INFLUENCE OF THE VOCAL TRACT MORPHOLOGY ON THE F1-F2 ACOUSTIC PLANE

Antoine Serrurier[1], Pierre Badin[2], Christiane Neuschaefer-Rube[1]

[1] Clinic for Phoniatrics, Pedaudiology & Communication Disorders, University Hospital and Medical Faculty of the RWTH Aachen University, Germany

[2] Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France
aserrurier@ukaachen.de

Abstract: In order to be understood, speakers must produce sounds that are recognised by listeners. Large inter-speaker morphology variability is however reported in the literature. This study explores the influence of seven morphological features on the F1-F2 formants: four related to global variations of the vocal tract and three to the palate shape. For this purpose, the contours of vocal tract articulators were manually segmented from midsagittal MRI images available for 11 French speakers sustaining 62 articulations. The morphology features were measured for each speaker on the mean articulation of the 62 articulations and their articulatory effect on the ten oral vowels derived by means of linear regressions. The set of articulations obtained for each vowel and morphology feature by varying at regular steps the values of the morphology features were used as input of an acoustic plane wave propagation model. Corresponding formants F1 and F2 were extracted and analysed. The results, in general agreement with the literature, confirm the phoneme-specific influence of the morphology features, and emphasise the articulatory impact on the whole vocal tract when the morphology of one region only varies and their acoustic consequences.

1 Introduction

In order to be understood in a language, speakers must share invariant acoustic codes. The oral vowels of a language are in particular known to be characterised by their two first formant frequencies [1]. At the same time, large inter-speaker variability has been reported in the literature from both articulatory [2,3] and acoustic point of views [4]. Speakers are indeed characterised by a specific morphology, *i.e.* the intrinsic size and shape of the speech articulators irrespective of the articulatory tasks. The morphological differences between males and females are well-documented (*e.g.* [5]): males have a longer vocal tract than females and the ratio of the lengths of the vertical part over the horizontal part of the vocal tract (initially defined and referred to as Larynx Height Index – LHI – by Honda *et al.* [6]) is also higher for males than for females. In addition to the morphological differences, speakers differ in their articulatory strategy, *i.e.* in the displacement and deformation of the speech articulators to perform speech tasks. This can be the result of an adaptation to the specific morphology in order to reach the invariant articulatory-acoustic goals or of pure idiosyncratic strategies [7].

Numerous studies that have analysed the acoustic impact of the variation of size of the vocal tract, the major male-female morphology difference, have naturally found that vocal tract lengthening decreases the resonance frequencies [1,8], although the scaling may not be uniform across phonemes [9]. It has also been shown that the proportion of the length of the vertical and horizontal parts of the vocal tract on the one hand and the palate shape on the other hand have influence on the articulations [2,3,10]. For instance, Lammert *et al.* [11] have characterised the various modes of variation of the palate shape across speakers and simulated their acoustic consequences. The previous studies usually consider limited portions of the vocal tract to perform the articulatory/acoustic analyses. Further, the acoustic outcomes are usually the result of modifications leading to area functions, *i.e.* the cross-sectional areas along the vocal tract midline, that cannot be related to specific speakers. Our objective in this study is rather to evaluate the acoustic impact of the variation of various morphological features based on articulations of speakers. In this data-based approach, a morphology feature may also encompass effects related to the speakers' strategy, but that speakers consistently implement. In addition, we intend to consider the entire vocal tract and all the surrounding articulators to ensure taking into account all possible consequences of local changes in the morphology on other regions of the vocal tract.

The article is organised as follows: the data and their processing are presented in Section 2, the various morphology features considered in the study and their acoustic influence are presented in Section 3, and a discussion and a conclusion are proposed in Section 4.

