

## Speech Production in Two Occlusal Classes

**Luis M. T. Jesus**

University of Aveiro  
Portugal

**André Araújo**

Instituto Politécnico do Porto /  
University of Aveiro  
Portugal

**Isabel M. Costa**

University of Aveiro  
Portugal

ONOMÁZEIN 29 (junio de 2014): 129-151  
DOI: 10.7764/onomazein.29.12



**Luis M. T. Jesus:** School of Health Sciences (ESSUA), University of Aveiro / Institute of Electronics and Informatics Engineering of Aveiro (IEETA), University of Aveiro. Portugal. Correo electrónico: lmtj@ua.pt

**André Araújo:** Escola Superior de Tecnologia da Saúde do Porto, Instituto Politécnico do Porto / Departamento de Comunicação e Arte (DeCA), University of Aveiro. Portugal.

**Isabel M. Costa:** School of Health Sciences (ESSUA), University of Aveiro. Portugal.

Fecha de recepción: junio de 2013  
Fecha de aceptación: mayo de 2014

## Abstract

**Background:** The influence of the occlusal class in speech production has been studied using the X-ray Microbeam Speech Production Database (XRMB-SPD).

**Objectives/aims:** This study aimed to relate the occlusal classes I and II with speech production adaptations.

**Methods:** The Modified A-Space method was used to select 4 speakers (1 male and 1 female class I, 1 male and 1 female class II). Articulatory and acoustic features of the vowels were studied using different tasks and methods. The articulatory and acoustic features of consonants in male and female speakers of class I and class II from the XRMB-SPD were also described in de-

tail. Measures extracted from multitaper spectra and articulatory data were used, to observe individual differences related with gender and dental occlusion.

**Results:** Results showed some structural differences related to occlusal class and variance in class II subjects' structures and articulatory adaptations. However, subjects showed a high adaptation capacity, being able to adjust their articulations to produce all vowels.

**Conclusions:** Speech production variability is related with orofacial structures' variance. Different structures produce various functional adaptations and distinct speech signals.

**Keywords:** occlusal class; speech production; X-ray Microbeam Speech Production Database (XRMB-SPD).

## 1. Introduction

Speech is the most commonly used communication modality in all human cultures (except, of course, the deaf culture) and is comprised of an inventory of sounds produced by actions of the articulatory and phonatory systems. Speech is, however, a secondary function of these systems, which were primarily developed to perform functions such as respiration, swallowing or mastication.

Articulatory structures can influence the accuracy of speech sound production and the corresponding sound pressure signal. Just as we differ from each other in terms of height, weight or facial characteristics, cranial structures and the articulatory systems may have multiple features that frequently condition the way we speak.

Several authors (Vallino & Tompson, 1993; Lee, Whitehill, Ciocca, & Samman, 2002) have shown the occlusal class (Angle, 1907; Daskalogiannakis, 2000) to be directly related to the articulatory perturbation of some speech sounds such as fricatives. Speech and Language Therapists' assessment protocols usually include the evaluation of the occlusal class and several other parameters that allow them to understand the relation between orofacial structures and stomatognathic functions, including speech.

Occlusal class refers to the manner that the upper (maxilla) and lower (mandible) dental arches relate. This relation was described by Angle (1907), who proposed a malocclusion classification based on the relative position of the maxillary first molar (Daskalogiannakis, 2000). Occlusal class has been shown to be directly related to articulatory perturbation of speech sounds such as fricatives and vowels.

The classification proposed by Angle (1907) included different types of malocclusion, based on the mesiodistal relationship of the perma-

nent first molars upon their eruption and locking, and comprised three classes. Normal occlusion, the reference used to classify malocclusion, is characterised by an adequate alignment of the maxillary and mandibular dental arches, and a normal molar relationship, or neutroocclusion, wherein the mesiobuccal cusp of the maxillary first molar occludes in the buccal groove of the mandibular first permanent molar (Daskalogiannakis, 2000). Malocclusion refers to the misalignment of teeth and/or incorrect relation between the teeth of the two dental arches.

Class I malocclusion presents a normal molar relationship, while the other teeth have problems like spacing, crowding, over or under eruption. Class I and Normal Occlusion are sometimes used as synonyms.

Usually, the distobuccal cusp of the maxillary first permanent molar occludes in the buccal groove of the mandibular first molar (Daskalogiannakis, 2000). In Class II malocclusion, the upper molars are placed not in the mesiobuccal groove but anteriorly to it. This occlusal class is frequently associated with retrognathic facial types, normally resulting in a reduction of the anterior-posterior area and in compensations such as: tongue dorsum elevation, anterior mandibular sliding, swallowing modifications, and speech sound distortions.

Class III malocclusion, or mesiocclusion, refers to an advancement of the lower dental arch, wherein the mesiobuccal cusp of the maxillary first molar occludes in the embrasure between the mandibular first and second permanent molars (Daskalogiannakis, 2000).

This paper aimed to: study the influence of different malocclusion classes (I and II) in speech production describing the articulatory structures involved; comparing acoustic features and articulatory processes in vowel and obstruent<sup>1</sup>

---

1 Obstruents are produced when the pulmonic airflow is disturbed by an obstacle somewhere along the vocal tract, resulting in two distinct categories of turbulent sounds: fricatives and stops. They have common articulatory and aerodynamic features (Crystal, 2008).

production (Crystal, 2008); characterising the structure of the functional adaptations found.

## 2. X-Ray Microbeam Speech Production Database

The XRMB-SPD is a speech production database, created at the University of Wisconsin – Madison, USA, that uses X-Ray Microbeam technology to collect a vast amount of coordinate data describing articulatory movements, which also includes acoustic and electroglottographic data collected simultaneously (Westbury, 1994).

The XRMB-SPD articulatory data are presented in a two dimensional xy mid-sagittal plane which includes: palate trace, Middle Pharynx Wall (MPW) line, lips, tongue, and mandible. The coordinates of each mobile structure refer to an 8 pellet system distributed through the oral cavity: lower lip (LL), upper lip (UL), mandibular incisor (MANi), mandibular first molar (MANm), and tongue (T1, T2, T3 and T4). All the pellet and reference data are transformed onto a coordinate system defined by the maxillary occlusal plane (x-axis) and the apex of the maxillary incisor (centre).

The speech samples result from different tasks, i.e., word and sentence reading, isolated productions, and non-verbal oral movements. There are 57 male and female (average age of 21 years old) American English speakers. The database includes individual parameters characterising each subject such as dental information and occlusal class, which allowed us to study the relations between speech production and occlusal class (Westbury, 1994). Other structure information in the database includes: cranial measures, mandibular measures (superior edge of the central mandibular incisors, gnathion, gonion, condyle centre, and coronoid process), palate height, and maxillary arch dimensions. These measures were useful in the characterisation of the Articulatory Oral Space (AOS) of each subject, as well as in their selection.

## 3. Method

### 3.1. Subject Selection and Characterisation

Four subjects, out of the 57 speakers in XRMB-SPD, were selected, representing four distinct groups regarding gender and occlusion (occlusal class I and II). The first selection procedures included various criteria related to gender, occlusal class, data availability (some data were missing from the freely distributed XRMB-SPD CDs), and each subject's individual information. The selection that followed this initial data analysis was based on the Modified A-Space method (Jesus, Araújo, & Costa, 2007), which is an extended and updated version of the A-space method proposed by Honda et al. (1996).

The swallowing, maximal tongue and lip protrusion and replicative jaw-“wagging” tasks were also characterised in the selected subjects using the outputs of *TF32* software (Milenkovic, 2001) to assess structural and functional differences not related to speech production (Araújo, 2007).

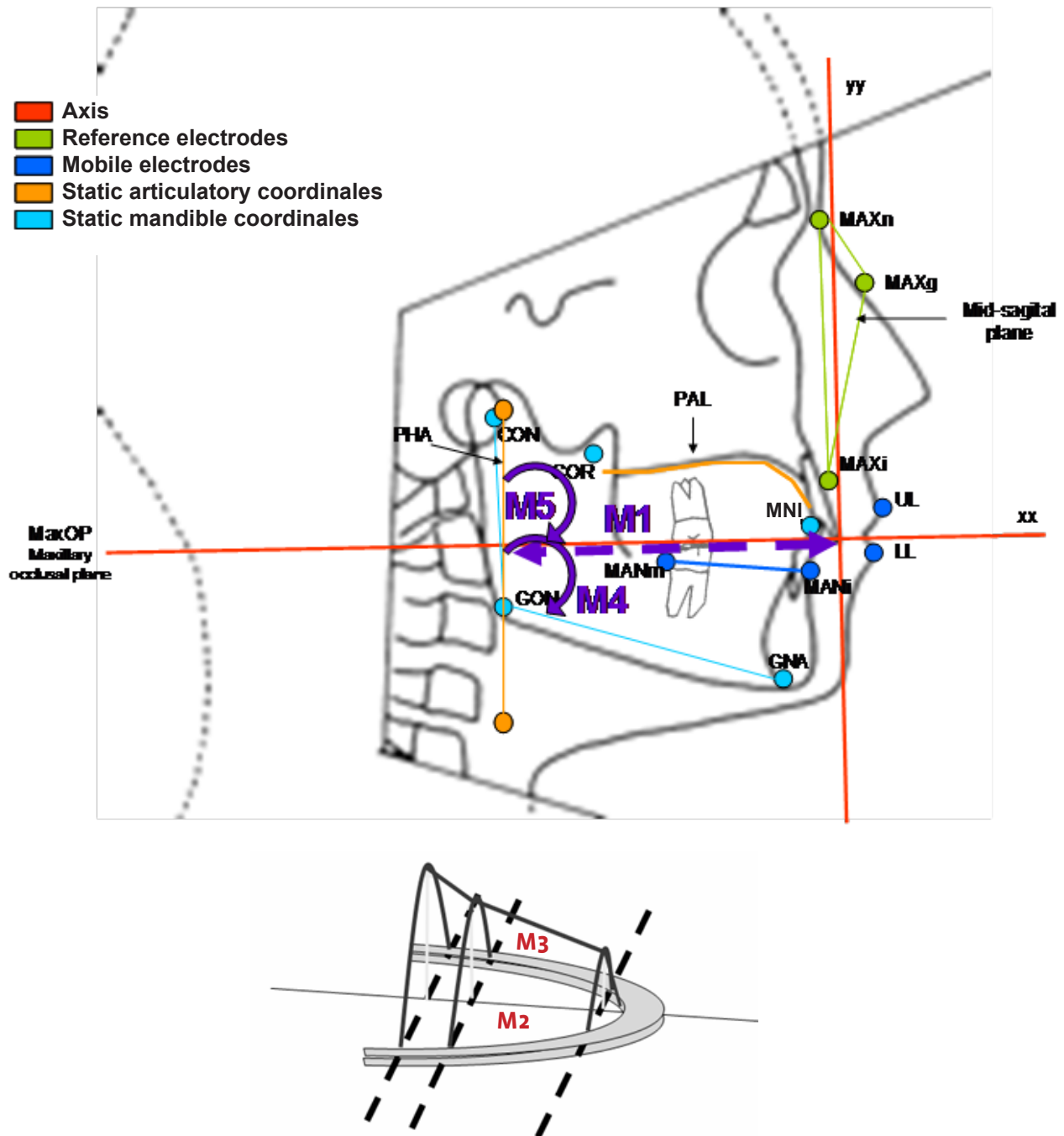
### 3.2. The Modified A-Space

XRMB-SPD provides occlusion classification, dental measures, anthropomorphic measures, reference pellets coordinates, biteplate records, and palatal outlines, for each of the 57 speakers. This was used to measure the articulatory oral space (AOS) in the absence of cephalometric analysis, based on the Modified A-Space described in Figure 1.

