Lip protrusion/rounding dissociation in French and English consonants: /w/ *vs.* /ʃ/ and /ʒ/

Martine Toda[1, 2], Shinji Maeda[3] Andreas J. Carlen[4] and Lyes Meftahi[4]

1. Laboratoire de Phonétique et Phonologie (UMR 7018), CNRS/Sorbonne Nouvelle, Paris, France
2. ATR Human Information Science Laboratories, Kyoto, Japan
3. ENST / CNRS (URA 820), Paris, France; 4. SyncMagic, Paris, France

E-mail: mtoda@his.atr.co.jp

ABSTRACT

In this study of facial movements during speech, we found a consistent tendency for two native French speakers and an American English speaker (also reported in Toda *et al*. [1]): there exist two separate lip gestures for consonants involving 'labialization'. The first one is a lip protrusion with rounding (narrow lip area) observed for /w/, and the second is a protrusion with a large lip opening observed for postalveolar sibilants /ʃ, ʒ/. Furthermore, this separation of lip patterns is supported by a factor analysis on the movement data, indicating that these gestures are functionally autonomous. We can relate this fact to the assumed modes of resonance for the consonants: rounding is a requirement for the Helmholtz' resonance of /w/ being responsible of low F2, whereas protrusion can be intended to lower the frequency of frication noise of postalveolars, enhancing the contrast with their alveolar counterparts. The acoustic/aerodynamic function of the large lip opening for postalveolars remains to be explained.

1. INTRODUCTION

English and French languages have consonants that are characterized by lip protrusion and/or rounding as principal or secondary articulation. We will refer to these lip gestures by the generic term of 'labialization'. Both in French and English, postalveolar sibilant fricatives /ʃ, ʒ/ are labialized [2, p.148]. The English rhotic approximant /ɹ/ also possesses labialized variants [2, p.234]. Finally, there are two labial approximants in French: /w/ and /ɥ/; and one in English: /w/. The status of lip gestures differs between the two languages, however. In French, lip shape contributes to phonological contrasts of front vowels, while in English, labialization is not a standing-alone feature.

The exact nature of lip gestures in labialized consonants is not clear, however. Anatomy of the lip muscles suggests their capability to form complex shapes (see discussion in Maeda *et al*. [3]). So far, a few studies used techniques that permit to capture three-dimensional (3D) articulatory data (examples are MRI, [4]; ultrasound, [5]; and Optotrak, [6]).

The present study compares 3D facial motion-capture data from three subjects (one American and two French) for labialized and non-labialized consonants uttered in nonsense VCV tokens. We measure the front-back position of the lips and approximate lip area in order to evaluate the amount of lip protrusion/retraction and 'rounding'. In parallel, we use a factor analysis to extract components of facial movements for each subject following a similar procedure as described in Maeda *et al*. [3]. This method permits us to extract relatively 'pure' components by isolating respective contributions of the jaw and the lips. Since the model is data-driven, the resulting factors can be considered as functional units of facial movements for speech. By this method, independently of subjects and their language, we obtain two functionally distinct labial components. We will discuss this articulatory fact with regard to the assumed underlying mechanisms of acoustic production.

2. METHOD

2.1 Subjects and Corpora

One American English subject (SE, male) and two French subjects (SF1, female and SF2, male) read a corpus containing nonsense words, short sentences in three speaking rates ('normal', 'slow' and 'fast' rate determined by subjects), and a short text, in the corresponding language. For protrusion and area measurements, data are from VCV nonsense words with symmetrical vocalic contexts, where C={dentals /t, d/, alveolar fricatives /s, z/ (also affricates /ts, dz/ for English), postalveolar fricatives /ʃ, ʒ/ (also affricates /tʃ, dʒ/ for English), and /w/ (only in non-labial contexts in French)}; and V=/a, i, u and y (for French)/. Dental stops and alveolar sibilants have been chosen as 'neutral lip' reference. All the available data served in the functional analysis, a total of 137s of speech (containing pauses) for subject SE, 271s for SF1 and 193s for SF3.

2.2 Data acquisition

A Vicon Motion Capture device has been used for data acquisition [7]. The data, a set of 3D coordinates of markers in a sequence of frames were acquired in two sessions. In the first session (English), six infrared cameras were placed around subject's face and recorded 65 markers glued on the subjects' face (61 markers) and on the neck (4) at the frame rate of 120 Hz. In the second session (French), eight cameras were used and the number of face markers was 63 instead of 61, but 61 of them were placed approximately at

ISBN 1-876346-48-5 © 2003 UAB

the same place on the face of subjects as in the first session. Instrumental precision of facial motion capture is estimated to be less than 0.2 mm.

