Labial-Coronal *vs.* Labial-Velar Consonant Sequences: An Articulatory Study

Manon Carrissimo-Bertola[1], Nathalie Vallée[1], Ioana Chitoran[2]

(1) Gipsa-Lab, (2) CLLILAC-ARP

manon.carrissimo-bertola@gipsa-lab.grenoble-inp.fr, nathalie.vallee@gipsa-lab.grenoble-inp.fr, ioana.chitoran@univ-paris-diderot.fr

Studies on universal trends in languages, both from different genetic and geographic origins, have shown favored combinations of non-adjacent consonants in words across various syllabic structures. Regarding place of articulation, sound sequences are more often organized with a labial consonant followed by a coronal one, rather than the opposite order. McNeilage & Davis (2000) called this trend the Labial Coronal (LC) Effect, and since then, it has been observed in twenty or so language lexicons (Vallée & al. 2009, G-ULSID database). This preferred organization is attested (i) between heterosyllabic onsets of two consecutive open syllables - CVCV, (ii) between tautosyllabic onset and coda - CVC (Table1), and (iii) in children productions at the first stage of words (McNeilage & Davis, 2000). Carrissimo-Bertola (2010) looked for other favored consonant place orders, in particular a potential labial-velar (LV) trend (Table2). A labial-velar effect was hypothesized because it shares with the labial-coronal order an anterior-posterior order of constrictions, and the use of two independent articulators. No labial-velar preference was found. We therefore propose to compare the articulatory strategies for both consonant sequences according to the places of articulation in order to understand why some of them are more widespread than others.

Table 1: Mean LC/CL ratios estimated in four studies (* means *no data*).

Studies	Number of languages	CVC	CV.CV	Dissyllabic words
MacNeilage & al. (1999)	10	*	*	2,23
Rousset (2004)	10 to 13	1,44	1,73	2,39
Vallée & al. (2009)	17	1,89	1,68	2,79
Carrissimo-Bertola (2010)	19	6,59	1,70	2,56

Table 2: Mean LV/VL ratios for the 19 G-ULSID languages, and number of languages by trend (* *means no data*).

	CVC	CVCV
Mean ratio LV/VL	1,48	1,55
LV Trend	9	4
VL Trend	6	7
Without tendency	*	7
*	4	1

In the framework of the *Frame, Then Content Theory* (McNeilage &Davis, 2000), the explanation given for the LC Effect is based on the simplest articulation first: A labial closure (produced with a single jaw rising) is less complex to produce than a coronal one (which needs jaw rising plus tongue tip rising). More recent studies claimed that production and perception constraints are involved in the LC Effect (Lancia & Fuchs, 2011; Nazzi & al., 2009; Rochet-Capellan & Schwartz, 2005; Sato & al., 2007; Tsuji & al., 2012) showing a stronger stability of LC sequences compared to CL. Analyzing the phasing of gestures from French speakers' productions, Rochet-Capellan & Schwartz (2005) provided evidence that a possible anticipation of the coronal gesture during the realization of the labial closure should be the basis of the LC Effect, whereas aerodynamic and acoustic constraints do not allow a labial closure during the production of a coronal consonant. However, a similar study carried out with German speakers showed no overlapping between lip and tongue tip gestures, the labial and coronal closures remaining sequential, in anti-phase even at fast speech rate (Lancia & Fuchs, 2011).

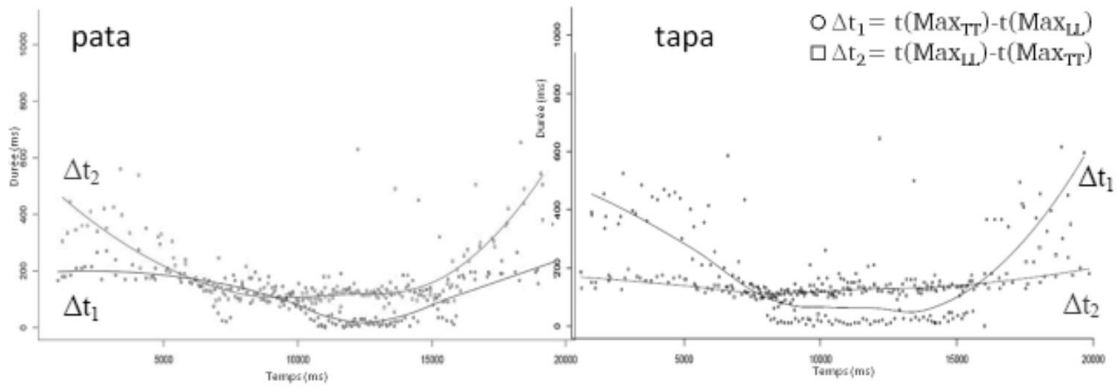
The study presented here deals with a similar experiment on French speakers. The articulatory stability of both LC vs. CL sequences, and LV *vs*. VL sequences were examined to test the following hypothesis: If the coronal gesture anticipation in a labial closure is the explanation for the LC Effect as claimed by

Rochet-Capellan & Schwartz, then LV sequences should also be favored over VL, as it is possible to anticipate the velar consonant during the labial closure.

Five native French speakers without hearing and speech impairment were asked to repeat during 18s a nonsense CVCV word with instructions to increase the speech rate then to decrease it by following rhythm of a visual metronome as much as possible. The interval between flashes increased from 600ms at the beginning to 100ms for the acceleration phase, then after a stability phase, the rhythm decreased in mirror-image to the acceleration phase. The four nonsense words were /paka/ *vs.* /kapa/, and /pata/ *vs.* /tapa/. Two distractors /taka/ and /kata/ were inserted on the list. The 6 dissyllables were randomly arranged inside a new list for each trial and three trials were presented to each speaker. Recordings took place in an anechoic room and articulatory data were collected with a Carstens® 2D-EMA at a sampling rate of 200 Hz. Lips, jaw, and tongue movements were captured by attaching 8 small sensors (2 of them provided reference data). The trajectory and velocity of each articulator were calculated, and the phasing of lips, jaw and tongue gestures were analyzed with Matlab® tools developed by Mark Tiede (Haskins Labs). We compared two duration measures (with Δt_i higher than 0): (i) Δt_1 and Δt_3 from the realization of /p/ (time where the lower lip achieves its highest position) to that of /t/ or /k/ (time where the tongue tip or tongue dorsum is at its highest position), (ii) Δt_2 and Δt_4 from the realization of /t/ or /k/ to that of /p/.

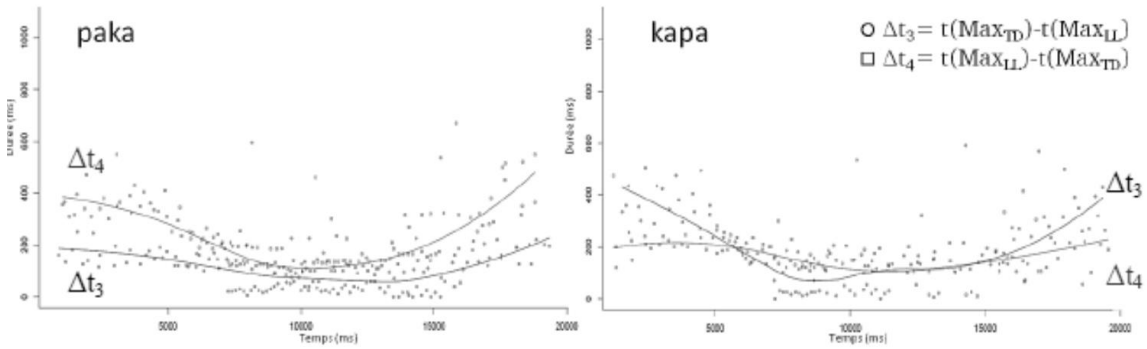
We observed that the accelerated speech paradigm involved a strong articulatory reorganization for both /pata/ and /tapa/ stimuli. At a slow rate, jaw movement trajectory showed two complete oscillation cycles: one synchronized with the tongue tip rising, the other one synchronized with the lower lip rising. At a fast speech rate, two behaviors appeared for both speakers: In some trials, the jaw seemed blocked and no oscillation was visible on the trajectory curve, while in other trials the two cycles of jaw movement evolved towards a single one. In the second case, labial and coronal gestures were no longer in anti-phase and moved together very closely, whereas they remained in anti-phase in the first case. Both for /pata/ and /tapa/, the time duration between the labial closure and the coronal closure was shorter than the time interval from /t/ to /p/ (figure1). For these two patterns, the speakers performed the task rather successfully: They highly increased the speech rate and maintained an alternation between /p/ and /t/.