The Modified A-Space method (Jesus et al., 2007) allowed the detailed characterisation of the XRMB-SPD speakers not just in terms of mid-sagittal-plane area, but also in terms of antero-posterior distance, occlusal plane area, posterior pharynx wall tilt, mandible arch width, and oral cavity volume. This last measure has proven to be far more reliable and dependent of individual speaker's characteristics than the measure previously proposed in Honda et al. (1996).

**FIGURE 1**

Top – Mid-sagittal-plane coordinates included in the XRMB-SPD<sup>2</sup>. Bottom – a three dimensional representation of the maxillary arch and mid-sagittal palate height of the anterior oral cavity (from the distal-buccal cusp tip of the second molar to the lips). The Modified A-Space measures M1, M2, M3, M4, and M5 are also represented.



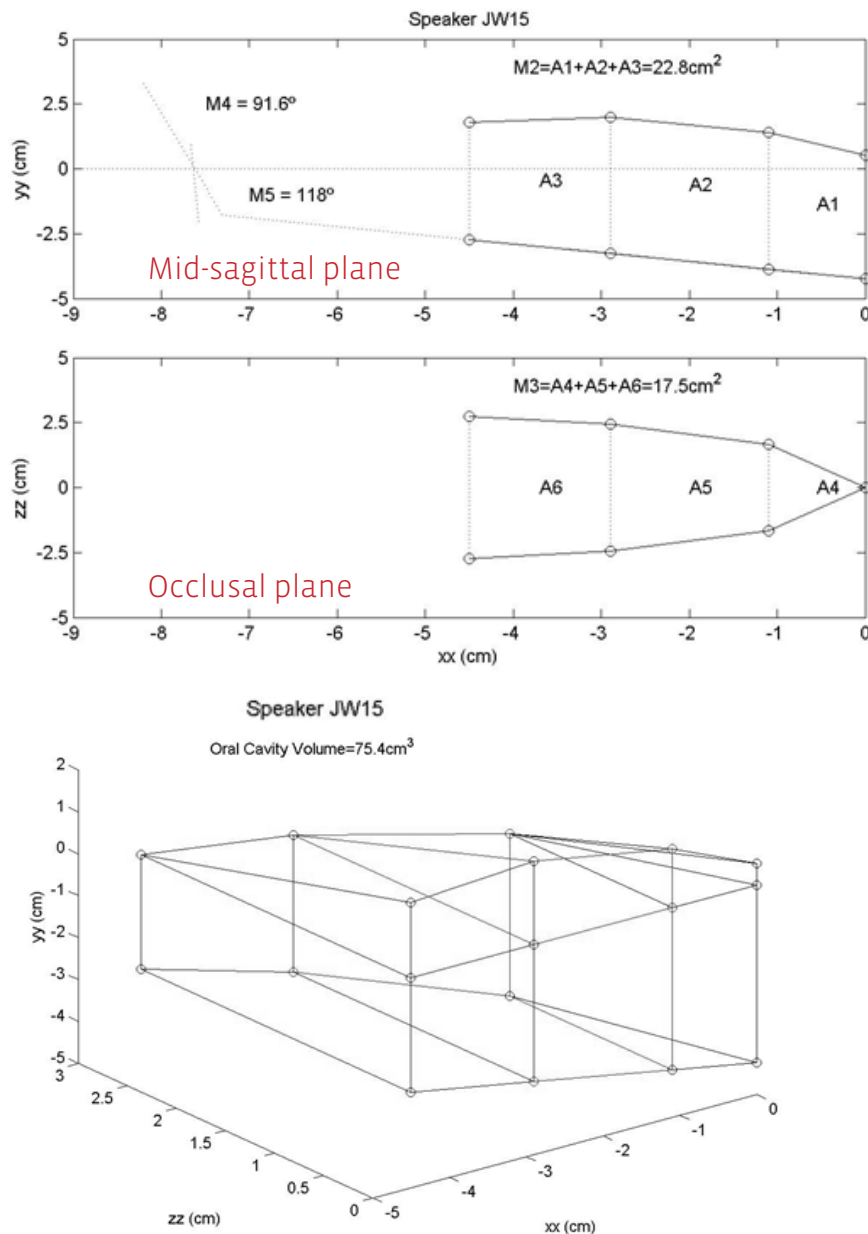
2 MAXn and MAXg – bridge of the nose; MAXi – buccal surface of the maxillary incisors; MANm – juncture between the first and second mandibular molars; MANi – buccal surface of the central incisors; LL – lower lip; UL – upper lip; PAL – palate; PHA – middle pharynx wall; CON – condyle; COR – coronoid process; GON – gonion; GNA – gnathion; MNI – lingual surface of the maxillary incisors.

We extracted the following measures of the AOS, as shown in Figure 1 and Figure 2: M1 – antero-posterior distance, calculated from the upper incisors to the posterior pharynx wall; M2 – mid-sagittal plane area, from the mandible to the palate midline; M3 – occlusal plane area, from the distal-buccal cusp tip of the second molar to the lips; M4 – posterior pharynx wall tilt, i.e.,

the angle between the pharynx and the occlusal planes; M5 – mandible arch angle, calculated with several mandible points; M6 – anterior oral cavity volume. Areas of trapeziums (A1, A2, A3, A5, and A6) and a triangle (A4), and volumes of convex hulls of cubes and tetrahedrons were used to estimate the AOS, as shown in Figure 2.

**FIGURE 2**

Measures M2, M3, M4, M5 and Oral Cavity Volume for speaker JW15, showing the mid-sagittal and occlusal planes (top) and half of the oral cavity volume as reconstructed using the Modified A-Space (bottom)



### 3.3. Corpus

Vowels [i]<sup>3</sup>, [ɛ], [A], and [u] were selected from several tasks of XRMB-SPD, in various phonetic contexts (Wells, 1997; Araújo, Jesus, & Costa, 2007). Acoustic analysis was based on vowels produced in isolation, preceded by [s] and followed by [d] (the words [sid], [sɛd], [sAd], and [sud]), and several productions in various words, totalising 10 [i], 7 [ɛ], 5 [A] and 5 [u] productions.

Two classes of consonant sounds were also selected for the study: the voiceless fricatives /f, s, S/ and the voiceless stops /p, t, k/. Articulatory analysis was based on materials from a citation VCV<sup>4</sup> task, where the speaker was asked to produce each consonant between two vowels (initial /V/ and final /a/). The consonants were produced in a stressed intervocalic context: [V<sup>"</sup>fa]<sup>5</sup>, [V<sup>"</sup>sa], [V<sup>"</sup>Sa], [V<sup>"</sup>pa], [V<sup>"</sup>ta], and [V<sup>"</sup>ka]. In the acoustic analysis the same nonsense words from the citation VCV task were used, and three additional words of each consonant were selected from a few other tasks: [f] in the words ‘before’, ‘beautiful’, and ‘information’; [s] in the words ‘yourself’, ‘himself’, and ‘conversation’; [S] in the words ‘special’, ‘information’, and ‘conversation’; [p] in the word ‘people’ (we analysed three repetitions of this word because there were no other real words with [p] available); [t] in the words ‘between’ and ‘dormitory’ (repeated twice); and [k] in the words ‘second’, ‘become’, and ‘because’. The XRMB-SPD contains rich variation in consonantal productions, especially related to a large variety of coarticulation phenomena and various contexts of each consonant in the words.

### 3.4. Annotation and Acoustic analysis

Acoustic analysis of vowels was performed manually using *TF32* (Milenkovic, 2001) functions *Wave plot*, *TimeFreqA*, and *Spec*. Formant frequencies  $f_1$ ,  $f_2$ , and  $f_3$ <sup>6</sup> were extracted from a stable region in the spectrogram of each vowel through the peaks of LPC spectra and cross-checked with values extracted from spectrograms. Formant values were then converted from Hertz (Hz) to Bark (Zwicker & Terhardt, 1980) and used to represent each subject’s vowel space, in which perceptually equal intervals are represented as equal distances along the x and y axes.

Selected fricatives and stops were manually annotated using *TF32*, and text files were produced containing a series of marked speech events for later analysis using *Matlab*. The phases considered during fricative annotation were: the “stable” phase (middle) of the previous vowel, the start of the vowel-fricative transition, the start of the fricative, the start of the fricative-vowel transition, the start of the following vowel, and the “stable” phase of the following vowel. The acoustic phases considered in the stop annotation were: the “stable” phase of the previous vowel, beginning of the closure, beginning of the release, the start of the following vowel, and the “stable” phase of the following vowel.

Multitaper spectra was also calculated using an 11 ms window centred at the middle of each fricative and left aligned at the beginning of each stop release (Lousada, Jesus, & Pape, 2012; Zygis, Pape, & Jesus, 2012). We estimated the power spectral density (PSD) via the Thom-

3 Phonetic transcriptions in SAMPA (Wells, 1997) in accordance to the International Phonetic Association guidelines. The vowel and consonant sounds in American English words from the XRMB-SPD are transcribed using the following conventions: /i/ - vowel in ‘seed’; /ɛ/ - vowel in ‘sad’; /A/ - vowel in ‘sod’; /u/ - vowel in ‘sued’; /f/ - second consonant in ‘before’; /s/ - second consonant in ‘yourself’; /S/ third consonant in ‘special’; /p/ - first consonant in ‘people’; /t/ - fourth consonant in ‘dormitory’; /d/ - first consonant in ‘dormitory’; /k/ - second consonant in ‘second’.

