a word-internal syllable coda [19].

The purpose of our study is to highlight acoustic cues of unreleased plosives in Thai in examining more specifically duration and spectral characteristics of vowel-consonant (VC) transitions as in previous studies cited above.

2. HYPOTHESES

Based on results of published studies mentioned in Section 1, we suggest that acoustic differences in VC transitions in Thai, where C is an unreleased plosive, existed (1) according to the place of articulation of the plosive, and (2) depending on whether the unreleased stop is in word-medial (i.e. coda of the first syllable of a dissyllabic word) or in word-final position (in case of monosyllabic word). We target the first three formants ($F_{1\text{-}3}$), the fundamental frequency (f_0), intensity (I), and the acoustic duration of the rhyme.

3. METHODOLOGY

3.1. Corpus and speakers

Thirty-four Thai words, 17 C_1aC_2# monosyllables and 17 $C_1aC_2.C_3VC$# dissyllables (the dot indicates the syllable boundary), were selected such as each target syllable occurred as a monosyllabic word and as the first syllable in a dissyllabic word. All target syllables carried the high rising tone. For compounds' second syllable, tones are chosen (high rising for syllables with final /p t k/, mid falling for those with final sonorants) to minimise the tonal frequency gap. The low vowel /a/ was chosen for better contrast with the target stops. C_2 could be /p/, /t/ or /k/. Consonants in C_3 were chosen voiced and could be /l/, /d/, /j/, /b/ or /w/ to aid C_2 segmentation, except for ซับซ้อน [sap˧.sɔ:n˧] *'complicated'* (no words with a voiced C_3 were found with [sap˧] in the Thai lexicon, due to the high rising tone constraint). Finally, the 34 selected words were inserted into the carrier sentence พูดว่า __ ดีๆ [pʰu:t˥ wa:˥ __ di:˩ di:˩] *'Say __ carefully'*.

Data from five male and five female native Thai speakers were analysed. The participants were from the Bangkok area and spoke the central Thai dialect. At the time of the study, they were 20 years old and were all third-year students in the French Department at Thammasat University (Rangsit Campus). During the experiment, the participants were seated in front of a 22" screen on which the sentences were presented to them one by one. They were instructed to read them aloud at a normal and fluent rate. The recording took place in a sound proof-room, with a Marantz PMD 670 recorder, an AKG C-1000S microphone. The corpus was digitized at 44.1 kHz on 16 bits mono in .wav format. The results presented below are from 34 target words ∗ 3 repetitions ∗ 10 speakers giving 1,020 occurrences in C_2 of the syllable-final target plosives (300 /p/, 360 /t/, 360 /k/).

3.2. Measurements and analysis

The acoustic signal was manually segmented and annotated with Praat [21] in terms of phonemic, phonetic and sub-phonetic labels. When C_2 did not correspond to a production of the phoneme, it was discarded. In total, 9 realizations were excluded (4 for /k/ and 5 for /t/). Moreover, we detected a burst in 19.7% of the plosives produced, of which 9.7% were in word-final position, and 10% were in word-internal syllable coda. These productions were not included in our analyses.

Measurements of the following acoustic parameters were performed automatically using scripts we wrote in Praat software:

- Duration of the coda C_2 and duration of the preceding vowel /a/;
- Temporal evolutions of the first three formants (ΔF_1, ΔF_2, ΔF_3), of the fundamental frequency (Δf_0), and of the intensity (ΔI) measured from the time points of 40% and 50% of the whole vowel duration to the time points of 60%, 70%, 80%, 90% of the vowel length (which corresponds to the acoustic transition between /a/ and the stop closure).

Before proceeding to the statistical analyses, all duration measurements were normalized to the speaker's speech rate (in syllables per minute) following the method used by [19].

3.3 Statistical Analyses

Statistical analyses were performed using several mixed linear models (LMMs) from R software to estimate, on one side, the effects of consonant place, time-point, burst, and word type on ΔF_1, ΔF_2, ΔF_3, Δf_0, and ΔI mean values for each gender, and, on the other side, the effect of consonant place, burst and word type on duration, without distinction between gender. The *lme* function of the *nlme* package was used to construct LMMs. Once the model was established, contrast analyses were performed with the *glht* function of the *multcomp* package, according to the method presented by [22]. Multiple comparisons were performed for the dependent variables 'formants', 'f_0', and 'intensity'. We targeted the following differences:

- Between consonants (/p/, /t/, /k/) for each combination (word type - time point);
- Between the two-word types (simple and compound) for each combination (consonant - time point);
- Between two successive time points (e.g. between 50% and 60%, 60% and 70%, etc.) for each combination (consonant - word type).

