F1 AND F2 FORMANT VARIATIONS AND INTER-SPEAKER ARTICULATORY VARIABILITY: A PRELIMINARY ANALYSIS

Antoine Serrurier, Christiane Neuschaefer-Rube

Clinic for Phoniatrics, Pedaudiology & Communication Disorders, University Hospital and Medical Faculty of the RWTH Aachen University, Germany
aserrurier@ukaachen.de

Abstract: Oral vowels are mainly acoustically characterised by their two lowest formants F1 and F2. It is well accepted that raising F1 is achieved by lowering the jaw and the tongue whereas lowering F2 is achieved by moving the tongue backward and protruding the lips. Conversely, it is well known that speakers differ in morphology and articulatory strategy. This study intends therefore to characterise the inter-speaker articulatory variability associated with the variation of F1 and F2. For this purpose, static midsagittal MRI from 21 speakers sustaining oral vowels were considered. The vocal tract contours were manually segmented and aligned per speaker and per articulation. The formants F1 and F2 were obtained by means of modelling. The F1 and F2 articulatory components for each speaker were obtained by a linear regression of the articulations on F1 and F2. Qualitative analyses showed a large inter-speaker variability of the F1 and F2 components. Averaged F1 and F2 components per sex revealed a similar strategy for males and females with subtle differences. Quantitative analyses of the F1 and F2 components' variability did not reveal clear organisation nor speaker cluster. Speakers tended to differ in the degree of use of the jaw and of the larynx and in the region of the tongue used for the constriction. These results suggest that the variations of the F1 and F2 components are not much driven by morphological constraints and may rather be purely idiosyncratic. Further analyses may include a larger number of speakers associated with deeper quantitative analyses.

1 Introduction

Vowels are acoustically characterised by their formants, among which the first (F1) and second (F2) formants play the major role. Since decades, the relationship between the F1 and F2 formants and the articulatory strategies of the vocal tract are well understood: raising F1 is achieved by lowering the jaw and the tongue whereas lowering F2 is achieved by moving the tongue backward and protruding the lips (*e.g.* [1]).

On the other side, it is also well-known that speakers differ in morphology and articulatory strategies [2]. The relationship between the acoustic and articulatory vowel spaces has for instance been largely analysed in the literature (*e.g.* [3,4]). More rarely were analysed the relationship between the vocal tract shape and the variations of F1 and F2 across speakers. Most studies focus on kinematic data and analyses of the tongue alone (*e.g.* [5]). Surprisingly, little is therefore known about the inter-speaker articulatory variability of the vocal tract related to the variations of F1 and F2. This study aims at filling this gap by providing a data-based preliminary analysis.

The objective of the study is to observe and characterise the inter-speaker articulatory variability related to the variation of F1 and F2. In other words, it aims at analysing the various articulatory strategies used by the speakers to modify their F1 and F2 formants. It aims in addition at considering the entire vocal tract, from the larynx to the lips. In general, a better understanding of the articulatory-acoustic relationships and their individualities may help in the future for the assessment and treatment of speech articulatory disorders.

This study relies on medical imaging articulatory data and a modelling approach. The section 2 presents the data and the methods, including qualitative and quantitative analyses, the section 3 presents the related preliminary results and the section 4 a short discussion and conclusion.

2 Material and methods

2.1 Data and corpora

The data consist in two datasets of static midsagittal Magnetic Resonance Images of the vocal tract recorded under auditive control of the experimenter. The first dataset consists of 11 French speakers (6 males, 5 females) sustaining artificially the 10 oral vowels [i e ɛ a y ø œ u o ɔ], and the second dataset of 10 German speakers (7 males, 3 females) sustaining artificially the 8 oral vowels [a: e: ɛ: i: o: u: y: ø:]. The contours of all articulators surrounding the vocal tract on each image have been manually segmented and the resulting articulations aligned per speaker on the contour of the hard palate and between speakers on corresponding landmarks taken on the palate bone. These data constitute a subset of the data presented in Serrurier *et al.* [6], for which more details regarding collection and processing procedure lie in Serrurier *et al.* [7]. The data consist in the end in two datasets of contour coordinates of size 11×10×1037×2 and 10×8×1037×2 corresponding to the 11 French and 10 German speakers, the 10 French and 8 German articulations, and the 1037 contour points of 2 x-y dimensions. They encompass altogether data of 13 male and 8 female speakers. The articulation contours will be sometimes referred to as articulations in this manuscript for simplicity reasons.