4 Vowel consonant vowel (VCV).

5 /<sup>"</sup>/ marks the beginning of a stressed syllable.

6 Vowels are perceptually distinguished from each other by two ( $f_1$  and  $f_2$ ) characteristic overtones known as formant frequencies. A formant is an acoustic resonance generated during vowel production that can be observed as a spectral peak, using, for example, LPC spectra or spectrograms (Crystal, 2008).

son multitaper method (linear combination with unity weights of individual spectral estimates and the default FFT length) available in Mathworks Signal Processing Toolbox Version 6.2. There are several methods (e.g., time averaging and ensemble averaging) for reducing the variance, but Blacklock (2004) has shown that multitaper analysis provides increased control over the resolution-bias-variance trade off, and spectra calculated this way can be used to estimate the underlying system variance, previously swamped by estimate errors.

Spectral peaks are due to the poles of the vocal tract frequency response. Spectral troughs are due to the zeros of the vocal tract frequency response. Moving the articulators alters the shape of the vocal tract, which in turn changes its frequency response (Shadle & Scully, 1995: 59-60). Spectral peaks tend to be the most prominent feature in the spectra of speech; they will, therefore, be referred to as one of the most important acoustic characteristics. Shifts in any pole or zero frequency affects all peaks and troughs, sometimes substantially (Stevens, 1998: 130-137), and radiation impedance increases bandwidth as frequency increases, but much more for front cavity resonances (Stevens, 1998: 152-156). The fact that there is noise excitation during the production of obstruents increases the difficulty of determining cavity affiliation of peaks and troughs. However, peaks and troughs which appeared consistently were used to describe the spectra of American English fricatives and stops.

Due to the extreme difficulty of picking the troughs and peaks “automatically”, all data presented in the following sections resulted from a careful examination of the obstruent spectrum and a “manual” measurement of peak frequencies and amplitudes. This time consuming research method was deemed necessary for a reliable characterisation of fricatives and stops in the frequency domain.

The most relevant peaks and troughs were

manually identified, using the same criteria throughout the corpus. We then extracted the centre frequencies of the following number of peaks and broad peaks that could be consistently observed for each sound: one broad peak in [S]; two peaks in [f], [t], and [k]; and three peaks in [s] and [p]. The peaks were manually marked using two straight lines that closely modelled each peak, as can be seen in Figure 9, allowing the accurate calculation of the central frequency, especially for blurred or low amplitude peaks. The peak frequencies were exported into *Excel*, to allow the comparison with bibliographic data and analysis of inter-speaker variability, enabling us to relate the acoustic results with data from the articulatory analysis.

### 3.5. Articulatory Analysis

Articulatory analysis of vowels was based on productions where the vowels were produced in preceded by [s] and followed by [d]. The coordinates of all pellets in the middle of the vowel were exported to text files to allow further processing with *Matlab*. Images and measures describing the articulatory configuration of each vowel produced by each subject were also exported and superimposed to allow a comparative evaluation of speech production characteristics. Four parameters were analysed after image editing: tongue posture, tongue elevation, mouth opening, and lip configuration. These data were then related with acoustic analysis results.

Articulatory analysis of obstruents was based only in citation VCV task materials. We developed a method to study the articulatory differences between the four subjects during the production of the selected sounds, which included the analysis of the place of articulation, articulators’ configuration, and trajectories.

During the annotation phase, we exported xy trajectory plots, such as the one shown in Figure 8, from the *TF32* graphical display, representing the articulatory configuration at the middle of fricatives and at the middle of the stops’ closure. The trace function of *TF32* has also been used, to



visualise a line that represented the movement of each pellet, from the anterior to the posterior vowel (see Figure 8). The trajectories of the selected speakers were overlapped, which allowed a comparative description.

The coordinates at the mid-fricative and mid-stop closure points were exported and analysed in *Matlab*, where graphical representations such as those shown in Figure 3 were produced, and the following measures extracted (see Figure 4): distance from pellet T1 to the palate (T1-P), distance from pellet T2 to the palate (T2-P), distance from pellet T3 to the palate (T3-P), distance between the place of maximum constriction (between the tongue and the palate) and the front teeth (PtX), and distance between the upper and lower lip (L-L).

## 4. Results

### 4.1. Subject Characterisation and Selection

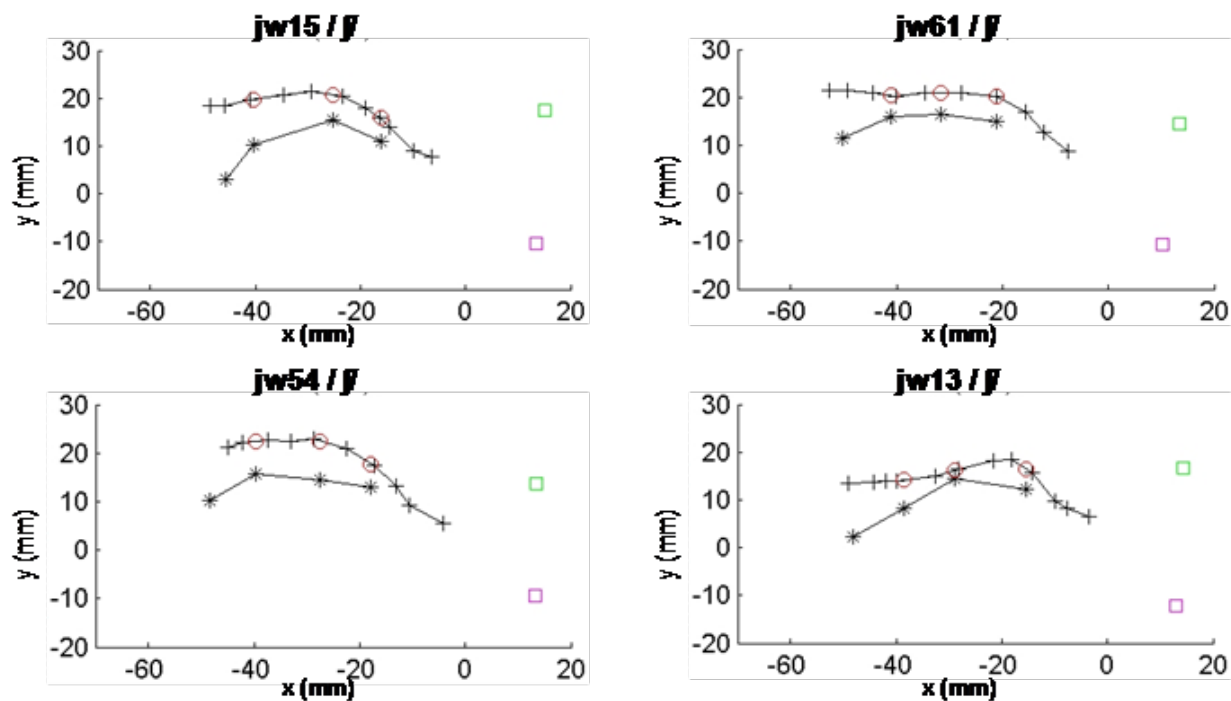
Materials available in the XRMB-SPD were thoroughly inspected and speakers' directories

with missing or incomplete task files were not considered. All remaining subjects were then grouped according to their occlusal class and gender, and then scatter plots (e.g., M1 vs. M2 and M1 vs. M3) of various Modified A-space measures were analysed. The most representative (closest to the central/median value of the scatter plots) speakers of each group were set apart for further analysis.

The four selected subjects were: jw15, class I male; jw61, class II male; jw54, class I female; and jw13, class II female (Araújo, 2007: 44-47). Class II malocclusion subjects present a slight AOS reduction, due to the anterior and tipped position of the posterior pharynx wall, and a posteriorised tongue apex during swallowing. However, other functional behaviours suggest that speakers jw61 and jw13, both class II, may have other occlusal differences besides the occlusal class, since speaker jw61 places his tongue apex further back than speaker jw13, and speaker jw13 frequently advances his jaw, suggesting a deep

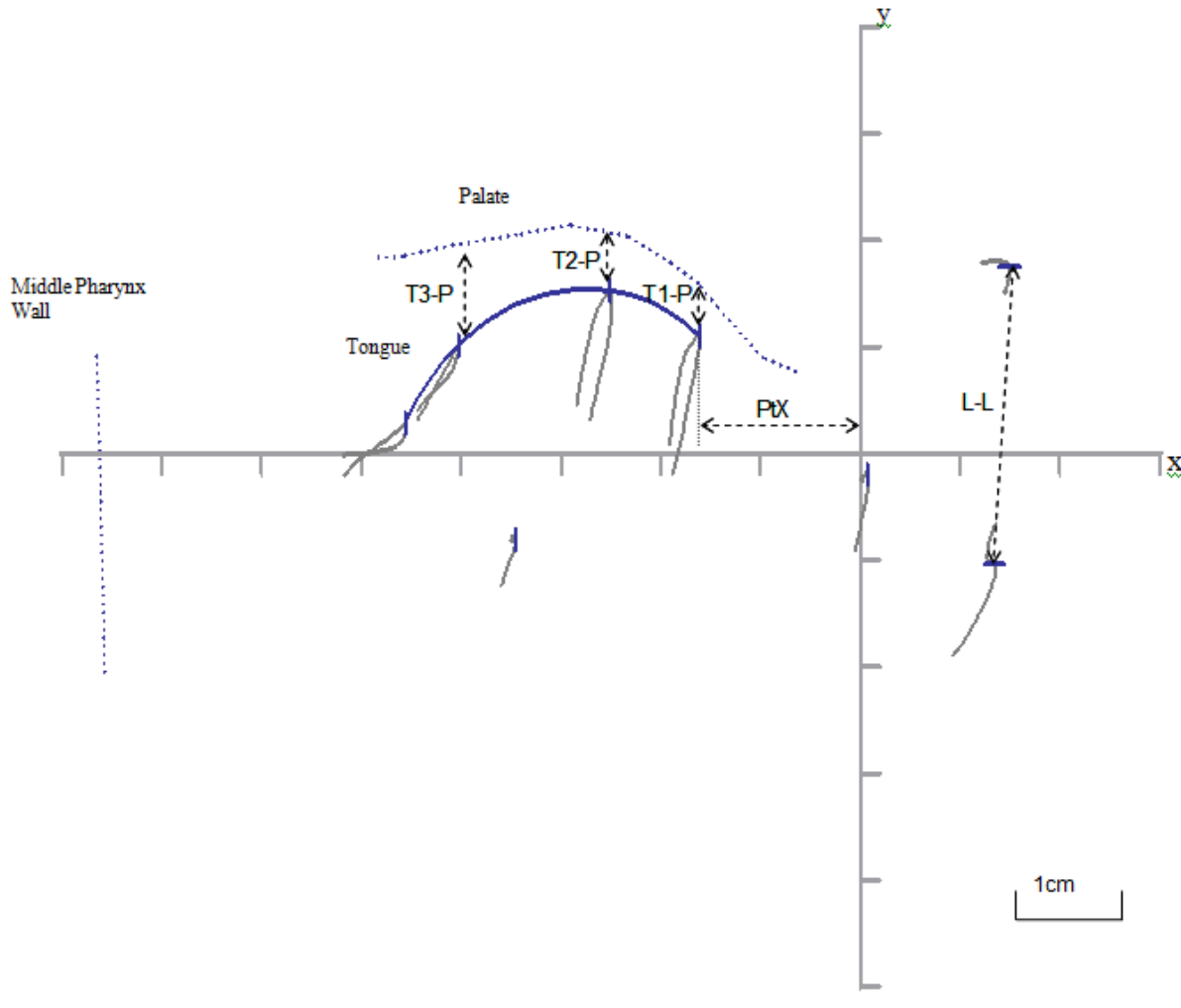
**FIGURE 3**

Articulatory configurations at the middle of [S] production by speakers jw15, jw61, jw54, and jw13