4. RESULTS

In terms of statistical significance, the factor place of articulation has the strongest effect on the dependant variables, except on Δf_0. This important and general

result is in agreement with what had been reported in Vietnamese by [19] and is consistent with the general proposal that in case of unreleased stops, the perceptual-acoustic cues for identifying the place of articulation that enables listeners to categorize the consonant are located in the transient part between the vowel and the unreleased plosive [23].

4.1. Acoustic Cues for Articulation Places of Unreleased Final Consonants

An effect of place of articulation is present in segment duration regardless of word-type: [kˀ] is in average longer (21.3ms in C_2#, 17.9ms in $C_2.C_3$) than [pˀ] (15.8ms in C_2#, 15.2ms in $C_2.C_3$) and [tˀ] (19.8ms in C_2#, 15.8ms in $C_2.C_3$). However, significant differences are only found for /k/ *vs* /p/ in C_2# [z=3.5; p<0.01] and for /k/ *vs* /t/ in $C_2.C_3$ [z=3.3; p<0.05]. These observations are similar to those made on Vietnamese by [19] who found a significantly longer duration of [kˀ] relative to [pˀ] in monosyllables, and to [tˀ] in dissyllables. However, unlike [19], no significant differences in vowel duration were found in relation to the place of articulation of coda consonants in Thai.

Also, formant transitions show a consonant articulation place effect. F_1 values go down more for /t/ than for the bilabial and velar stops. The difference is found significant in monosyllables for the male speakers (/t/ *vs* /k/ from 80% [z=4.1; p<0.01] to 90% [z=3.9; p<0.05] and /t/ *vs* /p/ at 90% [z=4.2; p<0.01]), and in dissyllables for the female speakers (/t/ *vs* /k/ from 50% [z=3.6; p<0.05] to 90% [z=6; p<0.01]).

Three specific directions of F_2 transitions, which are consistent between speaker genders and word-types, are noted according to the stop place of articulation (Fig. 1 and 2). Differences are significant for /p/ *vs* {/t/, /k/}: /p/ *vs* /t/ from 60% and /p/ *vs* /k/ from 70% of the vowel duration, for both female and male speakers (p<0.01) and whatever the stop position in coda (C_2# or $C_2.C_3$) (p<0.01). Concerning /t/ *vs* /k/ differences, a significant F_2 deviation is found in monosyllables from 70% of the vowel duration for the male [z=4.3; p<0.01] and female speakers [z=-4.6; p<0.01], whereas in compound words, the deviation of F_2 is significantly different only for the female speakers at 90% of the vowel duration [z=-4.2; p<0.01]. F_2 transition for /k/ shows relatively stable values independently of speakers and word-types. However, it should be noted that we have not observed the expected theoretical ascending F_2 transition for /a/ with a rather steeper slope in velar context as proposed by [24] from synthetic speech.

Regarding F_3 transitions, significant differences are present between /t/ *vs* {/p/, /k/} for both female and male speakers, but with different parts of evolution depending on word-type. For female speakers, as proposed by [24] with synthetic speech, an ascending F_3 evolution in the second half of the vowel duration is noted for the coronal stop (from 3 to 93Hz), significantly attested from 60% until the end of the vowel in C_2# and from 80% in $C_2.C_3$. For male speakers, ΔF_3 of /t/ has a slight increase (C_2#: from 6 to 37Hz; $C_2.C_3$: from 6 to 26Hz). ΔF_3 of /t/ differs significantly from /p/ from 60% of the vowel duration independently of word-type (p<0.01), and deviates from /k/ remarkably from 60% to 80% in C_2# (p<0.05), but only at 60% in $C_2.C_3$ position [z=-3.5, p<0.05]. A difference in F_3 evolution between /p/ and /k/ is significantly attested at only 90% of the vowel duration for male speakers in compound words.

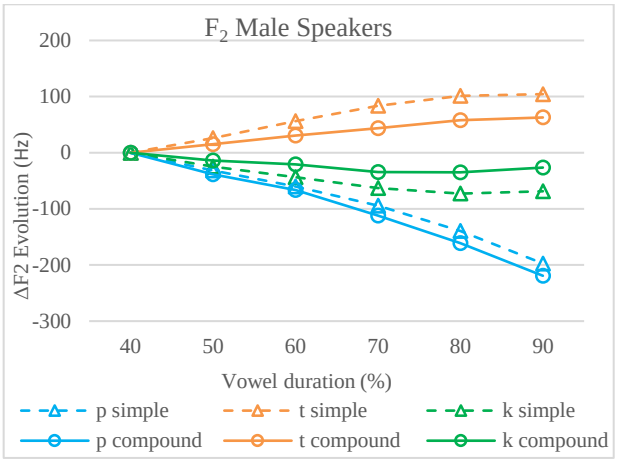

Figure 1: Temporal evolution of ΔF_2 estimated from 40% to 90% of the vowel duration (expressed in Hz) for male speakers as a function of word type (simple or compound).