2.2 Method

2.2.1 Acoustic modelling

The study relies on a modelling approach to represent the articulatory components corresponding to the formant variations. As the considered data do not contain any acoustic recordings, the formants themselves have been obtained by modelling. The chosen approach relies on the simulation of the propagation of a plane wave in the vocal tract [8]. For this purpose, the following steps have been performed on each articulation of the datasets: (1) calculation of the sagittal function, *i.e.* the variation of the distance between the upper and lower vocal tract contours along the midline from the glottis to the lips, (2) calculation of the area function, *i.e*. the variation of the area of the vocal tract from the glottis to the lips, from the sagittal function using a α-β model [9], (3) calculation of the acoustic transfer function from the area function using an electrical equivalent modelling approach [10] and (4) calculation of the frequency of the two first formants F1 and F2 from the acoustic transfer function. Such acoustic simulations have already proven their accuracy in the past literature for low frequencies such as for F1 and F2 [11]. By this method, the F1 and F2 formants for each oral vowel and each speaker can be calculated and represented.

2.2.2 F1 and F2 components

The articulatory components corresponding to the variations of F1 and F2 for each speaker have been obtained by means of linear regression. For this purpose, a linear regression of the articulation contours of the oral vowels on the z-scored F1 and F2 values for each speaker has been performed. It resulted in a set of two components, referred to as *F1 component* and *F2* *component*, for each speaker, capturing the linear articulatory variations associated with the variations of F1 and F2.

2.2.3 Variability analysis

The objective of the study is to investigate the articulatory variability of the F1 and F2 components across speakers, *i.e.* the different existing articulatory strategies to lower or raise these two formants. Qualitative and quantitative analyses have been performed to characterise this variability.

First, the F1 and F2 components have been averaged over the male and female speakers, and nomograms, *i.e.* the articulatory contours corresponding to a regular variation of F1 and F2 between the minimal and maximal values found in the data, have been calculated. This allows the general description of the F1 and F2 components and the observation of possible differences between the male and female strategies.

Second, the F1 and F2 components for the 21 speakers have been qualitatively observed by means of nomograms. A few samples have been displayed in this article as illustration of the variability. As the F1 component is expected to correspond mainly to an opening-closing of the jaw, the articulatory component corresponding to the jaw opening and closing, thereafter *Jaw Height (JH) component*, has also been calculated for each speaker to observe potential differences with the F1 component. It aims at measuring the contribution of the jaw movement in the variations of F1 and whether the F1 variation is mainly achieved by this movement or by more complex articulatory strategies. The JH component has been obtained by linear regression of the oral vowel articulations on the y-dimension of the jaw contour for each speaker.

Third, quantitative analyses have been performed in an attempt to understand the variability of the F1 and F2 components across speakers. A Principal Component Analysis of the 21 sets of coefficients obtained by the formerly described linear regression has been performed for each F1 and F2 components to extract the principal components of each of these components. This represents a sort of second level of modelling and is inspired by our work on the concept of model of models [7]. It aims at emphasizing the main modes of variations of the F1 and F2 components across the speakers. A correlation analysis between those second levels components and some morphology features [12] has been performed in order to determine whether the principal components explaining the variations of the F1 and F2 components could be related to morphological constraints. Finally, the arrangement of the 21 sets of coefficients for the F1 and F2 components have been displayed by means of t-SNE [13] to uncover possible clusters of speakers regarding the F1 and F2 components.

3 Results

3.1.1 Acoustic modelling

The vowel spaces encompassing the F1 and F2 formants for the male and female speakers are represented Figure 1. We can observe a smaller vowel space for the male speakers, in general coherence with the literature and supporting the validity of the approach (*e.g.* [14]).

3.1.2 F1 and F2 components

The nomograms of the F1 and F2 components averaged over the male and female speakers are displayed in Figure 2. No major differences are observed between the males and the females.