Figure 1: Interval duration between /p/ and /t/ (left) and /t/ and /p/ (right) for the 3 trials in function of time (speaker 1).



For the pseudo-dissyllabic words with /p/ and /k/, increasing the speech rate while maintaining a correct alternation between /p/ and /k/ was more difficult for speakers. Despite this difficulty, they reorganized their articulatory gestures and a reduction of the time interval between the two consonant closures was observed with, however, a different attainment. In the case of /paka/, the /p/-/k/ succession remained stable in spite of the fast rate of speech, whereas for /kapa/, no pattern seemed the most stable (figure2). Therefore, we cannot conclude that a strong stability works for LV over VL consonant sequences as LC over CL, and thus cannot claim the presence of a LV Effect.

Figure 2: Interval duration between /p/ and /k/ (left) and /t/ and /p/ (right) for 3 trials in function of time (speaker 2).



This investigation nuances Rochet-Capellan & Schwartz‘s observations about the reorganization of jaw movement under the effect of speed which constrains the production of two CV syllables within a single jaw cycle. Our results show that the mandibular oscillation can be discarded during the acceleration phase of the repetition task given a minimal role of the jaw in an extreme condition of speech production. Moreover, if the two speakers took on a similar behavior for both /pata/ and /tapa/, both switching more frequently towards the LC pattern, such an articulatory strategy used at a faster speech rate seemed less clear for sequences with Labial and Velar consonants. This point might be related to the absence of a LV effect in lexicons of the world’s languages, also to the most frequent double plosives /k_p/ and /g_b/ (e.g. in African languages) which have a shorter duration than other consonant clusters (Ladefoged & Maddieson, 1995). Undoubtedly, timing of articulatory gestures needs to be further investigated and cross-checked with fine aerodynamic measurements.

References:

Carrissimo-Bertola, M. (2010). *Structures syllabiques des unités lexicales : « the fronting effect »* (Master1). Université Stendhal, Grenoble.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world’s languages*. Oxford: Blackwell.

Lancia, L., & Fuchs, S. (2011). The labial coronal effect revised. In Proceedings of the 9th International Seminar on Speech Production (ISSP) (p187-194). *9th International on Speech Production* (ISSP), Montréal.

MacNeilage, P. F., & Davis, B. (2000). On the origin of internal structure of word forms. *Sciences*, Vol. 288.

Nazzi, T., Bertoncini, J., & Bijeljac-Babic, R. (2009). A perceptual equivalent of the labial_coronal effect in the first year of life. *Journal of the Acoustical Society of America*, p. 1440-1446.

Rochet-Capellan, A., & Schwartz, J.-L. (2005). The Labial-Coronal effect and CVCV stability during reiterant speech production: An articulatory analysis. In *Proceedings of the9th Interspeech* 05 (p. 1013-1016). Lisbon.

Sato, M., Vallée, N., Schwartz, J.-L., & Rousset, I. (2007). A perceptual correlate of the labial-coronal effect. *Journal of Speech, Language, and Hearing Research*, 50, 1466-1480.

Tsuji, S., Gomez, N. G., Medina, V., Nazzi, T., & Mazuka, R. (2012). The labial–coronal effect revisited: Japanese adults say pata, but hear tapa. *Cognition*, 125(3), 413-428.

Vallée, N., Rossato, S., & Rousset, I. (2009). Favored syllabic patterns in the world’s languages and sensori-motor constraints. In F. Pellegrino, E. Marsicoa, I. Chitoran, & C. Coupé (Éd.), *Approaches to Phonological Complexity* (p. 111-140). Berlin: Mouton de Gruyter.

***

Syllable affiliation of CVC-sequences in two Portuguese varieties: an articulatory account and some perceptual data

Conceição Cunha

Institute of Phonetics and Speech Processing, LMU Munich

cunha@phonetik.uni-muenchen.de

The focus of this paper is on consonant coordination and syllable affiliation of a second consonant after vowel deletion in Brazilian (BP) and European Portuguese (EP) within the framework of Articulatory Phonology (Browman & Goldstein 1986, 1988, ff). More specifically, this paper compares the temporal and spatial organization of lexical CVC-sequences realized as CVC sequences in BP and as phonetic clusters in EP, which often arise following high vowel deletion in unstressed position. The results of the production study were tested in a perception experiment with an identification task, involving participants of the same Portuguese varieties.

European Portuguese high vowels [ɨ, u] in unstressed position are mostly deleted in connected speech (Mateus & Martins 1982; Martins et al. 1995). Consequently rise post lexically many clusters that are similar to lexical clusters resulting in near homophones (e.g. /k(ɨ)rer/, "to want" and /krer/, "to believe"). The distinction of such word pairs differing on the presence or absence of a pretonic vowel has been shown to be possible in English, since the consonants in the sequences with the deleted vowels were coordinated further apart than in lexical clusters and showed no c center effect (Browman & Goldstein 1992; Davidson 2006). The resulting issues to test are 1) whether there are fine phonetic timing and spatial differences between clusters and CVCs also in EP that may be scarcely audible and 2) whether timing differences can be perceived in perception, i.e. how good is the match between production and perception.

A further aim is to test whether there was a greater likelihood of vowel deletion and increasing overlap associated to the place of articulation of the first consonant. Following the place-order hypothesis (Chitoran et al. 2002, Kühnert et al. 2006) front-to-back cluster (e.g. /pr/) could be more overlapped than its back-to front counterparts (e.g. /kr/), because they are more easily recoverable in perception and in some cases are produced with independent articulators (e.g. lips and tongue tip for /pr/, following Kühnert et al. 2006). In the EP case, this would contradict some previous research that showed a smaller difference on perception between clusters and CVCs with velar C1 /k/ than with the bilabial one (*perece* vs. *prece* comparing with *crer* vs. *querer*, Mateus & Martins 1982). Consequently, we ran a physiological study and a perception experiment with the aim of testing whether both consonants of the lexical CVC-sequence show a comparable gesture overlap as it has been shown for lexical clusters in other languages (e.g., Byrd 1996; Marin & Pouplier 2010).

Physiological movement data were recorded using a 3D electromagnetic articulograph from five first language speakers of EP and four speakers of BP. The sensors were fixed mid sagittally on the lips (upper and lower lip), jaw and three on the tongue (tip, mid, back). The remaining were reference sensors. The speakers repeated every target word embedded in a carrier sentence 8 times. The stimuli consisted of lexical words containing velar or bilabial plosive in C1 position and liquid in C2 (i.e. /pr/, /pɨr/, /pur/, /kr/, /kɨr/, /kur/ and /pl/, /pɨl/, /pul/, /kl/, /kɨl/, /kul/ in initial position. The following vowel and the consonant were held constant for each stimulus set. For the measures of consonant overlap we defined a) the interval between the end of the constriction plateau of the first consonant (C1) and the beginning of the gesture movement of the second (C2) (Chitoran et al. 2002; Gafos 2010, Kühnert et al. 2006) and b) the time interval between the end of the constriction plateau of C1 and the begin of the constriction plateau of C2. For the c-center the interval between C2 and the consonant after the stressed vowel (C3) was taken. For the spatial analysis we defined the magnitude of the tongue tip between the highest and lowest point in the movement of the closing gesture.