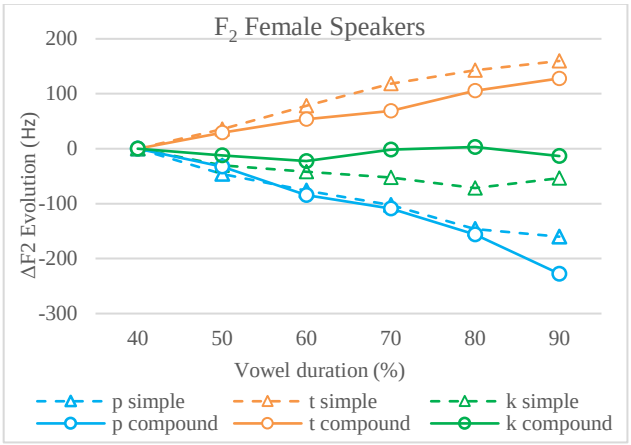

Figure 2: Temporal evolution of ΔF_2 estimated from 40% to 90% of the vowel duration (expressed in Hz) for female speakers as a function of word type (simple or compound).

No significant effect of place of articulation was found on Δf_0 values for both genders. This result is inconsistent with the observation of [19] on Vietnamese in which a significantly increase in Δf_0

from 70% until the end of the vowel was found before a non-released velar coda.

Our results also show that place of articulation impacts the temporal evolution of intensity. For all speakers, intensity significantly slopes down more for /k/ than for /p/ ($p<0.01$) and /t/ (♀: $p<0.05$; ♂: $p<0.01$) in $C_2.C_3$ from 70% of the vowel duration, and in C_2# for ♀ from 80% ($p<0.01$) and for ♂ only at 90% ($p<0.01$). For both word-types, the drop of ΔI for /k/ is significant from 40% until vowel endings ($p<0.01$). These results for ΔI are in agreement with what was observed in Vietnamese by [19].

4.2. Effect of Word Type on Final Stop Production

Analyses of our data show that the unreleased coronal stop /t/ is significantly longer in C_2# (20ms on average) than in $C_2.C_3$ (12ms) [$z=-3.6$, $p<0.01$], regardless of the speakers' gender. /k/ and /p/ are also longer in word-final position C_2# but the difference according to word types is not statistically proven (respectively $p=0.6$ and $p=1$). The duration of /a/ is also significantly different depending on whether the consonant is in C_2# or in $C_2.C_3$: /a_p/ (14.5ms *vs* 11.9ms, $z=-3$, $p<0.05$), /a_t/ (15.8ms *vs* 11.6ms, $z=-5$, $p<0.01$), or /a_k/ (16.2ms *vs* 12.8ms, $z=-3.8$, $p<0.01$). These findings partially confirm results of a previous two-speaker study on Thai [20], in which difference in consonant duration as a function of word type was found for stops.

An effect of word-type is found on F_1 transition but is different according to speaker gender. We observed in male speakers that, relative to 40% of the vowel duration (reference point), ΔF_1 of /p/ declines significantly more at 60% (-13Hz) in single words than in compound words (-9.6Hz) [$z=-3.7$, $p<0.05$], whereas in female speakers, ΔF_1 of /t/ decreases further in compounds from 50% (-13Hz) to 70% (-70Hz) of the vowel duration than in simple words (respectively from 1Hz to -49Hz) [$z=-4.2$; $p<0.01$]. These results are partially consistent with [19]s' findings for Vietnamese, which show that ΔF_1 drops significantly when the unreleased stop is an inner final syllable ending ($C_2.C_3$) rather than a word ending (C_2#). Unlike to F_1 transition, we found no effect of word-type on F_2 and F_3 transitions.