2 Data, corpus and material

2.1 Data and corpus

The data consist in static midsagittal MRI images of the vocal tract collected by Valdés Vargas [12] for 11 French speakers (6 males, 5 females) sustaining artificially 62 articulations, considered as balanced and representative of the French articulatory repertoire: 10 oral and 2 nasal vowels [i e ɛ a y ø œ u o ɔ ɑ̃ ɔ̃], and each of the 10 consonants [p t k f s ʃ m n ʁ l] in the 5 symmetric vowel contexts [i e ɛ a u] (*cf.* Serrurier *et al.* [13] for further details).

2.2 Articulation contours

For each speaker, the contours of the bony structures (hard palate, jaw, including the upper and lower teeth, hyoid bone and C5 cervical vertebrae) have been manually segmented on one specifically recorded articulation where the lips and tongue are in contact with the teeth, and rigidly aligned individually on each of the 62 other images. The contours of the deformable articulators surrounding the vocal tract have then been manually segmented on all the images: upper and lower lips (extended on the face until respectively the nose and the neck), tongue, velum, pharyngeal wall, epiglottis and posterior supraglottis. The anatomical landmarks Anterior and Posterior Nasal Spine (ANS and PNS) have been manually identified on all images. PNS delimits also the posterior end of the hard palate and the anterior start of the velum. The 62 sets of contours have then been rigidly aligned to ensure an identical palate contour for all articulations. As ANS and PNS are anatomically rigidly attached to the palate, these landmarks are supposed to exactly overlap after this alignment. Two small clouds of 62 points have however been observed due to the inherent errors in the manual identifications: ANS and PNS have thus been approximated by their means to smooth out the manual errors. The articulations have finally been rigidly aligned across speakers so that the lower edge of the upper incisor is arbitrarily set to the same location (X=5 cm, Y=10 cm) and the ANS-PNS line is horizontal. Further details can be found in Serrurier *et al.* [13].

2.3 Speaker normalisation

As emphasised in the introduction, one of the major morphological differences between speakers is vocal tract length. This large difference may outweigh and mask more subtle differences. One way to deal with this issue is to normalise the articulations so that the mean vocal tracts of all speakers have the same length [3]. As the horizontal and vertical parts of the vocal tract can differ significantly between speakers, they are normalised separately. A *factor_x* is calculated for the mean articulation of each speaker as the ratio of the horizontal distance between ANS and the intersection of the ANS-PNS line with the best fitting line of the pharyngeal wall to a reference length, defined as the mean of this distance on the 11 speakers. A *factor_y* is similarly calculated as the ratio of the vertical distance between the upper anterior corner of C5 and the ANS-PNS line to the mean of this distance on the 11 speakers. Finally, all articulations of each speaker are rescaled along both horizontal and vertical dimensions using the speaker's *factor_x* and *factor_y* parameters, while keeping the lower edge of the upper incisor as the same location. Depending on the morphology feature studied, the analyses performed in this study are carried out either on the contours before normalisation, or on these *normalised contours*. Note that this normalisation ensures that the lengths of the vertical and horizontal parts of the vocal tract are similar across speakers' mean articulation but does not fully handle the size of the articulators themselves.

3 Influence of morphology features

3.1 Morphology features

Both morphology and strategy features are involved when a speaker attempts to produce specific articulations. As stated by Serrurier *et al.* [13], one way to disentangle them and to obtain a contour free from the phoneme-specific strategies is to calculate the speaker's mean articulation over the set of the 62 articulations. The large and well-balanced corpus may ensure that the phoneme-specific strategies are compensated and that the resulting mean contour may be considered as representative of the morphology of the speaker. The speaker's morphology features are estimated from their mean articulation.

Seven different morphology features have been considered in this study. As done by Serrurier *et al.* [13], a Principal Component Analysis (PCA) has first been applied to the set of 11 mean articulations in order to determine the principal morphological modes of deformation of the speakers' articulations. The first two PCA components **MP1** and **MP2** have been retained, explaining respectively 64% and 24% of the variance, altogether 88%, and leading to a cumulative Root Mean Square (RMS) reconstruction error of 0.2 cm. The corresponding nomograms, *i.e.* the superposition of the contours predicted for the control parameter of MP1 (*resp.* MP2) varying at regular steps between the minimal and maximal values found in the data, are visible in Figure 1. MP1 is mainly related to the vocal tract length, both in the horizontal and vertical directions, and at the same time to the depth of the palate, while MP2 is related to the horizontal length of the vocal tract together with a rotation of the contours much related to the position of the head. These two parameters are related to the male/female differences and can be used to discriminate female and male speakers (*cf.* [13]). They constitute the first two morphology features considered in this study.