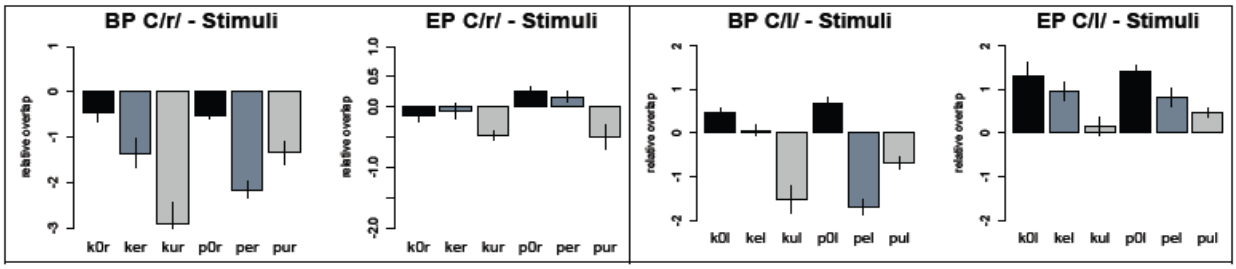
The distances and magnitudes were analyzed as independent variables with RM-ANOVA in R, first for all data and second separated for each variety, with lexical category, place of articulation and C2 as dependent variables and post-hoc t-test for each significant interaction between the factors. The results

showed high significant differences between both varieties and a clear main effect of lexical category for both C2 in BP. However, in EP the main effect was smaller for the lateral tokens and the overlap differences were not significant for the velar tokens with rhotics in the C2 position. Back-to-front CVCs (/ker, kel/) were overall more similar to lexical clusters than front-to-back PL clusters (/per, pel /).

37 native speakers of EP (11m/26f, originated from Porto and aged between 24-36 years) and 32 native speakers of BP (10m/22f, originated from Campinas and aged between 22-35 years) participated in the forced choice perception experiment testing the stimuli with rhotics. The results showed that listeners could not identify some of the CVC sequences when produced by EP speakers, resulting in rather a match between production and perception than between lexicon and production/perception.

Taking the results together, European CVC showed less overlap than consonant clusters, but the difference was found to be gradual and the lexical distinction was not always realized nor perceived even by native listeners. The proposed interpretation of the results is that the difference between clusters and CVC sequences was partially neutralized when produced by European speakers and this may result in the loss of the lexical contrast and an increasing number of homophones in EP.

Figure 1: Results of the consonants overlap measurement in the production of lexical clusters and CVC-sequences with the vowels /e, u/, realized as [e, u] in BP and [ɨ, u] in EP.



References

Cunha, C. Submitted. Portuguese lexical clusters and CVC sequences in speech perception and production.

Browman, C. P, and L. Goldstein. 1986. Towards an articulatory phonology. *Phonology yearbook* 3 (21): 9–252.

Browman, C. P and L. Goldstein. 1992: “Targetless” schwa: an articulatory analysis. In: G.J. Docherty and D.R. Ladd. Papers in *Laboratory Phonology II*: Gesture, Segment, Prosody. Cambridge: Cambridge University Press, 26-67.

Browman, C., and L. Goldstein. 2000. Competing constraints on intergestural coordination and selforganization of phonological structures. *Bulletin de la Communication Parlée* 5: 25–34.

Byrd, D. 1996. Influences on articulatory timing in consonant sequences. *Journal of Phonetics* 24: 209–244.

Chitoran, I., L. Goldstein, and D. Byrd. 2002. Gestural Overlap and Recoverability: Articulatory Evidence from Georgian. In *Laboratory Phonology*, 419–448. 7. Berlin, New York: Mouton de Gruyter.

Gafos, A. 2002. A grammar of gestural coordination. *Natural Language and Linguistic Theory* 20: 269–337.

Kühnert, B., P. Hoole, C. Mooshammer. 2006. Gestural overlap and C-center in selected French consonant clusters. *Proc. 7th International Seminar on Speech Production*, UFMG Belo Horizonte, pp. 327-334.

Marin, S., and M. Pouplier. 2010. Temporal organization of complex onsets and codas in American English: Testing the predictions of a gestural coupling model. *Motor Control* 14 (3): 380–407.

Mateus, M. H., and M. R. D. Martins. 1982. Contribuição para o Estudo das Vogais Átonas [əə] e [u] no Português Europeu. *Biblos* (58): 111–125.

Silva, D. J. 1997. The Variable Deletion of Unstressed Vowels in Faialense Portuguese. *Language Variation and Change* 9 (03): 295–308.

Implicit prosody in attribute attachment in Brazilian Portuguese

Marco Fonseca[1], Nivia Aniele Oliveira[2], Victor Hugo Medina Soares[2], Yuki Hirose[1]

[1]University of Tokyo, [2]Federal University of Minas Gerais

marcosilvafonseca@gmail.com, nivianiele@gmail.com, vhmedinasoares@gmail.com, hirose@boz.c.u-tokyo.ac.jp

This paper aims to present evidences for the Implicit Prosody Hypothesis (IPH) in Brazilian Portuguese (BP) by reporting on how subjects resolve syntactic ambiguity in NP1-V-NP2-attribute sentences, such as in (1):

(1) O repórter entrevistou o político sozinho. (*The reporter interviewed the politician alone.*)

In BP, two interpretations may be given to (1); that the politician was alone (high attachment) or that the reporter (low attachment) was alone. The ambiguity in (1) may be undone, in turn, by prosodic cues (Magalhães & Maia, 2006). This is because the prosodic parser tends to divide long sentences in order to maintain their balance (Fodor 1998; Lourenço-Gomes et al., 2005).

Although it is known that longer sentences tend to present high attachment (Lourenço-Gomes et al.,), the effect of the attribute length on attachment preference of NP1-V-NP2-attribute sentences has yet to be tested in BP. Bearing that in mind, we hypothesize that the longer the attribute of the sentence, the higher the probability of high attachment in BP. To test our hypothesis, we designed a questionnaire to verify whether 30 subjects judged that the nucleus of the NP1 or of the NP2 was related with the attribute. We presented 12 target sentences to the subjects in two conditions: long (3a) and short (3b):

(3a) O repórter entrevistou o político sozinho. (*The reporter interviewed the politician alone.*)

NP1 + V + NP2 + attribute

(3b) O repórter entrevistou o político claramente desolado e sozinho (*The reporter interviewed the politician clearly desolated and alone*).

NP1 + V + NP2 + intensity adverb + attribute 1 + attribute 2

Two questionnaires were created, and both conditions (long and short) were not present in the same questionnaire at the same time. In order to counterbalance the questionnaire, the same subject was not submitted to the same tests. After each sentence, one reading-verification question was asked to the participants, such as: “who was alone?” The answer to such questions showed the speaker’s attachment preference. We observed a general preference for low attachment (220/360=61.11%). However, the preference for low attachment was higher in short sentences (129/180=71.66%) when compared to long sentences (91/180=50.55%) ($x^2(1)=16.0013$, $p < 0.001$). These results indicate that length seems to play a role in attachment, confirming our hypothesis. We suggest this is due to the fact that the speakers tend to implicitly apply their prosodic patterns into silent reading.

In order to understand the nature of these prosodic patterns, we are conducting a production experiment using the same 12 sentences of the questionnaire. We recorded four subjects who did not take part in the first experiment. They were asked to read sentences such as in (1) with a contextualizing sentence prior to it, as in (4-7). The contextualizing sentence ensured that one of the two possible attachments was to be set. Overall, the subjects read four conditions of 12 sentences, in a total of 48 sentences for each subject.

(4) Short sentence, high atachment: Só havia um repórter no local. O repórter entrevistou o político sozinho. (*There was only one reporter there. The reporter interviewed the politician alone.)*

(5) Short sentence, low attachment: Só havia um político no local. O repórter entrevistou o político sozinho. (*There was only one politician there. The reporter interviewed the politician alone.)*

(6) Long sentence, high attachment: Só havia um repórter no local. O repórter entrevistou o político claramente desolado e sozinho. (*There was only one reporter there. The reporter interviewed the politician clearly desolated and alone*.)