Five markers, four on the forehead and one on the nose tip, served to eliminate head motion through frames. Six selected markers served for lip protrusion and area measurements in this study (see Figure 1). All of the 61 (English) and 63 (French) face markers were used in the factor analysis. The influence of the number of markers in the factor analysis is negligible (cumulative difference in percentage of explained variance across the first five factors extracted with either 61 or 63 face markers is about 1.0% for subject SF1 and 1.6% for SF2).

2.3a Protrusion and area measures

Protrusion was defined as the mean distance of representative lip markers (red dots) from a reference marker *N* (nasion) in the y-dimension (Figure. 1, left).

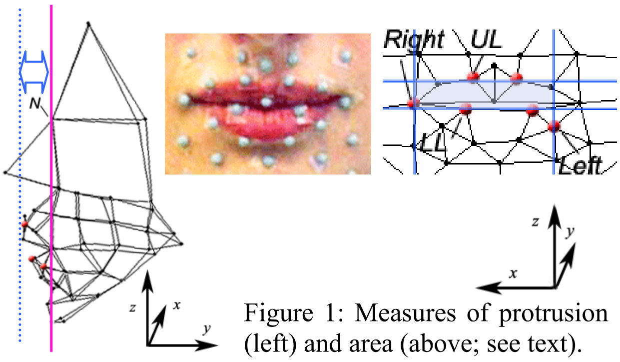

Figure 1: Measures of protrusion (left) and area (above; see text).

We calculated the area of a rectangle of height (LL_z-UL_z) × width ($Left_x$-$Right_x$) as an indicator of lip opening area (Figure 1, right). According to Abry and Boë [8], this measure is highly correlated with the true area. The selected lip and reference markers were always detected in the motion capture and thus had no drop of data. Protrusion and area were measured at the mid-point of consonants.

2.3b Functional analysis

The analysis consists of an arbitrary orthogonal factors analysis (AFA) to extract marker movements correlated with jaw motion (measured by choosing an appropriate marker, *J*, in Figure 3) in two dimensions (up-down and front-back), followed by a principal component analysis (PCA) on the residual correlations ([3] for details). In this study, we decided to limit the number of arbitrary factors to two instead of three (that would have exhaustively extracted movements correlated with the three dimensional movements of the jaw). The reason is that in an AFA with three arbitrary factors, the third factor of SF1 (also SF2 in a smaller proportion) contained an important lip component in addition to the left-right jaw movements. It can be explained by an 'accidental' correlation, resulting from the systematic use of asymmetrical movements in postalveolar sibilants (as shown in Figure 4), that are accompanied by a lip gesture, as attested by the measurement results. Therefore, we considered the use of two arbitrary factors to be the best compromise. The 'pureness' of the resulting components is still acceptable and they can be interpreted phonetically.

3. RESULTS

3a. Protrusion and area

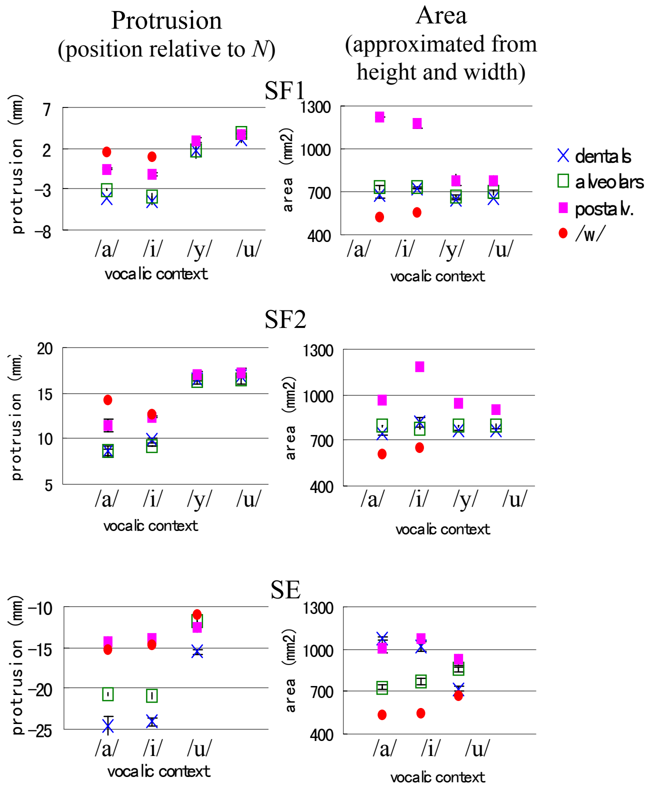

Figure 2: Average protrusion and area for 4 groups of consonants (dentals, alveolars, postalveolars and /w/) in various vocalic contexts. Contrasts of voicing and manner (for English) are ignored. (The value is not calculated from the same number of tokens for each dot.)