The next morphology feature, the vocal tract length, thereafter **VTL**, is a well-documented characteristic of a speaker [5] and deserves thus to be considered. VTL has been calculated for each speaker on the mean articulation as the length of the midline of the vocal tract. It appears highly correlated with MP1 (0.93). Since its effects are very similar to those of MP1, they have not been described in the rest of the article.

The vertical over horizontal length ratio (**LHI**) is a feature related to male-female differences, males presenting higher LHI than females – due to a lower larynx position for males -, but complementary to the formerly defined features MP1, MP2 and VTL. It has been calculated in this study on the mean articulation of each speaker according to the definition formulated by Honda *et al.* [6] as the ratio of LH over PD, where LH is the distance between the lower point of the pharyngeal wall and the ANS-PNS line and PD is the distance between ANS and the intersection of the ANS-PNS line with the pharyngeal wall. Note that LHI is not much correlated with MP1, MP2 nor with VTL (maximum correlation coefficient of 0.5). In order to visualise the variations of the mean articulation associated with LHI, a Linear Regression (LR) of the set of the 11 mean articulations on the z-scored vector of the 11 LHI has been applied. Figure 2 (right) displays the resulting nomogram. As expected, it is related to a variation of larynx height, and associated with a variation of the height of the tongue and of the epiglottis. It has however no influence on the palate, inducing a variation of the tongue-palate distance. LHI constitutes the fourth morphology feature retained in this study.

As emphasised in the introduction, the shape of the palate has a direct influence on the tongue articulation. The palate shape has been characterised in detail by Lammert *et al.* [11]. Their analyses were focused on the pure palate shape regardless of its size and orientation. On the opposite, the present study intends to assess the influence of the palate shape, including its size and orientation in relation to the ANS-PNS line, on the vocal tract acoustics. As the MP1 nomogram of Figure 1 shows, there is a correlation between the palate shape and the vocal tract length that is unexpected and might be ascribed to the limited number of speakers in the data set. To limit this effect, all the analyses related to the palate shapes are performed on the normalised contours (see Section 2.3). A PCA has been applied to the set of palates from the 11 mean articulations to extract the principal modes of deformation. The first three components have been retained, explaining respectively 84%, 10% and 6% of the variance, for a cumulated RMS reconstruction of 0.03 cm. Note that in comparison to the previous PCA applied to the global mean articulations, the third component was also retained here because it appears related to one of the modes of variation described by Lammert *et al.* [11]. In order to visualise the variations of the other articulators associated with these palate components on the mean articulation, LRs of the contours of these articulators (tongue, velum and pharyngeal wall) on each of the control parameters of these three components have been applied. The other articulators have been excluded to reduce as much as possible the corpus effects, *i.e.* the correlations likely ascribable to the limited set of speakers rather than to explicit strategy or morphology effects. Figure 2 displays the nomograms for the three resulting components **Pal1**, **Pal2** and **Pal3**. The first component Pal1, explaining alone 84% of the variance of the palate, is related to a flattening/doming of the palate, that also corresponds to the first mode of variation described by Lammert *et al.* [11] as the concavity mode. It is associated with a mirroring movement of the tongue that maintains the tongue-palate distance approximately constant; it is associated in addition with a slight shortening/lengthening of the horizontal size of the vocal tract, speakers with flatter palates having longer horizontal vocal tracts. The component Pal2 is related to a kind of rotation around the upper incisor with a slight concomitant increase of size together with the tongue, without changes in their shapes. The component Pal3 may be related to the second mode described by Lammert *et al.* [11] as the anteriority mode, *i.e.* "*whether the apex of the dome is positioned toward the anterior or posterior portion of the oral cavity*". Note that it is associated for the present data with a substantial rotation of the articulators of the back of the vocal tract, presumably related to the orientation of head, like MP2. Pal1 and Pal3 are in general agreement with the first two modes described by Lammert *et al.* [11], respectively the concavity and anteriority modes, although with different percentages of variance explanation, while Pal2 is related to the orientation of the palate in reference to the ANS-PNS line, whose variability was discarded by design in the study of Lammert *et al.* [11] . Pal1, Pal2 and Pal3 constitute the last three morphology features considered in the study.