**FIGURE 4**

TF32 graphical display of articulatory configuration and trajectories (in grey) during [S] production by speaker jw15 and articulatory measures (T1-P, T2-P, T3-P, PtX, and L-L) extracted with Matlab



bite. The palate line configurations were similar for all subjects except for speaker jw13, with a 0.5 cm reduction in height at its posterior end ( $x = -3$  cm), as shown in Figure 6. However, there is not a considerable reduction of the AOS related to this structural difference.

#### 4.2. Articulatory Measures of Obstruents

The articulatory analysis of obstruents presented in the following sections, which was based on comparative observations of overlapping articulatory configurations and trajectories of each phone production by the four speakers,

was complemented with the extraction of some measures shown in Table 1.

#### 4.3. Acoustic and Articulatory Analysis of vowels

Male's formant frequency values, shown in Table 2, were generally lower than female ones, as expected. There does not seem to be any considerable differences in male speakers related to malocclusion, as shown in Figure 5. However, the Class II female speaker JW13 used a considerably wider vowel space than the Class I female speaker JW54.

**TABLE 1**

Articulatory measures (median values) for [f, s, S, p, t, k] produced by speakers jw15, jw61, jw54, and jw13<sup>7</sup>

[f]	T1-P (mm)	T2-P (mm)	T3-P (mm)	PtX (mm)	L-L (mm)	[p]	T1-P (mm)	T2-P (mm)	T3-P (mm)	PtX (mm)	L-L (mm)
jw15	21	24	17	43	19	jw15	29	26	20	45	17
jw61	22	17	13	45	19	jw61	25	21	16	47	18
jw54	24	22	12	42	12	jw54	20	21	12	41	15
jw13	16	12	11	44	16	jw13	23	20	16	42	15

[s]	T1-P (mm)	T2-P (mm)	T3-P (mm)	PtX (mm)	L-L (mm)	[t]	T1-P (mm)	T2-P (mm)	T3-P (mm)	PtX (mm)	L-L (mm)
jw15	4	12	20	11	25	jw15	7	12	17	17	25
jw61	7	12	12	17	21	jw61	6	8	8	17	22
jw54	6	16	16	15	19	jw54	3	13	13	13	19
jw13	4	14	11	11	21	jw13	5	9	10	12	21

[S]	T1-P (mm)	T2-P (mm)	T3-P (mm)	PtX (mm)	L-L (mm)	[k]	T1-P (mm)	T2-P (mm)	T3-P (mm)	PtX (mm)	L-L (mm)
jw15	5	5	9	16	28	jw15	16	14	1	38	26
jw61	5	4	4	41	26	jw61	10	4	1	42	23
jw54	5	8	7	18	23	jw54	10	6	1	32	21
jw13	4	2	6	29	29	jw13	9	8	2	37	25

**TABLE 2**

Mean f1 (Hz) and f2 (Hz) of vowels produced by the four subjects and values from Peterson & Barney (1952)

	[i]		[ɨ]		[A]		[u]	
	f1 (Hz)	f2 (Hz)	f1 (Hz)	f2 (Hz)	f1 (Hz)	f2 (Hz)	f1 (Hz)	f2 (Hz)
P&B (1952) ♂	270	2290	660	1720	730	1090	300	870
JW15	321	2025	703	1737	726	1203	361	965
JW61	313	2062	698	1578	730	1142	375	963
P&B (1952) ♀	310	2790	860	2050	850	1220	370	950
JW54	395	2367	667	2015	811	1451	439	1123
JW13	332	2468	642	2250	919	1447	388	1035

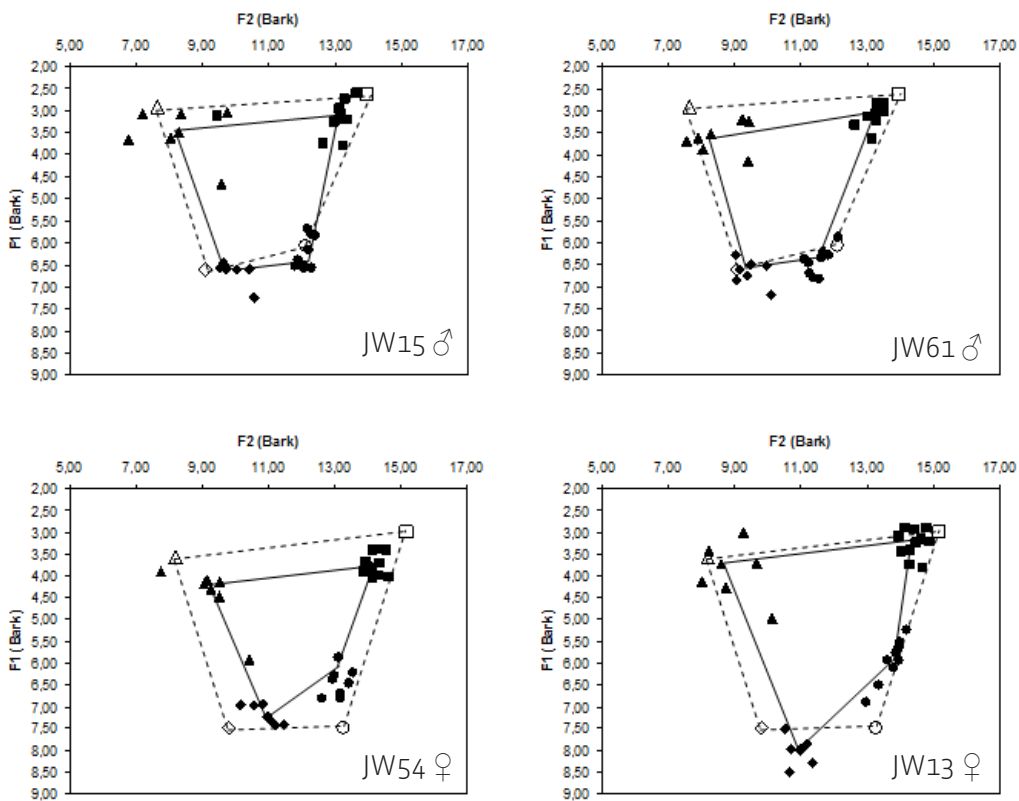
In [A] production we can observe a difference between Class II subjects (see Figure 5). Both have higher f2 frequency values than those reported in Peterson & Barney (1952).

Furthermore, f1 frequency values were lower for speaker JW54 and higher for speaker JW13 than the average values in Peterson & Barney (1952).

7 T1-P, distance from pellet T1 to the palate; T2-P, distance from pellet T2 to the palate; T3-P, distance from pellet T3 to the palate; PtX, distance between the place of maximum constriction and the front teeth; L-L, distance between the upper and lower lip pellets.

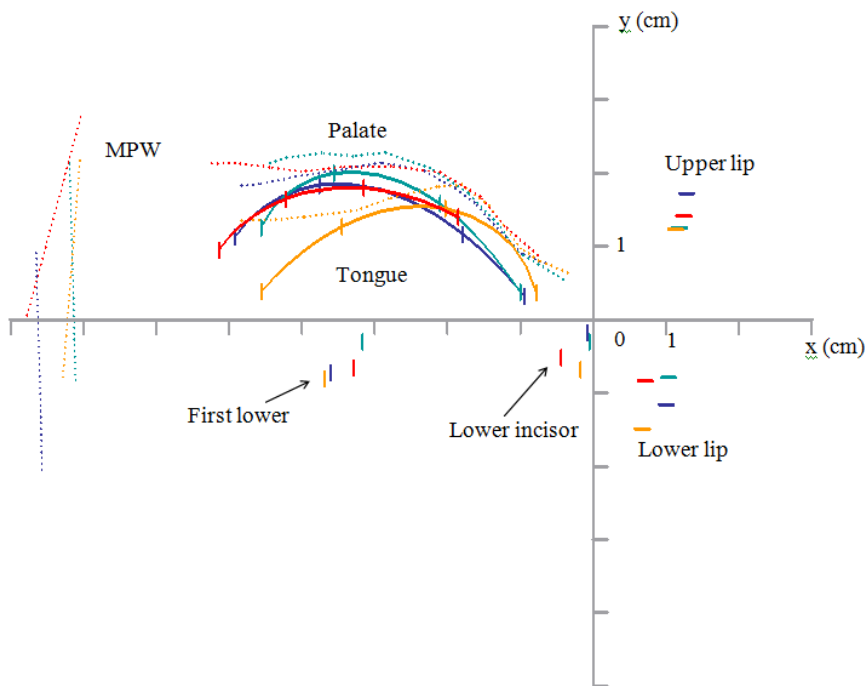
**FIGURE 5**

Vowel spaces of speakers JW15, JW61, JW54, and JW13 (filled lines), and from Peterson & Barney (1952) (dashed lines).  
 ■□ - [i], ●○ - [ɜ], ◆◇ - [A], and ▲△ - [u]



**FIGURE 6**

Articulation of [i] by speakers JW15 (blue), JW61 (red), JW54 (green), and JW13 (yellow).



Vowels [i, ʌ, u] produced by speaker JW54 had higher f1 frequencies and the ones produced by JW13 presented lower f1 frequencies than reference values (Peterson & Barney, 1952).