(7) Short sentence, low attachment: Só havia um político no local. O repórter entrevistou o político claramente desolado e sozinho. (*There was only one politician there. The reporter interviewed the politician clearly desolated and alone.)*

Previous research (D'Imperio et al. 2005, Frota et al. 2007) has shown that intonational phrases boundary is usually marked with high boundary tones, pre boundary syllable duration and pauses. We hypothesize that the parser is more likely to place prosodic boundaries in long sentences, so we expect to find prosodic markers in our set of long sentences when compared to the set of short sentences. At the present time, we measured duration pauses and found 41 pauses (mean=142.35 milliseconds, s.d.=96.6), which are distributed as follows:

	Long sentence	Short sentence	Total
High attachment	21	2	22
Low attachment	18	0	18
Total	39	2	

Table: distribution of paused found in the production experiment

The table shows that the pauses were mostly found in long sentences (39), independently of attachment. This might indicate that the prosodic boundary tends to place a boundary in long sentences in contrast with short ones. As for the location of the pauses, they were mostly found after the NP2 (27), but they were also found after the NP1 (7), and within the attribute (7).

In summary, our results indicate that the length of the sentence is relevant for the interpretation of NP1-V-NP2-attribute sentences. We also found a high number of pauses in long sentences when compared to short ones, which seem to indicate intonational boundaries. This might provide evidence for the IPH. In further research, we intend to measure syllable duration and to conduct a boundary tone analysis.

References

D'Imperio, M. G., Elordieta. S. Frota, S. O, Prieto. M,. Vigário, M. 2005. Intonational phrasing in Romance: the role of syntactic and prosodic structure. In: *Prosodies*. Berlin: Mounton de Greyter, 59-97.

Fodor, J. D. 1998. Learning to parse? *Journal of Psycholinguistic Research*, n. 27, 285-319.

Frota, S. Vigário, M. 2007 Intonational phrasing in two varieties of European Portuguese. In: Riad, T. Gussenhoven, C. (eds.) *Tones and Tunes*, v. 1. Berlin: Mouton de Gruyter, 265-291

Loureço-Gomes, M. C.; Maia, M.; Moraes J. 2005. Prosódia na leitura silenciosa: Um estudo com orações relativas estruturalmente ambíguas. (Prosody in silent reading: a study with structurally ambiguous relative sentences). In: Maia M.; Finger, I. (eds.) *Processamento da Linguagem* (Language Processing). Pelotas: Educat, 131-161.

Magalhães, J. O. Maia, M. (2006). Pistas prosódicas implícitas na resolução de ambigüidades sintáticas: um caso de adjunção de atributos (Implicit prosodic clues in the resolution of syntatic ambiguities: a case of attribute adjunctions). *Revista da ABRALIN*, v. 5, n. 1-2, 143-167.

***

Tongue contour for /s/ and /ʃ/ in ultrasound tongue imaging

Danira Tavares Francisco; Luciana de Oliveira Pagan Neves; Haydée Fiszbein Wertzner

Department of Physiotherapy, Communication Science & Disorders,
Occupational Therapy of the School of Medicine from the University of São Paulo

danira.francisco@usp.br; lupagan@usp.br; hfwertzn@usp.br

The present study provides qualitative evidence of the differences between the production of the sounds [s] and [ʃ] based on the ultrasound tongue imaging. The study reveals that the sound [s] has greater variability of tongue positioning (we observed four different tongue contours) than the sound [ʃ].

Previous research has shown that ultrasound imaging of the tongue shape can be used with various sounds, to address phonological questions, conduct phonetic fieldwork, and aid speech rehabilitation (Gick, 2002; Bernhardt et al., 2003; Bressmannet al., 2005; Stone, 2005; Davidson, 2006).

Ultrasound enables scanning of the tongue shape during the production of different sounds like vowels (Chen & Li, 2011); and consonantal sounds like /s/ (Stone et al., 2012), /ʃ/ (Zharkova, Hewlatt & Hardcastle, 2011). The study of the ultrasound tongue imaging (UTI) in adults may provide important information about speech sound production variations, enabling comparisons between adult production, typically developing children and children with speech sound disorder.

The aim is to o describe the tongue contour for the production of the sounds /s/ and /ʃ/ in adult speakers of Brazilian Portuguese.

The participants were 21 Brazilian Portuguese native speakers, aged 18:0 to 29:11 years old (four male and 17 female). Inclusion criteria were age, to be born in the state of São Paulo, have never undergone any type of speech/language therapy and do not present any distortions or phonological processing during the imitation of words task. Subjects were submitted to ultrasound analysis after the speech evaluation. The equipment used was the Mindray DP 6600 ultrasound machine connected to a computer by a video and audio synchronizer. The ultrasound was applied in an acoustic room and the endocavitary low frequency (6 kHz) transducer probe was positioned in the submandibular area to capture images of the tongue contour in a midsagital view. The transducer was fixed to the subject's head using an individually adjusted helmet. The acoustic signal was recorded directly into the computer using a Sennheiser microphone (Vocal Evolution model 817). The software Articulate Assistant Advanced (AAA) was used to capture and analyze the productions. The words /'sapu/ (frog) and /'ʃavi/ (key) were chosen because they are both accompanied by the tonic and neutral vowel /a/. The first step of the recording process was to ask the subjects to swallow a sip of water to trace the palate contour. The examiner showed in a computer screen the pictures corresponding to the target words five times each, and the subject was instructed to say the name of each picture. The pictures were presented during three seconds with an interval of 200ms between each one. The midpoint of the consonant was selected to avoid both the movement of tongue preparation to produce the target sound and the transition movement (coarticulation) between the target sound and the following vowel. The tongue contour was individually traced using the software AAA for the /s/ and /ʃ/ sounds for each repetition. The tongue contours were traced individually and after that, images were superimposed. The description of the tongue contour was based on the curvature and positioning of the tongue tip in relation to the tongue root.

The qualitative analysis of tongue contour for [s] showed that even though auditory perception was of the target [s], this sound was produced in four different ways. General observation demonstrated that the sound [s] was produced with a slight elevation of the tongue dorsum and great approach of the tongue tip to the alveolar ridge. The production of [ʃ] was not as variable as the production of [s]. The sound [ʃ] seemed to be produced by all the participants with the tongue dorsum closer to the palate and the tongue tip moving downwards. Comparison from the tongue contours traced for the two sounds indicated that

they are different and variable. In general, tongue contour for the production of [s] demonstrated that the tongue position was flatter with the tongue tip positioned closer to the alveolar ridge while the tongue contour for the production of [ʃ] indicated a more pronounced curvature of the tongue with an approximation of the tongue dorsum to the palate and a lower tongue tip. The superimposed images of the tongue contours for the production of [s] and [ʃ] for each subject pointed to a differentiation of the tongue contour. Such differentiation was evident for 47.62% of the subjects who demonstrated different positioning of the tongue tip, tongue dorsum and tongue root between the two target sounds. For the other subjects (52.38%) this differentiation was less evident with a different tongue positioning only in the anterior part of the tongue body and the tongue tip.

References

Bernhardt BM, Gick B, Bacsfalvi P, Ashdown J. Speech habilitation of hard of hearing adolescents using electropalatography and ultrasound as evaluated by trained listeners. *Clinical Linguistics and Phonetics*. 2003;17(3):199-216.

Bressmann T, Thind P, Uy C, Bollig C, Gilbert RW, Irish JC. Quantitative three-dimensional ultrasound analysis of tongue protrusion, grooving and symmetry: Data from 12 normal speakers and a partial glossectomee. *Clinical Linguistics and Phonetics*. 2005;19(6/7):573-588.