3.2 Acoustic simulations

In order to calculate independently the influence of each of the 6 retained morphology features presented above on the 10 oral vowels, independent LRs of each of these articulations for the 11 speakers on each morphology feature have been applied. Note that the normalised contours have been considered for the LR on the palate morphology features. This resulted in a collection of 6×10 articulatory components. For each of these components, articulatory nomograms have been calculated, resulting in a collection of 6×10×10 sets of midsagittal contours. Formants up to 5 kHz can be computed from the vocal tract midsagittal contours assuming plane wave propagation [1]. For this purpose, the midline of the vocal tract for each nomogram has been calculated as well as the sagittal function, *i.e.* the transverse sagittal distance between the upper and lower vocal tract contours along the midline The area function has finally been obtained using the model proposed by Soquet *et al.* [14], with a minimum area set to 0.05 cm² to avoid occlusions. Plane wane propagation in the tubes has then been simulated using an electric equivalent model [1] and the acoustic transfer function calculated. The two first formants F1 and F2 for each nomogram have finally been extracted.

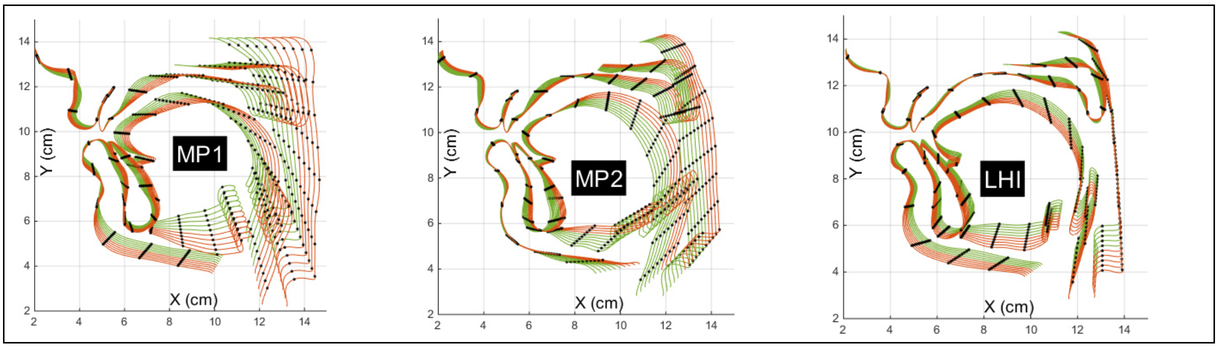

Figure 1 – Nomograms of the contours of the mean articulation for MP1 (left), MP2 (middle) and LHI (right) varying at regular steps between the minimal and maximal values found in the data. Contours with negative (*resp.* positive) predictor values are plotted in green (*resp.* orange). One every 25 point is plotted as black dot to emphasize deformation directions.