The [i] production results, shown in Figure 6, for speakers JW15 and JW54 present an elevation of the medium part of the tongue towards the palate. JW61 presents a similar elevation but a more posterior position of the tongue apex. JW13 also elevates the tongue at the most frontal region. Both Class II subjects (JW61 and JW13) present a more posterior position of lower incisors and lips than Class I subjects (JW15 and JW54). The mandible elevation seems to be similar in all the subjects.

Production results for vowel [A] have shown that speaker JW54 presents the highest tongue, mandible, and lower lip position. JW13 presents the lowest upper and lower lip and tongue position. JW61 presents an elevated tongue dorsum and tongue back position, relative to the other subjects. JW61's lower incisor and lip pellets, also indicate a more posterior position of the

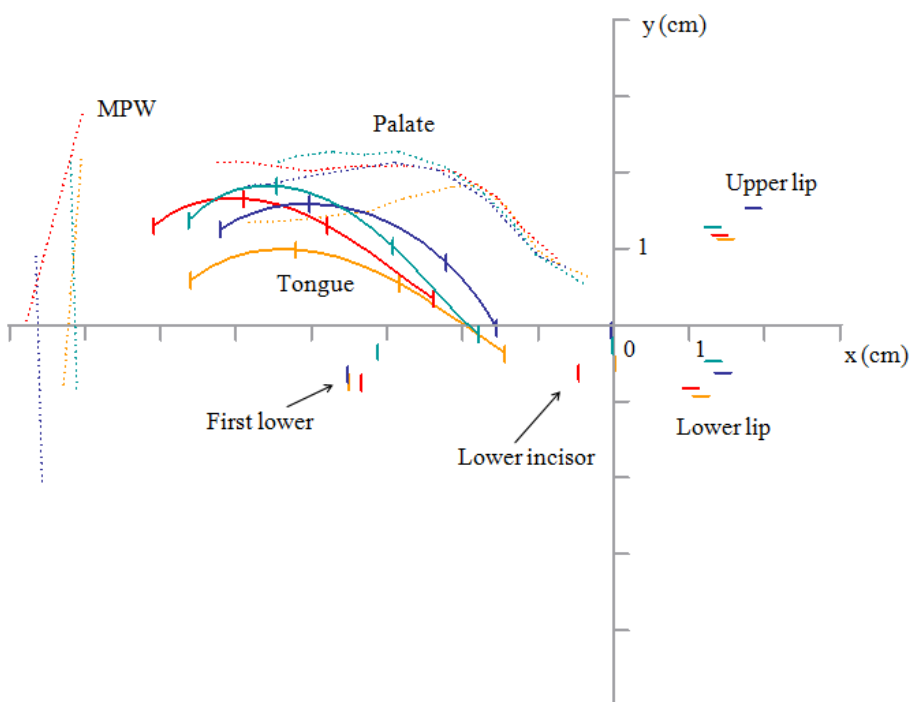
mandible than Class I subject JW15. Results also showed mandible advancement in JW13's production of [A].

Speakers JW15 and JW54 produced [ʌ] with similar tongue, mandible, and lip positions. However, the medium region of JW54's tongue had a higher elevation. JW61's articulatory position was more posterior than Class I subjects (JW15 and JW54): the tongue was almost horizontal and retracted from apex to the back of the dorsum, and the mandible and lower lip were also in a more posterior place. Speaker JW13 had a more anterior tongue position and wider mouth opening than Class I subjects.

In [u] production (see Figure 7) all subjects presented an elevation of the tongue dorsum. Although JW13 tongue position was lower than the other subjects, the distance to his palate was approximately the same. The mandible height was roughly the same in all subjects, and there was a slight lip advancement for this vowel when compared with [i, ʌ, A], as expected for a rounded vowel such as [u].

### FIGURE 7

Articulation of [u] by speakers JW15 (blue), JW61 (red), JW54 (green), and JW13 (yellow)



#### 4.4. Acoustic and Articulatory Analysis of [f]

The two peaks in [f] spectra selected for analysis were centred at frequencies that varied between 1.4 and 3.9 kHz, and 6.9 and 8.6 kHz. The median frequencies of each peak for the four speakers were: 2 and 8.2 kHz for jw15; 1.4 and 7.2 kHz for jw61; 1.7 and 8.2 kHz for jw54; and 2.9 and 8 kHz for jw13. These values do not seem to correspond to those described by Narayanan (1995), who only mentioned one broad peak around 10 kHz.

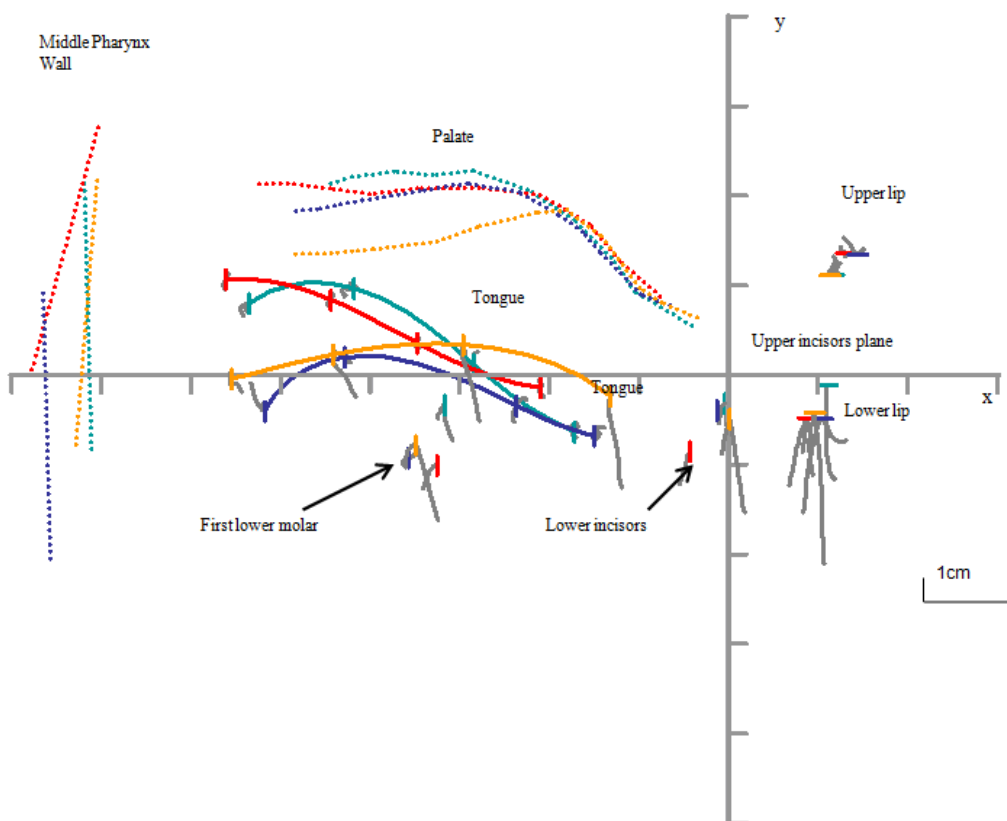
The comparative analysis of spectra and articulatory configuration did not show any relevant differences, in terms of gender and dental occlusion. We should bear in mind that [f] is a labiodental fricative, having a small front resonance cavity, which results in a wide range of low-amplitude and narrow-band peaks, that are difficult to identify. What makes a difference

when a speaker produces an [f] is the shape of the extremely short front cavity delimited by the upper teeth and the lower lip, and how efficiently we push air out of the oral cavity through it (Stevens, 1998: 389-398).

Figure 8 shows the overlapping articulatory configurations and trajectories of one of the [f] productions that have been studied. The lower lip shows the most significant displacement in all subjects and the mandible sustains part of the upward movement. The tongue does not show a posture change relative to the anterior vowel, confirming that it does not actively contribute to the production of this sound. The values of L-L, shown in Table 1, indicate that for [f], there is less lip opening than for the other fricatives. This lip closing is due to the upward movement and slight posteriorisation of the lower lip, as can be seen in Figure 8, since there is practically no movement of the upper lip.

**FIGURE 8**

Overlapping articulatory configurations and trajectories (grey) of [f] production by speakers jw15 (blue), jw61 (red), jw54 (green), and jw13 (orange)



Female speakers' (jw54 and jw13) upper and lower lips were almost vertically aligned, while the inferior lips of males (jw15 and jw61) were slightly anteriorised, what can also be observed in the backward movement of the lower incisors. This configuration can explain the fact that L-L is lower for female than for males (see Table 1).

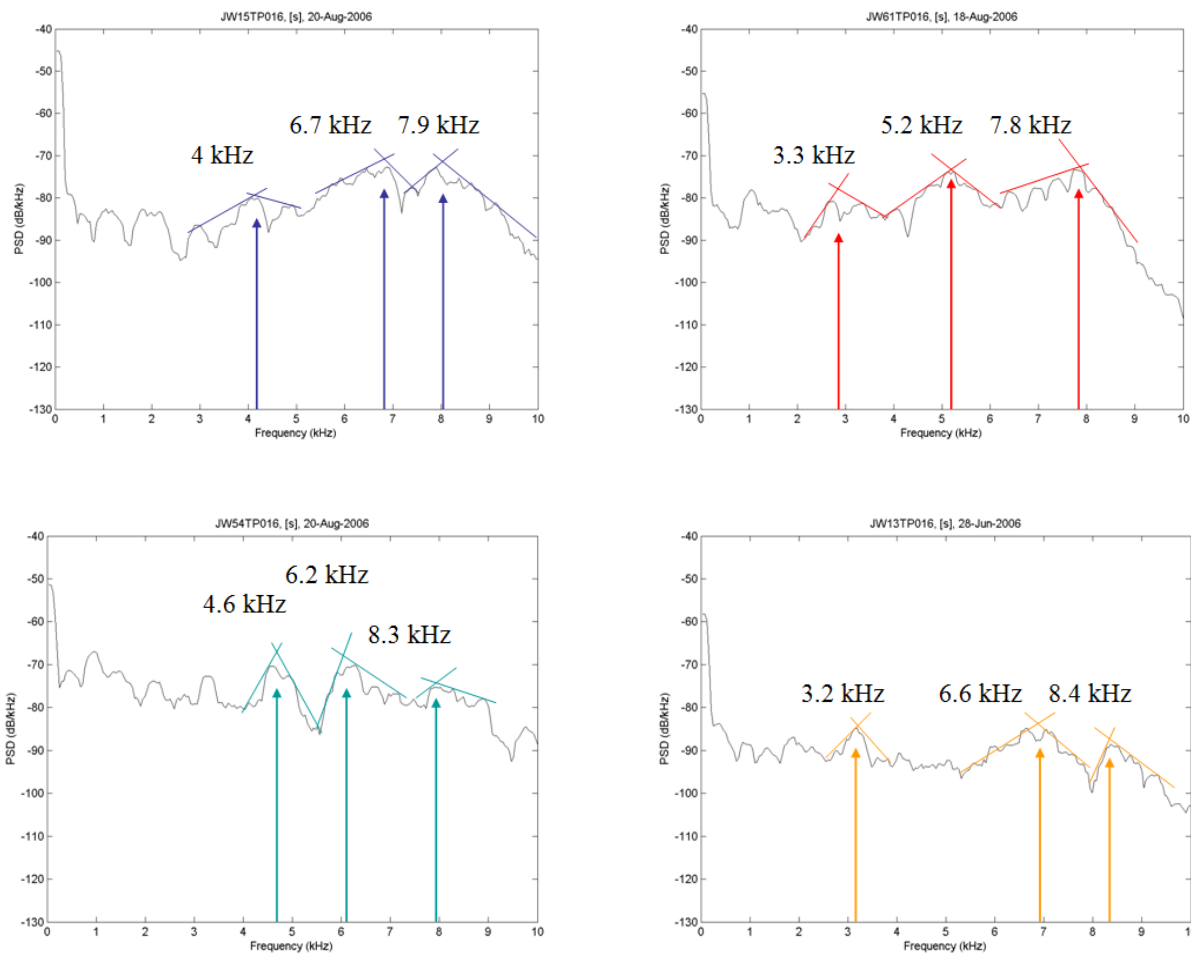
#### 4.5. Acoustic and Articulatory Analysis of [s]