Chen Y, Lin H. Analysing tongue shape and movement in vowel production using SS Anova in ultrasound imaging. *ICPhS XVII*. Hong Kong, 17-21 August 2011.

Davidson L. Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance. *Journal of the Acoustical Society of America*. 2006;120(1):407-415.

Gick, B. The use of ultrasound for linguistic phonetic fieldwork. *Journal of the International Phonetic Association.* 2002;32:113-122.

Stone M. A Guide to Analyzing Tongue Motion from Ultrasound Images.Clinical *Linguistics and Phonetics.* 2005;19(6-7):455-502.

Stone M, Rizk S, Woo J, Murano EZ, Chen H, Prince JL. Frequency of apical and laminal /s/ in Normal and Post-glossectomy Patients. In:

http://speech.umaryland.edu/Publications/StoneJMed_SLP_paper_2012.pdf

Zharkova N, Hewlett N, Hardcastle WJ. Coarticulation as an Indicator of Speech Motor Control Development in Children: An Ultrasound Study. *Motor Control*. 2011;15:118-140.

***

The perception of phonetic detail: a preliminary study about tap

Gustavo Nishida[1], Jeniffer Albuquerque[1], Maria Gabriel[2]

[1]Universidade Tecnológica Federal do Paraná, [2]Universidade Federal do Paraná.

nishida.utfpr@gmail.com, jeniffer.albuquerque@gmail.com, mgabriel.ufpr@gmail.com

Some of the landmarks in the field of speech perception are the collaborative studies of Liberman and his colleagues from the Haskins Laboratories throughout the 1950's (Liberman, 1957). These studies have based speech perception studies for three decades. Despite the suggestion of an articulatory cue for speech perception, it was only in the 70's that visual features of people's faces gained importance. Back then, the impact of the McGurk effect (McDonald & McGurk, 1976) made it possible to analyze speech perception through another perspective: what would be the importance of the articulatory cues, which are obtained through sight, to speech perception?

In addition, due to the improvement in the analysis' methods, the phonetic studies have embraced some changes in relation to what patterns would be considered language-specific or universal (Keating, 1985). Studies such as Keating's, point out the occurrence of phonetic detail which are language-specific and, which, therefore, should be taken into account in phonic grammars. An example of this finding is Nishida (2009). The author describes a vowel-like element next to Brazilian Portuguese's (BP) tap. This work suggested an intervocalic nature for the tap segment, independently from its position in the syllable. Nishida assumes in this work that the vowel-like element would be responsible for the tap's perception. Taking this into account, a possible way to test this hypothesis would be to conduct an experiment in which the participants would listen to an edited version of the tap segment, i.e., a tap without the vowel-like element.

Our experiment's hypothesis suggests that the tap in onset position is not perceived when it does not present the vowel-like element. However, we decided it would be highly important to insert visual information in this experiment, once the participants may profit from the "ambiguous" stimuli in order to decide over the presented data. This being said, if this hypothesis is confirmed, it would be an argument in favor of the realist approach, of Fowler's direct speech perception theory (Fowler, 1996), due to the multimodal nature of the proposal.

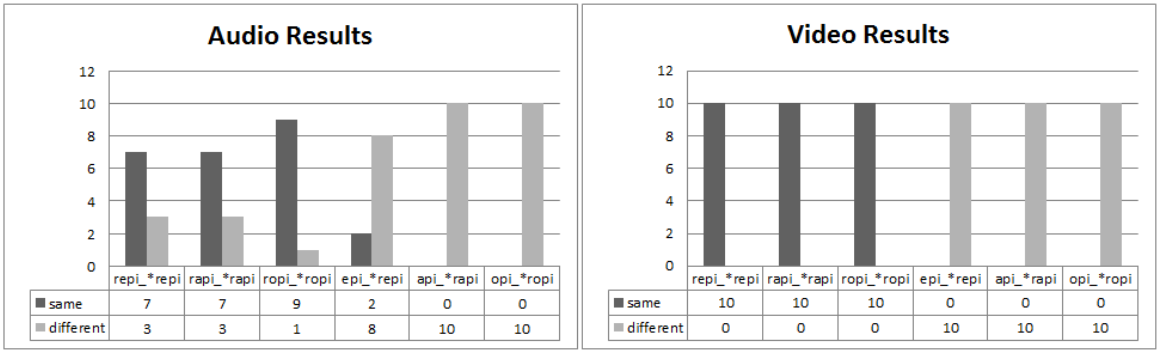
In order to test the hypothesis, a perceptual discrimination experiment (Categorial AX- same /different) was conducted. All of the 10 participants were students from the Language and Literaturecourse of UFPR (Federal University of Paraná): 4 men and 6 women. The corpus was composed of three groups of nonsense words: a) target words, which contained words with the tap sound, with the edited tap sound (*) and without the tap sound – *rapi*, *repi*, *ropi*, **rapi*, **repi*, **ropi*, *api*, *epi*, *opi* ; b) controlling words: a group formed by the same words of the first group, which were paired up in order to be equal (for example, the participants would listen or listen and watch a nonsense word like *repi* and the same data in the sequence); c) distractor words: were formed of voiced and voiceless plosives: *papi*, *tapi*, *kapi*, *bapi*, *dapi*, *gapi*, *pepi*, *tepi*, *kepi*, *bepi*, *depi*, *gepi*, *popi*, *topi*, *kopi*, *bopi*, *dopi* and *gopi*.

The data was recorded using the software *Audacity* 1.3.5. The audio data was segmented in *Praat* version 5.374 (Boersma & Weenink, 2014) and later submitted to normalization in *Audacity* 1.3.5, with a -5dB intensity. The videos were recorded with a Sony Cyber-shot DSC H200 Digital Camera (resolution 1280x 720 HD) and edited in *Adobe Premiere Pro CS6* (2001). The test was performed using the software *TP* (Rauber, Rato, Kluge, Santos, 2012). The participants listened and listened/ saw 30 sequences, 60 sequences total. Since the video data was too heavy, the 30 videos were presented in two sequences of 15.

As it was mentioned before, two types of information were used; therefore, we presented the audio and video data in a way to check whether different information sources (audio and visual) would interfere in the participants' discrimination task. So, 5 of the participants listened to the audio first and were presented to the video afterwards and the other 5 did the opposite. In addition, it is important to state that some of the data have passed through a dubbing process, i. e., a part of data was edited, by presenting

the image, the way it was recorded, and an edited audio. The preliminary results can be seen in the two graphics below.

Figure 1: Audio and Video Results



Although both graphics, the Audio Results and the Video Results images, display similar results, it is possible to notice that there were no divergences when the participants had the video as an input, independently of the order in which they were presented (whether the video or audio were presented first). This resultshows that the visual cue is helpful when discriminating information. Also, the participants' results towards the video may reveal that they take "ambiguity" into account when deciding if the data were the same or not, and this may lead us to assume the multimodal nature of speech. In addition, the reaction time was measured for the video and audio responses. Although this study has not conducted any statics tests yet, the first tests point out that the participants were able to discriminate the sequences faster with the audio than with the video. Yet, the video results were more homogeneous. As for the group of controlling words, the majority of results were consistent, i.e., the majority of the participants were able to spot the differences and similarities among the data.

Although this research has provided us with a better understanding of what is involved in the perception of the tap sound, it is still a preliminary study. Up to the moment, the results lead us to confirm the initial hypothesis that the participants did note recognized the tap without the vowel-like element. Yet, more data is being collected and another perception test has been conducted: an identification one.

References

Nishida, G. 2009. A natureza intervocálica do tap. In: *Cadernos de Pesquisas em Linguística* (PUCRS), v. 4, p. 67-79.

Boersma, P. Weenik, D. "Praat: Doing phonetics by computer" (Versão 5.3.74), 2014. Disponível em: http://www.praat.org/

Fowler, C. Listeners do hear sounds, not tongues. 1996. In: *The Journal of the Acoustical Society of America*, 99 (3), pp. 1730-1741.