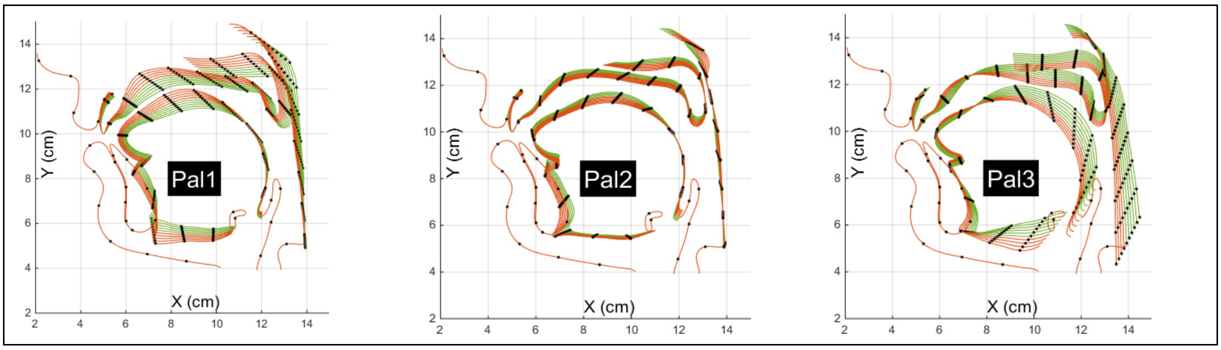

Figure 2 – Same as Figure 1 but for Pal1 (left), Pal2 (middle) and Pal3 (right).

3.3 Results

For each of the three morphology features MP1, MP2, LHI, the nomograms of the 10 oral vowels in the F1-F2 acoustic plane are visible in Figure 3. The nomograms for the morphology features Pal1, Pal2 and Pal3 present smaller ranges and less linear variations, making such display hardly readable. Instead Figure 4 displays the relative variations of F1 and F2 for the three corner vowels /a i u/ as functions of Pal1, Pal2 and Pal3, as already proposed by Lammert *et al.* [11]. Following their proposal, the relative variation δF1 of F1 (*resp.* δF2 and F2) is calculated as the difference between F1 (*resp.* F2) and the value of F1 obtained for Pal1/Pal2/Pal3 equal to 0 (*resp.* F2), expressed in percentage of this latter value. The corresponding articulatory nomograms could unfortunately not be displayed for space reason but nevertheless considered to help the interpretation.

For the three morphology features MP1, MP2 and LHI we observe a decrease of F1 and F2 for all the vowels as the size of the vocal tract increases. This is a well-documented phenomenon ascribed to the fact that a larger and longer vocal tract has longer cavities, associated with lower resonance frequencies [8,9]. This characterises typically the difference between males and females. The decrease of F1 and F2 as the size of the vocal tract increases comes also generally together with a decrease of the area of the triangle defined by the three extreme vowels [a i u] in the F1-F2 plane, related to the maximum vowel space of a speaker [8]. This is verified in our case for MP1 and MP2 for which we observe a linear decrease of the area of this triangle as the value of the parameters changes toward a longer vocal tract: the correlation coefficients between the triangle area and the parameter values are 0.96 for MP1 and 0.99 for MP2. Interestingly the highest correlation is observed for MP2, corresponding mainly to a shortening/lengthening of the horizontal part of the vocal tract, while the lowest correlation is observed for LHI (0.62), corresponding mainly to a shortening/lengthening of the vertical part of the vocal tract. This suggests that the area of the triangle may be primarily related to the size of the horizontal part of the vocal tract. The nomogram of MP1 shows a decrease by the same order of magnitude of F1 and F2 for all vowels, ranging from 16% to 33% depending on formant and vowel. This is in general agreement with the already documented effect of the vocal tract length on the formants F1-F2 of the vowels, derived for instance from vocal tract simulated changes based on articulatory models (*cf.* [8]). MP2 shows a smaller range of variation for F1 and F2, from 0% to 17% depending on formant and vowel. The front vowels ranging from [i] to [a] (*i.e.* [i e ɛ a]) show more linear variations and higher ranges of variation of F2 than other vowels; [u] does not show any variation of F2 for instance. The variations of the formants for LHI show higher disparities, from 2% for the F1 of [y] to 27% for the F2 of [u]. Again, the front vowels tend to show higher ranges of variation than the other vowels. In other words, middle and back vowels (*i.e.* [y ø œ u o ɔ]) tend to be more robust than front and low vowels to variations controlled by MP2 and LHI, *i.e.* variations mainly related to the size of either the horizontal or the vertical part of the vocal tract.