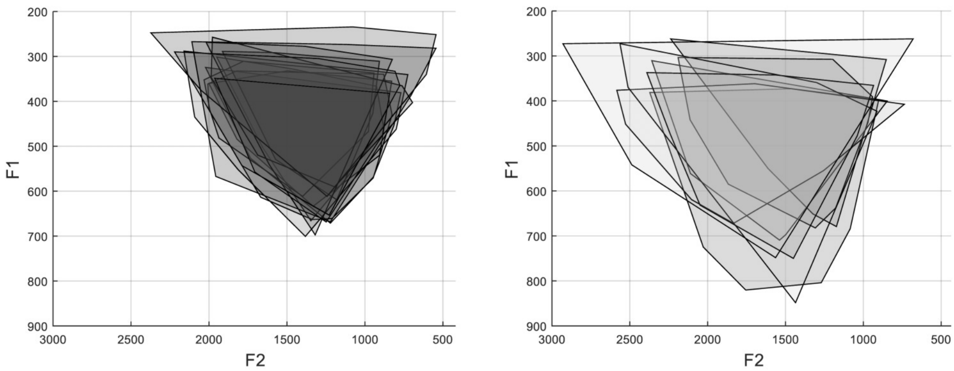

Figure 1 - Simulated vowel spaces for the male (left, in blue) and female (right, in yellow) speakers

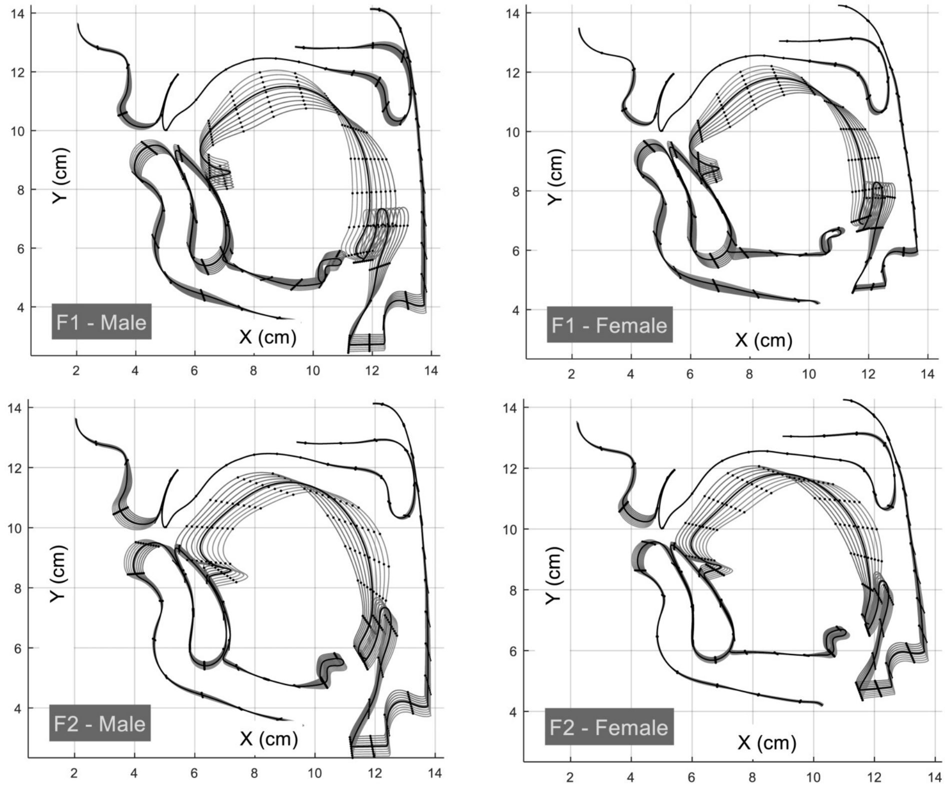

Figure 2 - Nomograms of the contours for the F1 (top) and F2 (bottom) components averaged over the male (left) and female (rights) speakers. Contours with negative (*resp.* positive) predictor values are plotted in green (*resp.* orange); one every 20 points is plotted as a black dot to emphasize deformation directions.