Our statistical analyses show significant differences for Δf_0 as a function of word type but only for female speakers. Indeed, Δf_0 curve decreases during the entire second half of the vowel duration for /t/, more in compound (from -1.7Hz to -6Hz) than in simple words (from -1.3Hz to -2Hz) ($p<0.01$). For /p/, Δf_0 curve is downward in compound words (from -1Hz to -2.5Hz) whereas it is slightly upward in monosyllables (from -1Hz to -0.06Hz). This Δf_0 difference is shown to be significant for /p/ ($p<0.01$). These results for the female speakers are consistent with those found in Vietnamese [19] which also reveal differences in the f_0 evolution from 60% until the end of the vowel depending on whether it is a simple word or a first syllable in a compound word. In terms of intensity, we did not observe an effect of word type on temporal evolution of ΔI that is partly inconsistent with what was observed in Vietnamese for which the intensity drop at 90% of the vowel duration is less important before a word-final stop than before a syllable-final within-word stop [19].

5. CONCLUSION

Knowing that burst helps listeners to discriminate the plosives of a system [25], and that native speakers are able to correctly identify a consonant even in the absence of its audible release [8] [17] [18], this study investigated acoustic cues of unreleased final stops in Thai in order to highlight perceptual-acoustic properties that allow listeners to categorize the unreleased allophones of stops. Based on a previous study [19], we also investigated effects of gender and of word type (simple monosyllables *vs* dissyllabic compounds) on acoustic cues. We analysed, as in [19], the duration of three target stops /p, t, k/ and of the preceding vowel /a/, as well as the temporal evolutions of I, f_0, and $F_{1\text{-}3}$ by considering place of articulation, word type and gender as independent variables.

Our findings confirm that differences in duration of unreleased stops and in other acoustic parameters of VC transitions (i.e. F_1, F_2, I, and in a lesser extent F_3) according to place of articulation are present in Thai. F_2 transition show three specific temporal evolutions depending on the articulation place (rising for coronal, flat for velar and downward for labial), thus partially confirming Delattre's work [24] in synthetic speech. Most of these results on Thai are consistent with those of [8] [19] for Vietnamese and provide information to suggest what acoustic parameter is useful for the categorization of unreleased stops and what is not. However, to be conclusive, they will need to be verified by perceptual tests. More broadly, the studies on relevant acoustic cues to the discrimination of unreleased stops contribute to validate the important role of acoustic transitions in speech perception (see e.g. [26] [27]).

Another result of this study concerns significant effects of word type on the production of syllable rhyme: the realization of a number of acoustic parameters (consonant duration, vowel duration, F_1, f_0) varies as a function of word-type, suggesting that it could probably be the result of different degrees of coarticulation between intra-word *vs* inter-word syllables in this language.

6. ACKNOWLEDGEMENTS

This study was conducted within the research framework of the SCALA project (*Stop Consonants in Asian LAnguages*) (reference ANR 15 IDEX 02).

REFERENCES

[1] N. Vallée, L. J. Boe, J. L. Schwartz, P. Badin, and C. Abry, "The Weight of Phonetic Substance in the Structure of Sound Inventories," *ZAS Papers in Linguistics*, vol. 28, pp. 145–168, Jan. 2002, doi: 10.21248/zaspil.28.2002.163.

[2] I. Maddieson, *Patterns of sounds*. in Cambridge studies in speech science and communication. Cambridge ; New York: Cambridge University Press, 1984.

[3] F. Argod-Dutard, *Éléments de phonétique appliquée: prononciation et orthographe en français moderne et dans l'histoire de la langue ; aspects prosodiques et métriques*, Réimpr. Paris: Armand Colin/Masson, 1996.

[4] Calliope, *La parole et son traitement automatique*. Paris Milan Barcelone Mexico: Masson, 1989.

[5] C. Meunier, Phonétique acoustique, in Les dysarthries, P. Auzou, Ed., Solal, 2007, pp. 164–173.

[6] P. Ladefoged and I. Maddieson, *The sounds of the world's languages*. Oxford: Blackwell, 1996.

[7] K. Tsukada, A Cross-linguistic Acoustic Comparison of Unreleased Word-final Stops: Korean and Thai. In *Proceedings of INTERSPEECH 2004 - ICSLP, 8th International Conference on Spoken Language Processing*, Jeju Island, Korea, 2004, pp. 1293–1296. doi: 10.21437/Interspeech.2004-76.