Liberman, A. 1957. Some results of research on speech perception. In: Liberman, A. *Speech: a special code*. Cambridge, MA: MIT Press, 1996.

McGurk, H., MacDonald, J. 1976. Hearing Lips and Seeing Voices. In: *Nature*, p. 264746-48.

Keating, P. 1985. Universal phonetics and the organization of grammars. In: FROMKIN, V. (Ed.). *Phonetic Linguistics: essays in honor of Peter Ladefoged*. New York: Academic Press, p. 115 132.

Rauber, A., Rato, A., Kluge, D., Santos, G. 2012. TP, v. 3.1 [Application software]

http://www.worken.com.br/tp_regfree.php/

***

The Relationship between Linguistic Stress and Musical Accent in Bossa Nova

Gabriela Ricci & Eleonora Albano

State University of Campinas

gabricci@hotmail.com, albano@unicamp.br

The aim of this study is to investigate the relationship between linguistic and musical prosody in *Bossa Nova*. This is a pilot study of accent accumulation of accents in musical scores and its relation to lexical stress in the lyrics of two emblematic songs of this genre. We also analyze the changes made by the interpreter in the score accent pattern and the contribution of the instrumental accompaniment to enhance accent accumulation in interpretation.

Bossa Nova is a musical genre built from variations of syncope, a rhythmic figure that moves the downbeat accents to the upbeat and extends them to the next downbeat. This genre is claimed to be a kind of "spoken song" because of its impressionistic closeness to speech in Brazilian Portuguese (henceforth BP) (Garcia, 1999). This study consists of analyses of *Chega de Saudade* (henceforth CdS) and *Bim Bom* (henceforth BB), two key songs of *Bossa Nova*. Both are sung by João Gilberto who accompanies himself with an acoustic guitar. João Gilberto is regarded as the greatest performer as well as one of the greatest *Bossa Nova* composers.

The hypothesis is that the incidence of musical accent should be higher in stressed syllables than in unstressed ones, approximating the interpretation of the lyrics to BP speech. This is, in turn, reinforced by the accompaniment. Musical scores with the lyrics of the two songs were coded by the first author, who is a trained professional musician, on the basis of Gilberto's recordings. At a later point of the study, consistency will be checked and measured through the addition of another two coders.

The syllables of the lyrics were classified as stressed and unstressed. The notes of the melody aligned with such syllables were defined as having tone, duration, downbeat and upbeat accents. The downbeat (henceforth, DB) is characterized by the first note of each musical measure, and can occur at a regular pace (henceforth rDB) – i.e., concomitant with the beginning of the measure – or at a syncopated pace (henceforth sDB) – i.e., displaced with respect to the beginning of the measure. The upbeats (henceforth, UB) are the beats that do not occupy the first place of a musical measure, and can occur at a regular pace (henceforth, rUB) – i.e., concomitant with the beginning of the beat – or at a syncopated pace (henceforth, sUB) – i.e., displaced with respect to the beginning of the beat.

Only rhythm was considered in the accompaniment because we believe that it is largely responsible for the characterization of *Bossa Nova*. Another important aspect of the accompaniment is that it only occurs in interpretation. In it, we observed whether rDB, sDB, rUB and sUB were aligned with the syllables of the lyrics and the notes of the melody.

The number of musical accents present in each melody note was counted in the score and in the interpretation. This count allowed us to observe the convergences and divergences between the former and the latter. Analysis of the accompaniment was added to that of the interpretation, and alignment with the melody and lyrics was observed and classified as rDB, sDB, UB or not aligned. The sDB alignment type suggests an interpretation most characteristic of *Bossa Nova*, due to its syncopated rhythm.

Figure 1: Proportions of musical accents in the syllables on CdS for score and interpretation

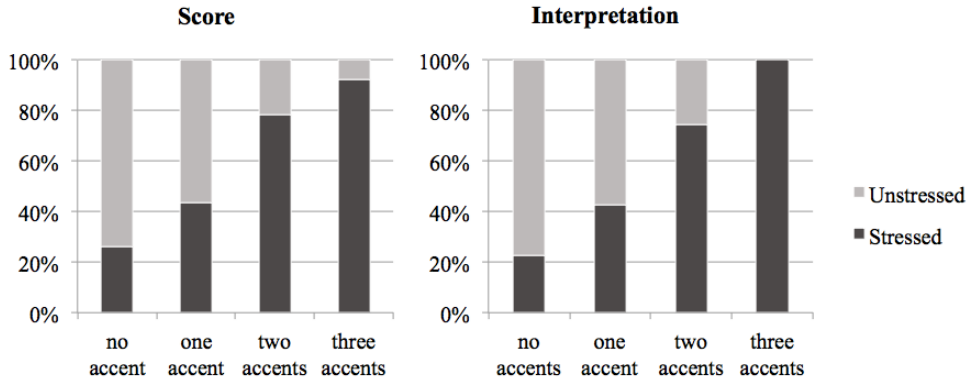
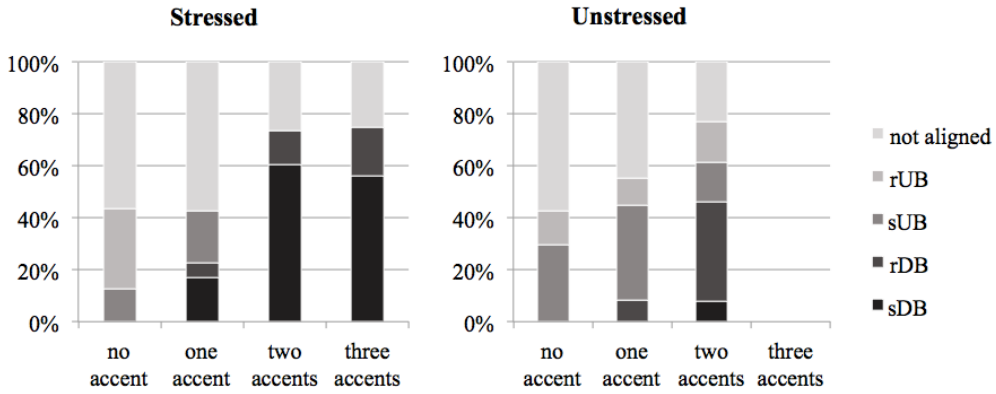


Figure 1 shows that both in the score and in the interpretation: (i) the occurrences of stressed syllables with musical accent are directly proportional to the number of accumulated musical accents; and (ii) that the unstressed syllables show the reverse pattern, i.e., their occurrences are inversely proportional to the number of accumulated musical accents.

Figure 2: Alignment of accompaniment with stressed and unstressed syllables on CdS for interpretation



Alignment of accompaniment with lyrics and melody reinforces the accumulation of musical accents, as can be seen in Figure 2. There is a higher incidence of alignment of the accompaniment with lyrics and melody in stressed syllables with greater accumulation of musical accents. Unstressed syllables contain a smaller number of musical accents and, in most cases, do not align with DB. sDB appears more frequently in stressed syllables with two and three musical accents. Data from BB maintain the alignment and musical accent pattern observed in CdS.

We can, therefore, conclude that the score, reinforced by the interpretation and the accompaniment, points to a dynamic that creates close interaction between linguistic and musical prosody, a characteristic feature of *Bossa Nova*.

References

Garcia, W. 1999. *Bim bom: a contradição sem conflitos de João Gilberto*. São Paulo: Paz e Terra.

***

Lexical Stress in Singing: a pilot study

Cássio Andrade Santos

University of São Paulo (USP)

cassioandradesantos@hotmail.com

Lexical stress in Brazilian Portuguese is characterized by longer duration of the stressed syllable nucleus (Massini-Cagliari, 1992). However, in a song, the segments' length – musical notes and/or syllables - is predetermined by the musical rhythmic structure. Thus, we have asked whether syllable duration in a song follows the western musical structure (half notes, quarter notes and eighth notes) or whether it follows the linguistic structure (stressed and unstressed syllables). It has been assumed, as an initial hypothesis, that the relation between vowel quality and vowel duration (Massini-Cagliari, op. cit.) is narrower in singing than in speech. We have assumed that the duration by itself will not suffice to characterize the stressed syllable compared to unstressed one in singing, and thus we have proposed to analyze the formant pattern in sung vowels.