The F1 component corresponds to a lowering of the jaw, the lower lip and the tongue. The tongue blade follows a slightly oblique trajectory, being slightly more frontward in upper positions whereas the back of the tongue follows a horizontal frontward-backward movement. A slight protrusion associated with a lowering of the larynx is observed for the males when the jaw closes. In general, a good correlation is observed between the y-dimension of the jaw contour and F1, with a mean value of 0.7 (range 0.22-0.94, standard deviation 0.19). For a few speakers, the articulatory variations associated to F1 are not correlated with a movement of the jaw: the variations of F1 are achieved without opening or closing the jaw. Two speakers with different strategies are displayed in Figure 3. In the upper row, the F1 and JH components are very similar: the variations of F1 correspond to an opening-closing of the jaw (correlation coefficient of 0.94). In the lower row, the variations of F1 are achieved by a movement of the vocal tract rather independent from the jaw movement (correlation coefficient of 0.43).

The F2 component corresponds mainly as expected to a frontward-backward movement of the tongue, slightly oblique from an upper front position to a lower back position. The backward movement of the tongue is clearly associated with a protrusion of the lips and a lowering of the larynx.

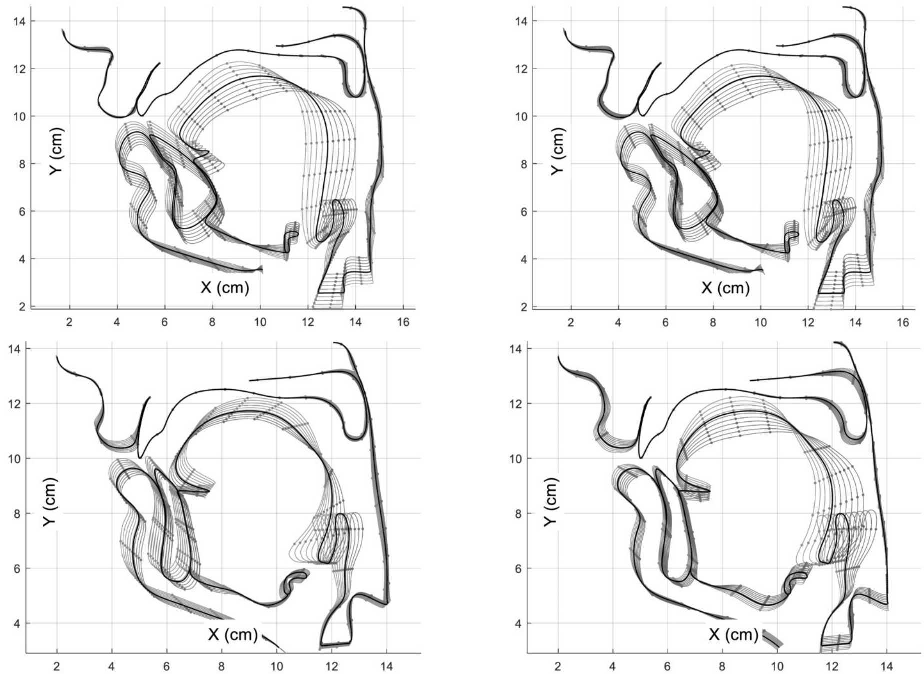

Figure 3 - Nomograms of the contours for the JH component (left) and the F1 component (right) for one French speaker (top) and one German speaker (bottom). Contours with negative (*resp.* positive) predictor values are plotted in green (*resp.* orange); one every 20 points is plotted as a black dot to emphasize deformation directions.

3.1.3 Variability analysis

Despite the high similarity of the F1 and F2 components between the male and female speakers (Figure 2), an individual analysis shows a high inter-speaker variability. As an illustration, the Figure 4 displays nomograms of the F1 and F2 components for three different speakers respectively. For the F1 component, we can observe for instance a difference in the range of variation of the tongue movement, in the directions of deformation and in the range of opening-closing of the jaw. For the F2 component, we can also observe a difference in the range of variation of the tongue movement, in the directions of deformation, as well as a difference in the amplitude of the lip protrusion and larynx raising-lowering.