The variations of F1 and F2 as a function of the palate morphology features Pal1, Pal2 and Pal3 are in general smaller, usually lower than 10%, as visible on Figure 4. While the formants for [a] do not appear affected by the morphological variations related to Pal1, F1 increases on the contrary for both vowels [i] and [u] with Pal1, *i.e.* from a flat to a domed palate. This can be ascribed in both cases to the decrease of the volume of the back cavity, acting in both case as a Helmholtz resonator [1]. The lengthening of the back cavity as Pal1 increases for [i] might also be the reason for the decrease of F2. Regarding Pal2, higher variations are observed for F2 than for F1. The slight increase of F1 for [u] as Pal2 increases, *i.e.* as the palate and the tongue rotate around the upper incisor, might be explained by the associated slight decrease of the volume of the back cavity, acting as a Helmholtz resonator. For F2, a detailed analysis revealed that an increase of Pal2 was interestingly associated with a lengthening of the vocal tract for [a] and [u] but a shortening for [i], probably responsible for the decrease observed for [a] and [u] on the one hand and the increase observed for [i] on the other hand. For [u], the observed decrease of F2 may additionally be accentuated by the slight increase of the volume of the front cavity, acting as a Helmholtz resonator, as Pal2 increases. For Pal3, we can observe a decrease of F1 and F2 for the vowels [a] and [i] than can be related to the lengthening of the vocal tract as Pal3 increases, *i.e.* as the back articulators rotate clockwise. In addition for [i], the tongue-palate distance decreases, leading to a smaller neck of the Helmholtz resonator back cavity, possibly reinforcing the decrease of F1. For [u], the tongue-palate distance for the section of the vocal tract between the front and back cavities decreases for high values of Pal3; this section acts as the neck for the front and back Helmholtz resonators and probably participates in the decrease observed for F1 and F2. The increase of F2 for lower values of Pal3 may be ascribed to the decrease of the volume of the front Helmholtz resonator cavity before the decrease of the tongue-palate distance mentioned above counterbalances this effect. The results obtained for the vowel [i] for the morphology features Pal1 and Pal3 can be compared to the simulations performed by Lammert *et al.* [11] for the concavity and anteriority modes on a high-front vowel. The same trend for the variations of F1 and F2 in relation to the palate shape is observed for Pal1/concavity mode, but with a lower range of amplitude in our case. The variations of F1 and F2 for Pal3 differ however from the variations obtained by Lammert *et al.* [11] for the anteriority mode. This could be partly ascribed to the large rotation of the back articulators observed in the present study.

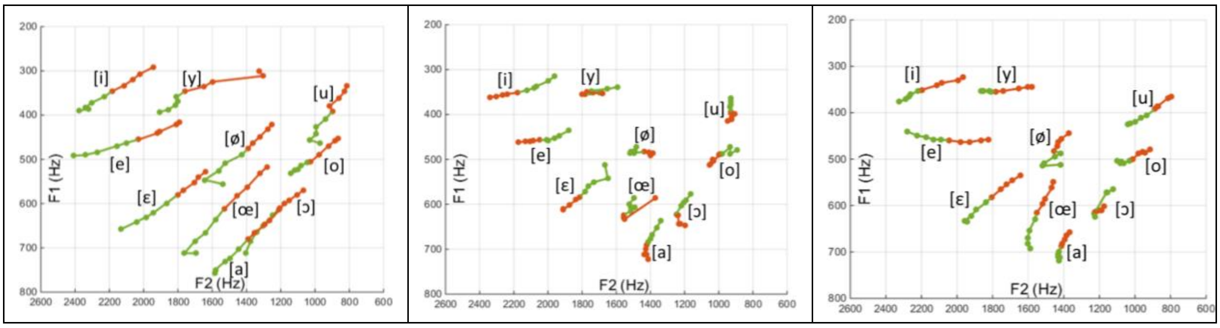

Figure 3 – Nomograms of the 10 oral vowels in the F1-F2 plane for MP1 (left), MP2 (middle) and LHI (right) varying at regular steps between the minimal and maximal values found in the data. F1-F2 values corresponding to negative (*resp*. positive) parameter values are plotted in green (*resp*. orange).