One of the theoretical basis for this study is Articulatory Phonology (AP) (Browman; Goldstein, 1986; Albano, 2001), which defines speech as a dynamic phenomenon: occurring in time and space (Fowler, 1980; Elman, 1995; Kelso, 1995). AP establishes a relationship – by the *articulatory gesture* as a linguistic unit - between the physics of speech and its symbolic nature. Thus, we have also asked, based on the problem of pattern complexity (Kelso, op cit.), if the articulatory gesture can be extended to the universe of song.

In our present research (doctorate research in progress) we are developing an experiment to verify how syllable length behaves in Brazilian popular music. This experiment compares speech and singing, using recurrent stress patterns found in Brazilianpopular music: stress syllables in down beat, pre-stressed syllable in upbeat and pos-stressed syllable in upbeat.

For this pilot study we recorded a competent male singer (the whole experiment will record 10 male subjects), *a capella*, without the use of a metronome or a tuning fork. Recordings were done in a sound attenuated booth with Shure head-mounted microphones. The subject sang and spoke the first verse of the song *A Banda* (The marching band), composed by Chico Buarque de Holanda, in which we inserted the following non-sense words:

/ˈpa.tɐ/, /ˈpe.tɐ/, /ˈpɛ.tɐ/, /ˈpi.tɐ/, /ˈpɔ.tɐ/, /ˈ'pu.tɐ/;

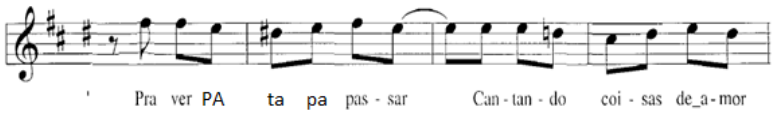
/ˈpa.tɐ/, /ˈpa.tɪ/, /ˈpa.tʊ/

and /pɑ.ˈta.pɐ/, /pe. ˈta.pɐ/, /pi. ˈta.pɐ/, /po. ˈta.pɐ/, /pu. ˈta.pɐ/,

In the figure below, there is an example of /pɑ.ˈta.pɐ/, as adapted to the lyrics without altering rhythm according to the original score.

***Figure 1**. Example of nonsense word inserted in the experiment song /pɑ.ˈta.pɐ/*

('to see /pɑ.ˈta.pɐ/ passing, singing love things')



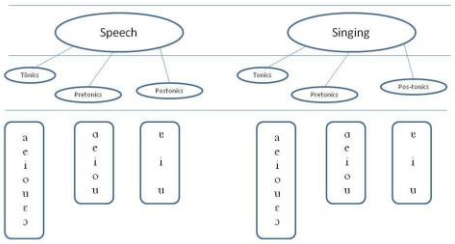
Each sentence was spoken and sung three times and saved in separate files. For this pilot study, which will guide us for the rest of the research in terms of data collection, organization and analyses, we chose only extreme vowels to measure. Duration and formant pattern measures (first 3 formants) will be done for vowels in the sung words, as follow:

/ˈpa.tɐ/, /ˈpi.tɐ/, /ˈpu.tɐ/, /ˈpa.tɐ/, /ˈpa.tɪ/, /ˈpa.tʊ/, /pɑ.ˈta.pɐ/, /pi.ˈta.pɐ/, /pu.ˈta.pɐ/,

Although we expect to find similar durations between sung vowels of different stress positions, we expect to find different formant patterns between vowels being compared (the stressed x pos-stressed). Values of duration and formant pattern will be submitted to descriptive statistical analysis, and then, after preliminary analyses a hypothesis test will be elected to compare the selected vowels.

Our aim with this preliminary study is to discuss the plausibility of our broad hypothesis (could the phonological gesture be accounted in the music domain?) and to check the feasibility of our experimental design that is resumed in figure 2.

Figure 2. Experimental design for the complete data collection and analyses.



The main result, for now, focuses on the mean duration of syllables. In the table below one can see that stressed syllables are longer than the no-stressed ones.

Table 1: Mean duration of stressed and non-stressed syllables**.**
Percentage refers to the measured syllable length in relation to sentence length

	Singing			Speech		
	SyllableDur (ms)	SentenceDur (ms)	%	SyllableDur (ms)	SentenceDur (ms)	%
Stressed	238	1694	14	208	1334	16
PreStressed	191	1805	11	156	1424	11
PosStressed	202	1811	11	130	1368	10

Sentence length was chosen as a parameter, because we do not control, at the time of the recordings, the speech rate / musical tempo, which we believe influences the raw data. Our first observation, without statistical analysis, is that stressed syllables may be longer than the unstressed ones in singing, as percentages indicates. Surprisingly, syllable length in singing is very close to syllable length in speech. We are facing unexpected similarity between singing and speech, maybe due to the segment chosen to be measured, and for this reason, the next step will be measuring isolated vowels. For now, we have to highlight that the difference between differently stressed syllables in singing appears to indicate an influence of speech on singing in terms of temporal structure as all sung syllables in the analyzed utterance should ideally possess the same duration because of their alignment with the notes.

References

Albano, E. C. (2001) *O gesto e suas bordas:* esboço de Fonologia Acústico-Articulatória do Português Brasileiro. Campinas: Mercado de Letras/ALB/FAPESP.

Browman, C.; Goldstein, L. (1986) Towards an Articulatory Phonology, *Phonology Yearbook*, 3: 219-252,.

Fowler, C. A. (1980) Coarticulation and theories of extrinsic time coarticulation and theories of extrinsic time. Journal of Phonetics 8., p. 113 - 133

Elman, J. L. (1995) Language as Dynamical System. In: Port, R. F., Gelder, T. van. p. 195 - 225

Kelso. J A. S. (1995) Dynamic patterns: the self-organization of brain and behavior. The MIT Press. London

Massini-Cagliari, G. (1992) Acento e ritmo. São Paulo : Contexto, (Coleção repensando a língua portuguesa).

***

An ultrasound study of Mandarin fricatives and affricate consonants

Xiu Ming & Didier Demolin

Gipsa-Lab, Université de Grenoble, Alpes

myengsu@hotmail.com

The most systematic study of retroflex consonants in Mandarin is Ladefoged & Wu (1984) who studied the Pekingese dialect. In their study, these consonants are described as 'so-called retroflex'. Their results show that there is no retraction and elevation of the tongue to the back. Ladefoged & Wu (1984) recorded static X-rays, palatograms and acoustic data. Ladefoged and Maddieson (1996) mention that sounds like '*...[ʂ] does not involve the tip of the tongue being curled up and backwards into the palatal region as in the Dravidian sub-apical retroflex stops, nor does it have the apical post-alveolar shape that occurs in the Hindi retroflex stops'*. For Ladefoged and Maddieson this sound is produced with the upper surface of the tip of the tongue, making it a laminal rather than an apical post-alveolar. They also suggest that the strategies of Mandarin speakers to produce sounds like [tʂ] are a bit more subtle. Indeed it seems that in this case the tongue is raised in the coronal region but without involving a backward movement into the palatal region.