The quantitative analysis did not reveal any major trend in the organisation of the F1 and F2 components across speakers. Three principal components were necessary to explain only 69% of the variance of the F1 component and 74% of the F2 component. The highest explanation was observed for the first principal component of the F2 component with 37%. More precisely, the largest trend observed for the variation of the F1 component, 34% of the variance explanation, seems to rely in the degree of use of the other articulators than the tongue: from using almost exclusively the tongue movement to raise F1 to lowering the jaw and raising concomitantly the larynx in addition. The largest trend for the variation of the F2 component, 37% of the variance explanation, seems to correspond to a variation of the backward movement of the tongue: from an association with a lowering of the blade to an association with a lowering of the back part of the tongue. All correlation coefficients between these principal components and the morphology features described in Serrurier *et al.* [12] remained below 0.6. Finally, the 21 points representing the sets of coefficients for the F1 and F2 components in the two-dimensional space defined by t-SNE sampled the overall space without revealing any cluster. These preliminary findings suggest that the inter-speaker variations observed for the F1 and F2 components reflect rather idiosyncratic strategies not related to the morphology and that common strategic trends could not be found between the speakers.

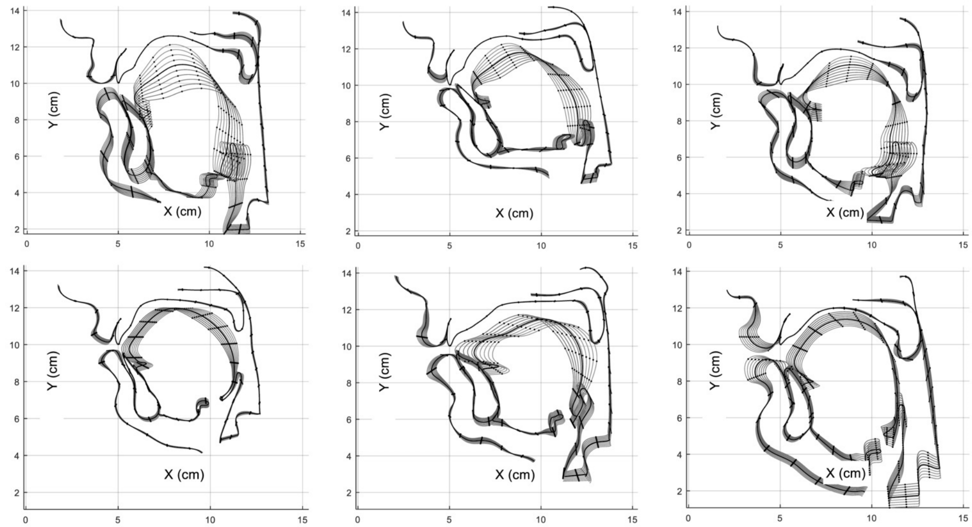

Figure 4 - Nomograms of the contours for the F1 (top) and F2 (bottom) components for six different speakers. Contours with negative (*resp.* positive) predictor values are plotted in green (*resp.* orange); one every 20 points is plotted as a black dot to emphasize deformation directions.

4 Discussion and conclusion

This study presented a modelling approach to characterise the variability of the articulatory strategy involved in the variations of the two first formants F1 and F2 for a set of 21 speakers. The acoustic variability of F1 and F2 obtained by means of acoustic simulations appears in general adequation with the literature. The articulatory components associated with a variation of F1 and F2, the F1 and F2 components, have been calculated and revealed a similar strategy for the males and the females. In average, lowering F1 is achieved by raising the jaw, the lower lip and the tongue blade and by bringing frontward the back of the tongue; it is slightly associated for the males with a protrusion of the lips and a lowering of the larynx. Lowering F2 is achieved pushing the tongue backward, protruding the lips and lowering the larynx. These general observations hide a high inter-speaker variability. Quantitative analyses of the variability of the F1 and F2 components across speaker did not reveal clear organisation nor speaker cluster. The largest trend observed for the variability of the F1 component lies in the degree of use of the jaw and the larynx in complement to the tongue movement. For the F2 component, it lies in the region of the tongue that is raised or lowered in the complement to the overall frontward-backward movement.