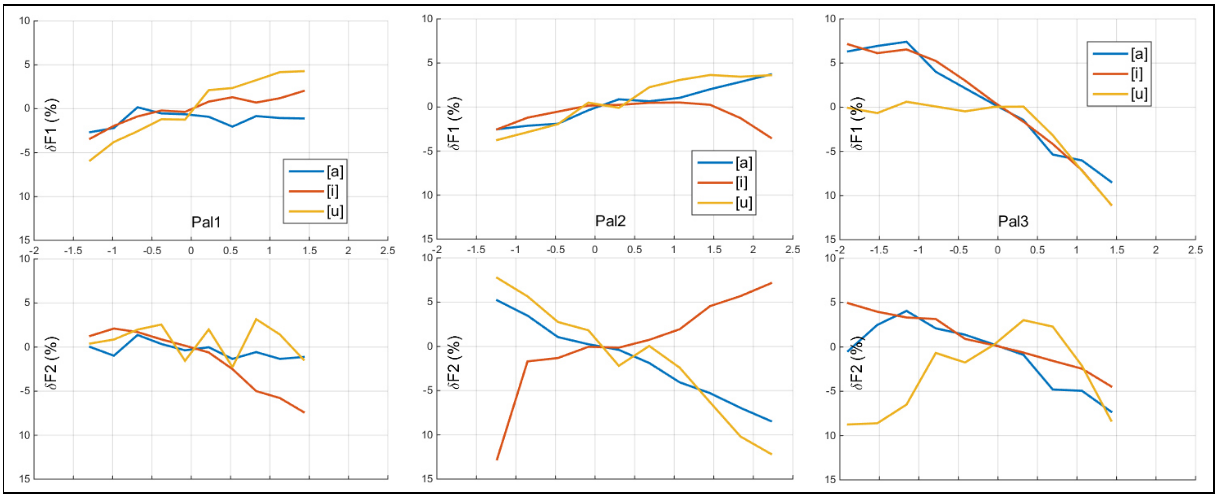

Figure 4 - δF1 (top) and δF2 (bottom) as a function of Pal1 (left), Pal2 (middle) and Pal3 (right) for the three vowels [a i u]

4 Discussion and conclusion

In this study, seven morphology features were measured for 11 French speakers on their mean articulation estimated over 62 recorded articulations representative of the French articulatory repertoire: four capturing global variations of the vocal tract (*e.g.* its size), and three related to the palate shape. For each of the 10 oral vowels, a LR of the 11 articulations on each of the 6 retained morphology features (out of 7) was applied to capture the articulatory variations related to the morphology features. Articulatory nomograms for the resulting 10×6 components were calculated, acoustic wave propagation was simulated and the formants F1-F2 extracted from the acoustic transfer functions. The nomograms of the formants F1-F2 for the 10 vowels for each of the morphology presented were then presented and analysed. This data-based approach confirms in general the existing results found in the literature. It emphasizes in addition that a given morphology feature can have very different articulatory and acoustic consequences for different phonemes, in agreement with Fant [9], either because of the different articulatory goals of the different phonemes, or because of the different articulatory strategy a speaker may use for different phonemes. Lastly, it was observed that local morphology variations, for instance the palate concavity, influence the whole vocal tract, for instance by lengthening or shortening the vocal tract. This confirms *a posteriori* that all the articulators and the entire vocal tract should be considered while performing acoustic analyses of morphology features. The main limitation of the study lies in the limited number of speakers. Some correlations between different articulators observed on the mean articulations might indeed be ascribed to the specific set of speakers considered in this study. Despite our attempt to limit these effects, for instance by carrying out analyses on normalised contours, the statistical analyses and results presented in this study should be considered with caution. Further analyses on a larger cohort of speakers should be performed in the future to validate these preliminary results.