It was possible to always clearly identify

three peaks in the spectra of [s], shown in Figure 9. The values varied between 2.5 and 4.6 kHz, 4.5 and 6.8 kHz, and 7.2 and 8.6 kHz. The median frequencies of each peak were: 4, 6.7, and 7.9 kHz for speaker jw15; 2.9, 4.8, and 7.8 kHz for speaker jw61; 4.2, 5.4, and 8.3 kHz for speaker jw54; 3.1, 6.3, and 8.4 kHz for speaker jw13. The resulting values are similar to those presented by Narayanan (1995), who described a broad peak between 5 and 6.6 kHz and secondary peaks between 1.6 and 1.8 kHz, 2.5 and 2.9 kHz, and 4.6 and 4.8 kHz.

**FIGURE 9**

Multitaper spectra of [s] produced in citation VCV task by speakers jw15 (upper left), jw61 (upper right), jw54 (lower left), and jw13 (lower right)



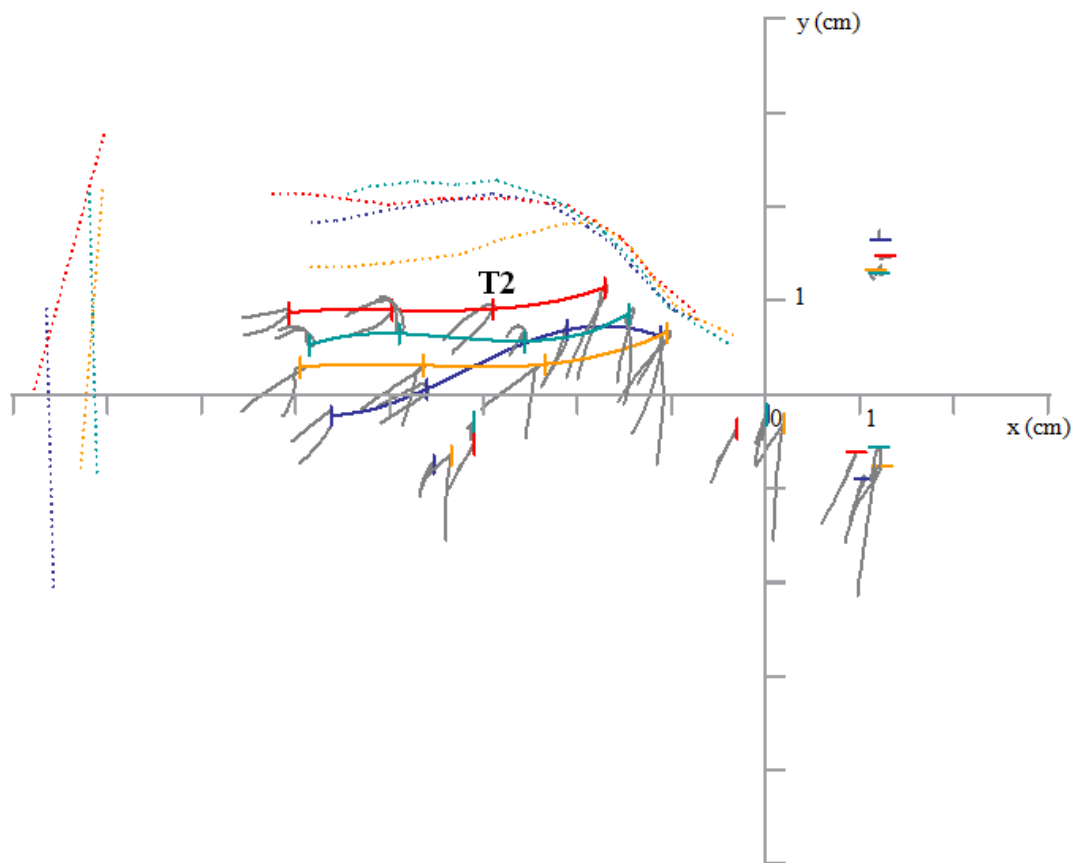
Spectral analysis of [s] revealed that the median frequency of the first peak (2.5-4.6 kHz) is related to dental occlusion since it was higher in class I (4-4.2 kHz) than in class II (2.9-3.1 kHz). The frequency of the third peak (7.2-8.6 kHz) varies with gender, i.e., medians are higher in women (8.3-8.4 kHz) and lower in men (7.8-7.9 kHz). The second peak of [s] spectrum seems to be related to the place of the constriction for each subject.

Figure 10 shows that the tongue apex moves towards the alveolar region and that the lower incisors move toward the upper incisors. The place of articulation is most anterior for jw15 and jw13, and gradually more posterior for jw54 and jw61. Only speaker jw15 shows a slightly different tongue configuration, with a higher T2 position.

Data in Table 1 shows that the smallest constriction occurs, as predictable, at the tongue apex (T1). The values of PtX (11-21 mm) also indicate the place of articulation to be around T1, as shown in Figure 10. Speakers jw15 and jw13 show the same place of articulation (PtX), 11 mm from the incisors and identical second peak frequency values (6.7 and 6.6 kHz respectively). Speaker jw54 presents a PtX measure of 19 mm and a decrease of the second peak's spectral frequency to 6.2 kHz. Speaker jw61 exhibited the most posterior place of articulation (21 mm) and a decrease of the second peak frequency to 5.2 kHz. The second peak frequency is therefore (inversely) related with a resonance of the anterior cavity, which increases in size with the place of articulation posteriorization (Stevens, 1998: 398).

**FIGURE 10**

Overlapping articulatory configurations and trajectories (grey) of [s] production by speakers jw15 (blue), jw61 (red), jw54 (green), and jw13 (orange)





#### 4.6. Acoustic and Articulatory Analysis of [S]

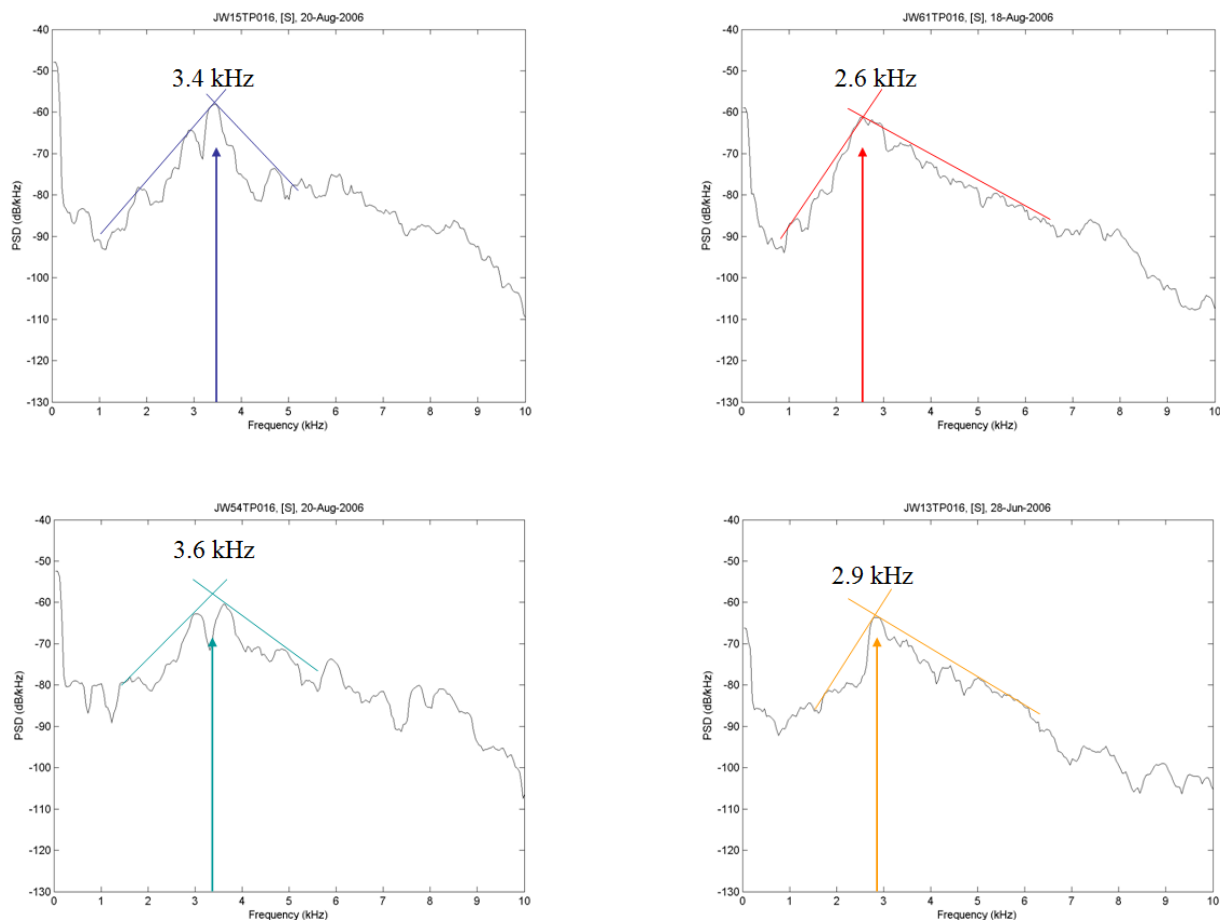
Fricative [S] was acoustically characterised in terms of only one broad-peak that results from a front cavity resonance (Stevens, 1998: 406). Its centre frequencies vary between 2.3 and 3.8 kHz, and the median values for each speaker were: jw15 – 3.1 kHz; jw61 – 2.5 kHz; jw54 and jw13 – 3 kHz. The analysis of citation VCV task spectra, shown in Figure 11, suggested differences in terms of dental occlusion.