This presentation describes the articulatory contrasts between alveolar, post-alveolar (retroflex), palatalized post-alveolar (alveo-palatal) fricative and affricate consonants in Mandarin using the utrasound technique. The aim is to compare tongue contours between the three sets of consonants (alveolar, post-alveolar, palatal and alveo-palatal) and to describe how they are articulated. The present state of the study is based on recordings made with 6 Mandarin speakers using the following set of data (numbers indicate tones). Mandarin speakers came from 3 different regions of China (Sichuan, Beijing and Mandchourie)

[tʂ] vs [ts] :	[tʂang4] (账 account)	vs [tsang4](藏 Tibet)
[tʂ] vs [ts] :	[tʂang1] (张 family name)	vs [tsang1](脏 dirty)
[tʂʰ] vs [tsʰ] :	[tʂʰeng2] (成 to become)	vs [tsʰeng2](层 floor)
[ʂ] vs [s] :	[ʂi1] (师 teacher)	vs [si1](思 to think)

Data were recorded with a portable ultrasound Terason T3000 system, one microconvex 140° transductor and a standard acquisition sound system. Data were recorded with Ultraspeech1 (http://www.ultraspeech.com/download/ultraspeech/ultraspeech_1_2_setup.msi) which processes ultrasound and audio data in parallel. The software was used to record simultaneously and in a synchronous way: the acoustic signal (16 bits, 16 kHz) and the ultrasound flux (320X240 pixels). The recording was made at 80 frames per/s. Three repetitions of each word was recorded and tongue contours were extracted in the middle of the fricative and retroflex and in the stop and fricatives parts for the affricates. The location was chosen from the corresponding waveform (Figure 3 and 4).

Results of our study confirm previous observations made by Ladefoged and Wu (1984). They also corroborate previous studies showing that the sub-lingual part of the tongue is not involved in the articulation of the so-called retroflex. Our study shows that the articulation of Mandarin fricatives and retroflex consonants varies in each region. Subjects from the North produce fricatives with a slight elevation of the tongue tip and some elevation of the back of the tongue. Retroflex consonants are articulated with an elevation of the tongue blade and apex (which are not differentiated in this case). The back of the tongue is also fronted. The constriction is in the tongue blade region. This is illustrated for the alveolar and retroflex fricatives in figure 1.

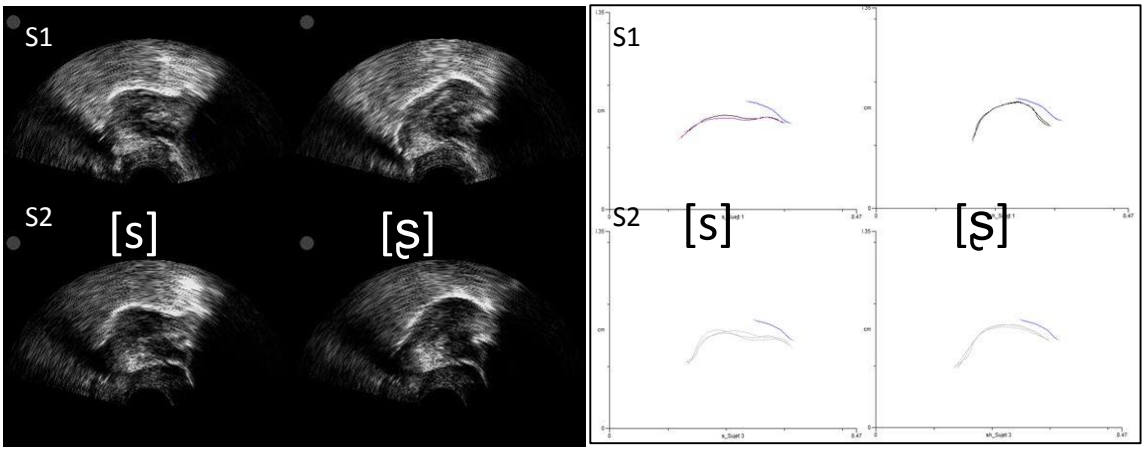


Figure 1. Ultrasound and extracted tongue contours (3 repetitions) from two subjects from the North (Mandchourie) for the alveolar and retroflex fricatives [s]-[ʂ] in the words [si1] 'to think' and [ʂi1] 'teacher'.

Subjects from the South (Sishuan) show that the tongue dorsum is flat with a slight elevation of the back of the tongue for the fricatives. Retroflex consonants are articulated with an elevation of the tongue blade and apex (which are not differentiated in this case). The back of the tongue is also fronted. The constriction is still in the tongue blade region but more in the front. This is illustrated in Figure 2. Similar observations can be made for the affricates (plain and aspirated). When compared to the palatalized post-alveolar fricatives and affricates [tɕ], [tɕʰ], [ɕ] it can be seen that there is much less variability for those consonants. This suggests that fricatives and affricates are more variable and that this might have a regional basis.

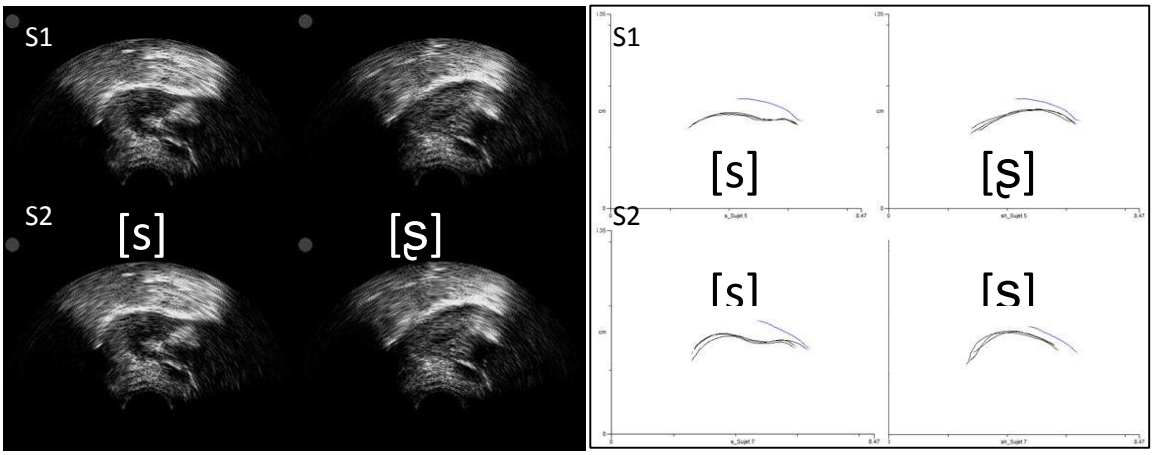
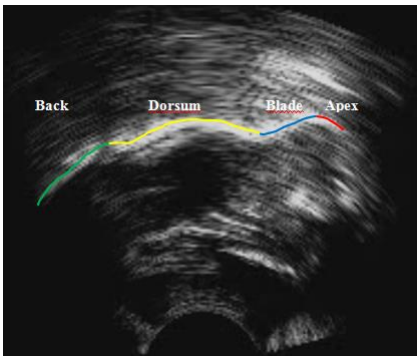


Figure 2. Ultrasound and extracted tongue contours (3 repetitions) from two subjects from the South (Sishuan) for the alveolar and retroflex fricatives [s]-[ʂ] in the words [si1] 'to think' and [ʂi1] 'teacher'.



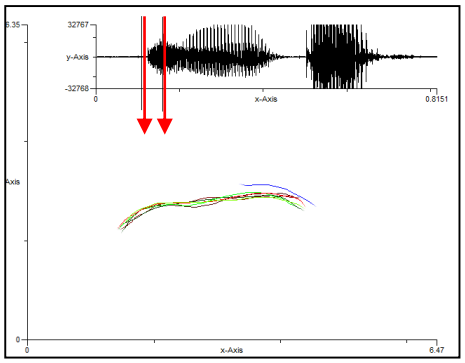


Figure 3 (left) shows the identification of the different parts of the tongue. Figure 4 shows (arrows) the place where the tongue contours were extracted from the corresponding waveform

References

Ladefoged, P.; Maddieson, I. (1996) *The sounds of the world's languages*. Oxford: Blackwell.
Ladefoged, P.; Wu, Z. (1984) Places of articulation: an investigation of pekingese fricatives and affricates. *Journal of Phonetics*, 12, 267–278.

***