Despite this limitation, this study presents a promising data-based methodology to analyse the influence of speakers' morphology to the acoustic domain. As the articulations produced by a speaker result from her/his morphology but also from her/his idiosyncratic strategy, a similar approach could be used to characterise acoustically the strategies of the speakers.

5 Acknowledgments

The authors are very grateful to L. Lamalle for the recording of the MRI, and to J.-A. Valdés Vargas and G. Ananthakrishnan for performing the majority of the initial tracings. This work has been partially funded by the French ANR (grant ANR-08-EMER-001-02 "ARTIS").

6 References

[1] FANT, G.: Acoustic Theory of Speech Production. The Hague: Mouton, 1960.

[2] BRUNNER, J., FUCHS, S. and PERRIER, P.: On the relationship between palate shape and articulatory behavior. In J Acoust Soc Am, vol. 125(6), pp. 3936–3949, 2009.

[3] GENG, C. and MOOSHAMMER, C.: How to stretch and shrink vowel systems: Results from a vowel normalization procedure. In J Acoust Soc Am, vol. 125(5), pp. 3278–3288, 2009.

[4] PETERSON, G. E. and BARNEY, H. L.: Control Methods Used in a Study of the Vowels. In J Acoust Soc Am, vol. 24(2), pp. 175–184, 1952.

[5] VORPERIAN, H. K., WANG, S., CHUNG, M. K., SCHIMEK, E. M., DURTSCHI, R. B., KENT, R. D., ZIEGERT, A. J. and GENTRY, L. R.: Anatomic development of the oral and pharyngeal portions of the vocal tract: An imaging study. In J Acoust Soc Am, vol. 125(3), pp. 1666–1678, 2009.

[6] HONDA, K. and TIEDE, M. K.: An MRI Study On The Relationship Between Oral Cavity Shape And Larynx Position. In Proc. ICSLP, vol. 2, pp. 437–440, 1998.

[7] LADEFOGED, P. and BROADBENT, D. E.: Information Conveyed by Vowels. In J Acoust Soc Am, vol. 29(1), pp. 98–104, 1957.

[8] MÉNARD, L., SCHWARTZ, J.-L. and BOË, L.-J.: Role of Vocal Tract Morphology in Speech Development: Perceptual Targets and Sensorimotor Maps for Synthesized French Vowels From Birth to Adulthood. In J Speech Lang Hear R, vol. 47(5), pp. 1059–1080, 2004.

[9] FANT G.: A note on vocal tract size factors and non-uniform F-pattern scalings. In STL-QPSR, vol. 7(4), pp. 22–30, 1966.

[10] FUCHS, S., WINKLER, R. and PERRIER, P.: Do Speakers' Vocal Tract Geometries Shape their Articulatory Vowel Space? In Proc. ISSP, pp. 333–336, 2008.

[11] LAMMERT, A., PROCTOR, M. and NARAYANAN, S.: Interspeaker Variability in Hard Palate Morphology and Vowel Production. In J Speech Lang Hear R, vol. 56(6), pp. 1924–1933, 2013.

[12] VALDÉS VARGAS, J. A.: Adaptation of orofacial clones to the morphology and control strategies of target speakers for speech articulation. PhD Thesis, Université Grenoble Alpes, 2013.

[13] SERRURIER, A., BADIN, P., LAMALLE, L. and NEUSCHAEFER-RUBE, C.: Characterization of inter-speaker articulatory variability: a two-level multi-speaker modelling approach based on MRI data. In J Acoust Soc Am, in review.

[14] SOQUET, A., LECUIT, V., METENS, T. and DEMOLIN, D.: Mid-sagittal cut to area function transformations: Direct measurements of mid-sagittal distance and area with MRI. In Speech Commun, vol. 36(3–4), pp. 169–180, 2002.