Pellet T1 is located more anteriorly during [S] production by speaker jw15 and jw13 than for jw54 and jw61, as had happened in [s] pro-

duction. However, considering the measures presented in Table 1, one can conclude that the most severe constriction normally occurs between pellets T2 and T3 for jw61, and around T2 for jw13, both occlusal class II speakers. This seems to be related with spectral results, since the broad peak values are lower for class II (2.6 and 2.9 kHz) than for class I (3.4 and 3.6 kHz) speakers, which can be related to a more posterior place of articulation and therefore a longer anterior resonance cavity. One other important element in this sound articulation seems to be labial protrusion, especially the upper lip when compared to other sounds.

**FIGURE 11**

Multitaper spectra of [S] produced in citation VCV task by speakers jw15 (upper left), jw61 (upper right), jw54 (lower left), and jw13 (lower right)



#### 4.7. Acoustic and Articulatory Analysis of [p]

The spectra of [p] consistently showed three main peaks whose frequency varied between 0.5 and 2.6 kHz, 2.8 and 4.9 kHz, and 7.1 and 8 kHz. The medians of each peak frequency, for each speaker, were: jw15 – 1.1, 3.4 and 7.3 kHz; jw61 – 1.2, 4.2 and 7.4 kHz; jw54 – 1.2, 3 and 7.8 kHz; jw13 – 1, 2.8 and 7.2 kHz. The peak values were very similar for all speakers, and the first peak frequency was slightly higher than the value (0.6 kHz) usually described in the literature (Raphael, Borden, & Harris, 2007: 150).

There was a significant displacement (closure) of both lips, as expected, and almost no tongue movement during the VCV articulation. Major lip and mandible pellet displacement were visible, e.g., the upper lip showed a steeper downward movement than any other obstruent. Despite the upward displacement of mandible pellets, supporting the occlusion of the oral cavity, the displacement of the lower lip was preponderant. There did not seem to be any difference related with occlusal class.

#### 4.8. Acoustic and Articulatory Analysis of [t]

Stop [t] had spectral peaks around 2-3.9 kHz and 7-8.6 kHz. The median frequencies were: 3.5 and 8.3 kHz for speaker jw15; 2.6 and 7.4 kHz for speaker jw61; 2.8 and 8.3 kHz for speaker jw54; 3.5 and 8.3 kHz for speaker jw13. The values of the first peak were very similar to those (3 kHz) described in classical literature (Raphael et al., 2007: 150). The second peak was quite similar for all speakers, with the exception of speaker jw61. The frequency of the first peak was extremely variable, and apparently not related to gender or dental occlusion.

The tongue apex approaches the alveolar region and also some mandible elevation,

during [t] production. All subjects show an alveolar place of articulation as described in the literature (Raphael et al., 2007: 146-152). Nevertheless, we can observe some variation amongst speakers. There is significant displacement of the tongue, especially in the anterior region, accompanied with an (apparently) passive ascending displacement of the mandible and lower lip. The pellet movement and tongue configuration were similar for all speakers.

The measures in Table 1 show that, for all speakers, the place of articulation was located around pellet T1. The distance measures between the lips did not seem to be relevant. Articulatory analysis pointed out similarities between individuals of the same gender, with more anterior places of articulation in women (jw54 and jw13) than in men (jw15 and jw61).

#### 4.9. Acoustic and Articulatory Analysis of [k]

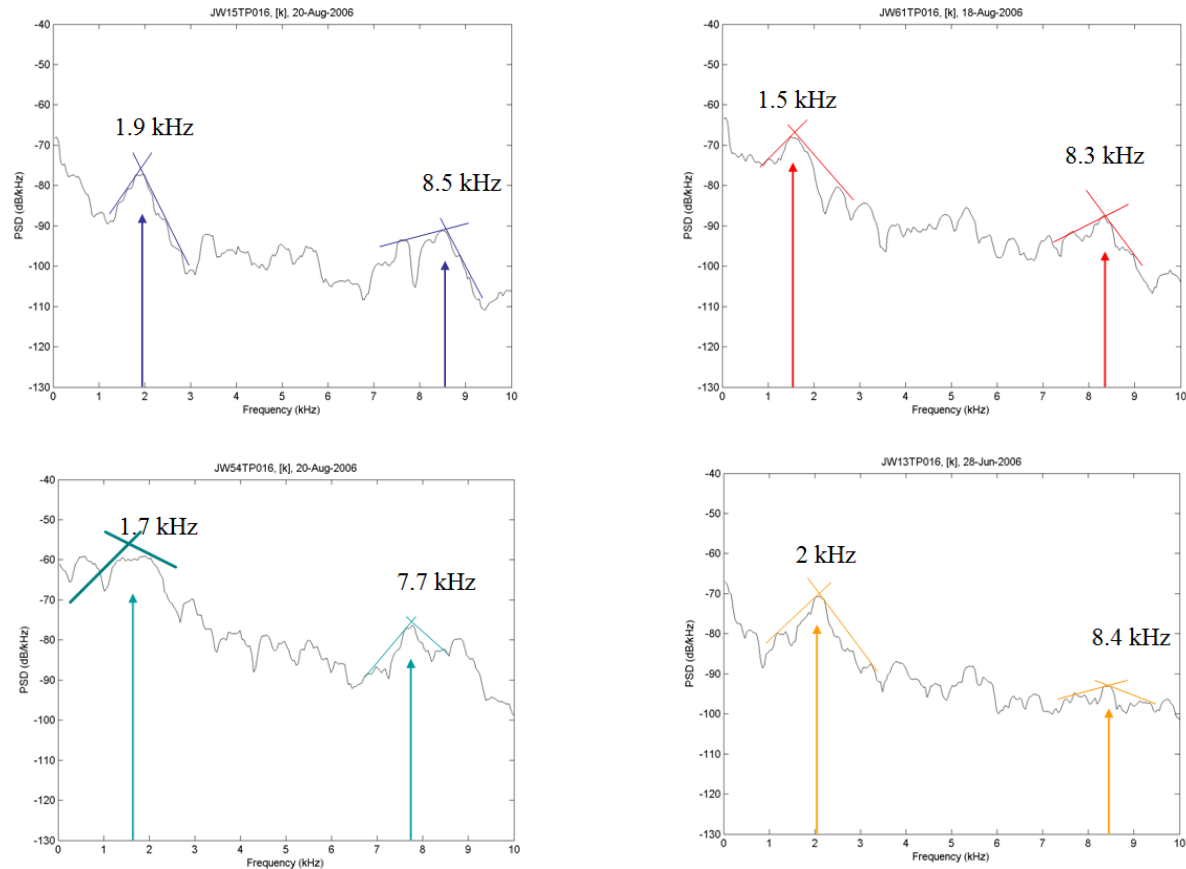
There was far more spectral variation for [k] than for [p] and [t], but we always clearly noticed two peaks at 1.5-2.8 kHz and 7-8.5 kHz. The medians of each peak were: jw15 – 2 and 8.3 kHz; jw61 – 2.2 and 7.4 kHz; jw54 – 2.2 and 7.8 kHz; jw13 – 2 and 8.1 kHz. The spectra of the four speakers (citation VCV task) are shown in Figure 12.

The articulatory representation in Figure 13 shows a movement of the tongue dorsum towards the hard and/or soft palate. The most constricted place was located between pellets T3 and T4, which corresponds to the transition between the hard and soft palate. Since the soft palate is not represented in Figure 13, it is difficult to analyse this part of the vocal tract. The trajectory of all tongue pellets seems to be directly related to the articulatory configuration of the vowels before (open-mid back [V]<sup>8</sup>) and after (open front [a]) the consonant. The tongue trajectory would probably be different in other vocalic contexts.

8 Vowel in 'but'.

**FIGURE 12**

Multitaper spectra of [k] produced in citation VCV task by speakers jw15 (upper left), jw61 (upper right), jw54 (lower left), and jw13 (lower right)



There seemed to be some similarity between speakers jw15 and jw54 in terms of the curvature of the tongue (see Figure 13). Class I speakers (jw15 and jw54) had a dome-shaped back tongue, and class II speakers (jw61 and jw13), a much flatter back tongue surface.

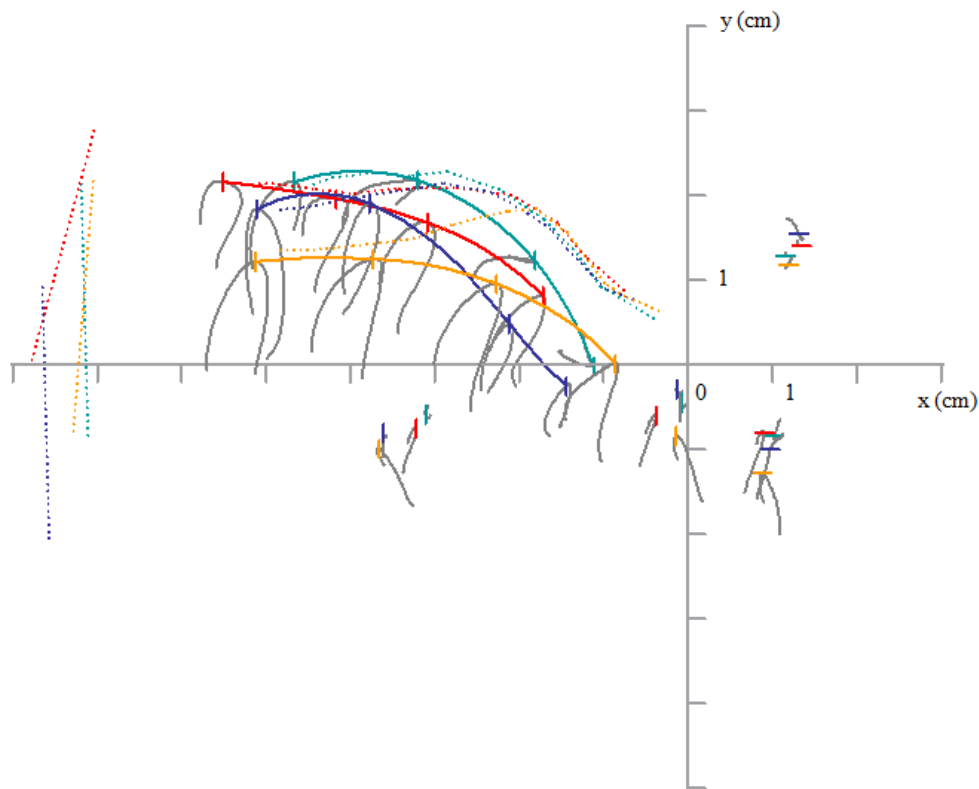
Table 1 measures confirm that pellet T3 was the closest to the palate. The distance to the palate progressively increased from the back to the front of the tongue in all speakers. PtX varied between 32 and 42 mm, with larger average values for females than for males. This information is rather insufficient, because the XRMB-SPD tracings do not include the soft palate, and so it was not always possible to extract

measures around pellet T4.