[8] T. T. H. Tran, *"Processus d'acquisition des clusters et autres séquences de consonnes en langue seconde : de l'analyse acoustico-perceptive des séquences consonantiques du vietnamien à l'analyse de la perception et production des clusters du français par des apprenants vietnamiens du FLE,"* Thèses, Université de Grenoble, 2011. Accessed: Jun. 03, 2021. [Online]. Available: https://tel.archives-ouvertes.fr/tel-01568326

[9] A. S. Abramson, Word-Final Stops in Thai. In *Tai Phonetics and Phonology*, J. G. Harris and R. B. Noss, Eds., Bangkok: Central Institute of English Language, 1972, pp. 1–7. [Online]. Available: http://purl.org/sealang/abramson1972word.pdf

[10] M. R. K. Tingsabadh and A. S. Abramson, Thai. *Journal of the International Phonetic Association*, vol. 23, no. 1, pp. 24–28, Jun. 1993, doi: 10.1017/S0025100300004746.

[11] J. Kingston, Lenition. In *Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*, Somerville, MA, 2008, pp. 1–31.

[12] T. Cho, S.-A. Jun, and P. Ladefoged. Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics*, vol. 30, no. 2, pp. 193–228, Apr. 2002, doi: 10.1006/jpho.2001.0153.

[13] T. T. Doan. *Ngữ âm tiếng Việt* (Vietnamese Phonetics) Hanoi National University Publishing House, pp. 99–148, 1999.

[14] I. Rousset, *"Structures syllabiques et lexicales des langues du monde. Données, typologies, tendances universelles et contraintes substantielles,"* Thèses, Université Stendhal - Grenoble III, 2004. Accessed: Mar. 17, 2021. [Online]. Available: https://tel.archives-ouvertes.fr/tel-00250154

[15] A. S. Abramson and K. Tingsabadh. Thai Final Stops: Cross-Language Perception. *PHO*, vol. 56, no. 3–4, pp. 111–122, 1999, doi: 10.1159/000028446.

[16] S.-C. Rhee. Onset-Coda Asymmetries in Mishmi and Other Southeast Asian Languages. In *Papers from the Seventh Annual Meeting of the Southeast Asian Linguistics Society*, Tempe, Arizona, 2003, pp. 137–152.

[17] R. Iwata, H. Hirose, S. Niimi, M. Sawashima, and S. Horiguchi, Syllable final stops in East Asian languages: Southern Chinese, Thai, and Korean. In *Proceedings of the 1990 International Conference on Spoken Language Processing*, Kobe, Japan, pp. 621–624, 1990.

[18] K. Tsukada, "Cross-language perception of final stops in Thai and English: A comparison of native and non-native listeners. In *Proceedings of the 10th Australian international conference on speech science and technology*, Sydney, Dec. 2004, pp. 563–568.

[19] T. T. H. Tran, N. Vallée, and L. Granjon. Effects of Word Position on the Acoustic Realization of Vietnamese Final Consonants. *Phonetica*, vol. 76, no. 1, pp. 1–30, 2019, doi: 10.1159/000485103.

[20] N. Yamlamai and T. Tran. Effet de la position de la syllabe sur la réalisation acoustique des consonnes finales du thaï. In *Actes des XXXIIe Journées d'Études sur la Parole*, Aix-en-Provence, France, Jun. 2018, pp. 151–159. doi: 10.21437/JEP.2018-18.

[21] P. Boersma and D. Weenink. Praat: doing phonetics by computer [Computer program]. Version 6.2.23. Oct. 08, 2022 [Online]. Available: http://www.praat.org/

[22] T. Hothorn, F. Bretz, and P. Westfall. Simultaneous Inference in General Parametric Models. *Biom. J.*, vol. 50, no. 3, pp. 346–363, Jun. 2008, doi: 10.1002/bimj.200810425.

[23] X. H. Cao, *Phonologie et linéarité: réflexions critiques sur les postulats de la phonologie contemporaine*. Paris, France: Selaf, 1985.

[24] P. C. Delattre, A. M. Liberman, and F. S. Cooper. Acoustic Loci and Transitional Cues for Consonants. *The Journal of the Acoustical Society of America*, vol. 27, no. 4, pp. 769–773, Jul. 1955, doi: 10.1121/1.1908024.

[25] W. Wang. Transition and Release as Perceptual Cues for Final Plosives. *Journal of speech and hearing research*, vol. 2, pp. 66–73, Apr. 1959, doi: 10.1044/jshr.0201.66.

[26] T. A. X. Tran, V. S. Nguyen, E. Castelli, and R. Carré. Production and perception of pseudo-V1CV2 outside the vowel triangle: speech illusion effects. In *Proceedings of INTERSPEECH 2013*, Lyon, France, Aug. 2013, pp. 407–411. doi: 10.21437/Interspeech.2013-124.

[27] R. Carré, P. Divenyi, and M. Mrayati. *Speech: A dynamic process*. De Gruyter, 2017. doi: 10.1515/9781501502019.