The results obtained in this study tend to show that the variations of the F1 and F2 components are not much driven by morphological constraints and that no underlying organisation nor causes could be uncovered. This suggests that these variations may rather be purely idiosyncratic.

Several limitations can be ascribed to this study: a limited set of 21 French and German speakers, limiting the possible quantitative analyses and the power of generalisability, a limited set of 8 to 10 artificially sustained articulations per speaker, and a lack of simultaneous acoustic recordings, obtained *in fine* by means of simulations. Despite these limitations, this study proposes an interesting methodological framework and preliminary results towards the objective. In a short term, further research may push forward the quantitative analyses to uncover potential characteristics in the organisation in the inter-speaker variability of the F1 and F2 components. Longer-term research may consider more extensive and representative acoustic and articulatory data.

Acknowledgements

The authors are very grateful to P. Badin for providing the FR data and to L. Lamalle and S. Romanzetti for participating at recording the FR and DE data. The DE corpus has been designed with the help of Bernd Kröger and Cornelia Busch. The initial tracing was performed by J.-A. Valdés Vargas, G. Ananthakrishnan, Mohammad Eslami, Arne Hülsdünker and Frederic Stepp. This research project is supported by the START-Program of the Faculty of Medicine, RWTH Aachen University. The recording of the FR data has been partially funded by the French ANR (grant ANR-08-EMER-001-02 'ARTIS').

References

[1] LINDBLOMB B., and SUNDBERG, J.: *Acoustical Consequences of Lip, Tongue, Jaw, and Larynx Movement*. In *The Journal of the Acoustical Society of America*, vol. 50, 1971, pp. 1166–1179.

[2] LADEFOGED, P., and BROADBENT, D.E.: *Information Conveyed by Vowels*. In *The Journal of the Acoustical Society of America*, vol. 29, 1957, pp. 98–104.

[3] WEIRICH, M., and SIMPSON, A.: *Articulatory vowel spaces of male and female speakers*. In *Proc. of the 10th International Seminar on Speech Production ISSP*, Cologne, Germany: 2014, pp. 449–452.

[4] LEE, J., SHAIMAN, S., and WEISMER, G.: *Relationship between tongue positions and formant frequencies in female speakers*. In *The Journal of the Acoustical Society of America*, vol. 139, 2016, pp. 426–440.

[5] DROMEY, C., JANG, G.-O. and HOLLIS, K.: *Assessing correlations between lingual movements and formants*. In *Speech Communication*, vol. 55, 2013, pp. 315–328.

[6] SERRURIER, A. and NEUSCHAEFER-RUBE, C.: *Towards articulatory comparison between French and German from MRI data*. In *Proc. of the 12th International Seminar on Speech Production ISSP*, Providence, UK: 2020.

[7] 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 *The Journal of the Acoustical Society of America*, vol. 145, Apr. 2019, pp. 2149–2170.

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

[9] 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 Communication*, vol. 36, 2002, pp. 169–180.

[10] BADIN, P., and FANT, G.: *Notes on vocal tract computation*. In *Speech Transmission Laboratory - Quarterly Progress and Status Report - Stockholm, Sweden*, vol. 25, 1984, pp. 53–108.

[11] STEVENS, K.N.: *Acoustic Phonetics*, MIT Press, 2000.

[12] SERRURIER, A., BADIN, P., and NEUSCHAEFER-RUBE, C.: *Influence of the vocal tract morphology on the F1-F2 acoustic plane*. In *Proc. of the 30th Conference on Electronic Speech Signal Processing ESSV*, Dresden, Germany: 2019, pp. 312–319.

[13] VAN DER MAATEN, L., and HINTON, G.: *Visualizing Data using t-SNE*. In *Journal of Machine Learning Research*, vol. 9, Nov. 2008, pp. 2579–2605.

[14] WHITESIDE, S.P.: *Sex-specific fundamental and formant frequency patterns in a cross-sectional study*. In *The Journal of the Acoustical Society of America*, vol. 110, 2001, pp. 464–478.