There seems to be some variation amongst subjects, related to the second peak, which might be potentially connected to the distance between lips. In fact, measure LL-UL and this peak's frequencies seem to be related: jw54 – LL-UL = 21 mm and Peak 2 = 7.7 kHz; jw61 – LL-UL = 23 mm and Peak 2 = 8.3 kHz; jw13 – LL-UL = 25 mm and Peak 2 = 8.4 kHz; jw15 – LL-UL = 26 mm and Peak 2 = 8.5 kHz. It is important to note that above 5 kHz non-planar modes begin to propagate, the radiation impedance decreases, and losses due to radiation thus decrease (Jesus & Shadle, 2002: 445), when interpreting these results.

**FIGURE 13**

Overlapping articulatory configurations and trajectories (grey) of [k] production by speakers jw15 (blue), jw61 (red), jw54 (green), and jw13 (orange)



## 5. Discussion

Acoustic analysis has shown that Class I and II female speakers use a smaller vowel space than the one generated by the reference values in Peterson & Barney (1952). The four vowels studied give us the four corners of a space showing the relative auditory qualities of vowels. Therefore, speaker JW13 had a wider perceptual separation of vowels than speaker JW54.

The MPW of female subjects were in a more anterior location than male, which may explain the higher second formant frequency values due to a reduction of the vocal tract length. The differences found in  $f_1$  frequencies between female subjects may be related to the dimensions of the posterior region of the vocal tract. In [i,  $\xi$ , u] productions the back tongue pellet was located more anteriorly for JW13, suggesting that

the pharyngeal cavity may be larger, producing lower  $f_1$  values. In [A] production this pellet is in a more posterior region for JW13, resulting in a higher  $f_1$ .

Articulatory analysis has also shown some variations between JW15 and JW54's productions, mostly related to tongue height in lower vowels [A] and [ $\xi$ ]. However, the general postures presented great similarities pointing to a representative mode of articulating of vowels in Class I subjects. On the other hand, Class II subjects showed very particular forms of articulating vowels. JW61 seems to be a typical Class II division 1 subject (Daskalogiannakis, 2000), presenting in all productions a retraction of the mandible and lower lip relative to the upper structures and to Class I subjects. His tongue was usually in a more posterior position and the back tongue

was often more elevated than for other subjects. This may be related also to his tipped MPW that requires an adapted posture to compensate the reduction in the AOS with the creation of extra space in the posterior region. These sorts of compensations have been described previously as typical of class II sagittal facial type (Daskalogiannakis, 2000).

JW13 tends to use a more open posture of the mouth and to move the mandible forward. This may be explained by the presence of a deep bite or by a division 2 of Class II malocclusion (Daskalogiannakis, 2000). Another common functional adaptation in JW13 is related to the tongue proximity to the palate. Since his palate has a different form, there is the need to adjust the tongue posture during vowel production, in order to create a resonance tube capable of producing appropriate formants for each vowel. This explains the anterior elevation of the tongue in [i] production and the lower tongue position in the other vowels' production.

The differences found in Class II subject's articulation show the great adaptation ability of the human vocal tract to adjust functional skills involved in speech production to structural variations. A better functional description could be achieved if XRMB-SPD data were complemented with cephalometric analysis in order to measure cranial structures involved in speech production. These adaptations could be related to the muscular groups involved in the production of some vowels' articulation, which could be quite different from those usually described for "normal" speech. As an example, in [i] production, speaker JW13 seems to use the superior longitudinal tongue muscle to elevate the tongue tip, which is not usually activated in "normal" productions. The study of these variations would be of particular relevance to support the clinical practice of speech and language therapists dealing with articulatory perturbations.

Obstruents [f] and [p] acoustic and articulatory results did not show any significant dif-

ferences between speakers. However, we have shown that some particular sounds' acoustic and articulatory measures are related: the frequency of the second peak of [s] and the place of articulation; the frequency of the broad peak in [S] spectra and the place of maximum constriction; the second peak frequency in [k] spectra and the distance between lips.

As for the studied variables (gender and dental occlusion), we noticed that: in [s] there were different adaptations by class II speakers; [S]'s place of articulation was more posterior in class II than in class I speakers; the articulatory configuration for [t] was related to gender; in [k] production there were also differences related to gender, concerning the curvature of the back tongue.

## 6. Conclusions

The Modified A-Space provided additional information, allowing the characterisation of cranio-facial features and the selection of a uniform set of speakers in studies (Jesus et al., 2007) involving XRMB-SPD. This method combines anatomical data and biomedical signals producing a reference dataset for research into speech production. We believe that this method may provide additional information to regular cephalometric analysis.

Vowel production did not seem to present acoustic differences related to occlusal class. However, the Class II female speaker had lower [i, ʌ, u] first formant frequencies than the Class I female subject.

There was great variability in terms of the articulatory processes used by the four subjects in this study, but mostly in Class II malocclusion subjects. Class II subjects used different articulatory postures to functionally adapt speech to their structural configuration (occlusal class and palate). The type of adaptations found should be described using cephalometric data contributing to a better understanding of normal and pathological speech production.

This study has shown that some of the known speech production variability is related with orofacial structures' variance. Different structures produce various functional adaptations, and distinct speech signals. When studying speech production, it is important to have in mind these characteristics to properly describe the speakers' sample and produce valid normative data.

## 7. Bibliographic references

ANGLE, E., 1907: *Treatment of malocclusion of the teeth* (7th ed.), Philadelphia: SSW.

ARAÚJO, A., 2007: *A influência de diferentes tipos de oclusão dentária na produção de sons da fala [The influence of occlusal class in speech production]*. M. Sc. Thesis, University of Aveiro, Aveiro, Portugal.

ARAÚJO, A., L. JESUS, & I. COSTA, 2007: "Vowel Production in Two Occlusal Classes" in *Proceedings of InterSpeech 2007*, Antwerp, Belgium, 1418-1421.

BLACKLOCK, O., 2004: *Characteristics of variation in production of normal and disordered fricatives, using reduced-variance spectral methods*. Ph.D. Thesis, University of Southampton, Southampton, UK.

CRYSTAL, D., 2008: *A Dictionary of Linguistics and Phonetics* (6th ed.), Oxford: Blackwell.

DASKALOGIANNAKIS, J., 2000: *Glossary of orthodontic terms*, Berlin: Quintessence Publishing.

HONDA, K., S. MAEDA, M. HASHI, J. DEMBOWSKI, & J. WESTBURY, 1996: "Human palate and related structures: their articulatory consequences" in *Proceedings of the 4th International Congress on Spoken Language Processing (ICSLP 96)*, volume 2, Philadelphia, USA, 784-787.

JESUS, L., A. ARAÚJO, & I. COSTA, 2007: "Articulatory Oral Space Measures Using the Modified A-Space" in

*Proceedings of the 5th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA 2007)*, Florence, Italy, 167-170.

JESUS, L. & C. SHADLE, 2002: "A parametric study of the spectral characteristics of European Portuguese fricatives", *Journal of Phonetics* 30(3), 437-464.

LEE, A., T. WHITEHILL, V. CIOCCA, & N. SAMMAN, 2002: "Acoustic and perceptual analysis of the sibilant sound /s/ before and after orthognathic surgery", *International Journal of Oral and Maxillofacial Surgery* 60(4), 364-372.

LOUSADA, M., L. JESUS, & D. PAPE, 2012: "Estimation of stops' spectral place cues using multitaper techniques", *DELTA* 28(1), 1-26.

MILENKOVIC, P., 2001: *TF32 user's manual*, Madison: University of Wisconsin.

NARAYANAN, S., 1995: *Fricative consonants: an articulatory, acoustic and systems study*. PhD. Thesis, University of California at Los Angeles (UCLA), Los Angeles, USA.

PETERSON, G. & H. BARNEY, 1952: "Control Methods in a Study of Vowels", *Journal of the Acoustical Society of America* 24(2), 175-184.

RAPHAEL, L., G. BORDEN, & K. HARRIS, 2007: *Speech Science Primer: Physiology, Acoustics, and Perception of Speech* (5th ed.), Baltimore: Lippincott Williams & Wilkins.

SHADLE, C. & C. SCULLY, 1995: "An articulatory-acoustic-aerodynamic analysis of [s] in VCV sequences", *Journal of Phonetics* 23(1-2), 53-66.

STEVENS, K., 1998: *Acoustic Phonetics*, Cambridge: MIT Press.

VALLINO, L. & B. TOMPSON, 1993: "Perceptual characteristics of consonant errors associated with

malocclusion”, *Journal of Oral and Maxillofacial Surgery* 51(8), 850-856.

WELLS, J., 1997: “SAMPA computer readable phonetic alphabet” in D. GIBBON, R. MOORE, & R. WINSKI (eds.): *Handbook of Standards and Resources for Spoken Language Systems*, Berlin: Mouton de Gruyter, 684-732.

WESTBURY, J., 1994: *X-ray microbeam speech production database user’s handbook. Version 1.0*, Madison: University of Wisconsin.

ZWICKER, E. & E. TERHARDT, 1980: “Analytical expressions for critical-band rate and critical bandwidth as a function of frequency”, *Journal of the Acoustical Society of America* 68(5), 1523-1525.

ZYGIS, M., D. PAPE, & L. JESUS, 2012: “(Non)retroflex Slavic Affricates and Their Motivation: Evidence from Czech and Polish”, *Journal of the International Phonetic Association* 42(3), 281-329.