For all the subjects, /w/ and postalveolars are protruded in comparison with alveolar sibilants and dental stops, in contexts /a-i/ (Figure 2, left column). Subject SF1 seems to use a supplementary distinction of the degrees of protrusion between /w/ and postalveolars, where /w/ is more protruded than postalveolars. Alveolar sibilants and dental stops show a very similar behavior across the subjects: not protruded in contexts /a-i/ and protruded in contexts /y-u/. The difference in the degrees of protrusion between alveolars and postalveolars is totally neutralized (SF1 and SF2) or partially (SE) in contexts /y-u/. We can consider that these tendencies are reliable, since the dispersion is very small, although the number of utterances is limited.

The area data (Figure 2, right column) show a very different pattern in comparison with the protrusion data. For SF1 and SF2, postalveolars have the largest area in contexts /a-i/, and even in contexts /y-u/ for SF2. In the opposite, /w/ show the smallest lip area for all the subjects. For SF1 and SF2, alveolar sibilants and dental stops have a similar lip area through all vocalic contexts, whereas the area of dental stops for SE is closer to the postalveolars' than to the

ISBN 1-876346-48-5 © 2003 UAB

alveolars'. In both cases, postalveolar sibilants have a larger lip area than alveolar sibilants, at least in /a-i/ vocalic contexts. If we consider the area as a measure of 'rounding', we can conclude that (1) /w/ is rounded; and (2) postalveolars are not rounded, or rather 'open'.

These observations comfort us in the assertion that there exist at least two distinct patterns of labialization [9]. Moreover, we can safely assume that some postalveolar sibilants (though it could depend on language) are produced with a special effort for enlarging the lip opening area (the opposite of 'rounding'), since we observe a larger area than 'reference' consonants (dental stops) for two of the subjects (SF1 and SF2).

3b. Functional analysis

Factors resulting from the analysis are shown in Figure 3. We discarded factors 6 and higher ones because their contribution to the total variance became very small (less than 3 %). Factors 1 and 2 contain movements correlated with high-low (factor 1) and front-back (factor 2) dimensions of the marker *J* located on the chin. The pattern of the arbitrary factors for SF2 and SE (not shown here) are very similar to those of SF2 except the percentage of total variance explained by them (see legend).

Factor 3, the first factor extracted from PCA, contains a movement of 'protrusion with rounding' vs. 'retraction with spreading' for all the subjects and explains from 26% to 34% of the total variance ('PR factor' hereafter). Among the following factors, factor 5 of SF1 and SF2, and factor 4 of SE contain mainly left-right 'asymmetric' movements of the jaw and the surrounding tissues ('left-right jaw factor' hereafter). Recall that this dimension of jaw movement had intentionally been left by the AFA because of an unexpected strong lip protrusion associated with it for SF1 (and SF2 in a smaller extent), which we wanted to separate. Factor 4 for SF1 and SF2, and factor 5 for SE contain another lip component, which is perpendicular to that of factor 3 (referred to as 'vertical lip factor'). These factors contain also a somewhat curious up/down movement of cheeks, especially for SF1 and SF2. Although no phonetic contribution may be accounted for by such gestures, it is possible that they result secondarily from articulation, such as upper lip rising and lowering intended to control lip area. Indeed, as shown in Figure 4, the vertical lip factor is globally a good measure to isolate postalveolars from other consonants. We just demonstrated that a large lip area characterizes postalveolars and we assume that the upper lip gesture contained in the 'vertical lip factor' has an important role in the achievement of this target.

By using the inverse of factors, we can calculate the value of each factor for every frame from the normalized motion data ([3] for details). Figure 4 plots the value of factors at the mid-point of consonants.

As for the difference of lip area between consonants in 'non-labialized' contexts (/a/ and /i/), two individual strategies can be pointed out. SF2 and SE keep the jaw height (factor 1) almost constant for all the sibilants and use a vertical lip gesture to vary lip area. On the other hand, both jaw opening and vertical lip gesture are involved for SF1 in making a large lip area (although the jaw position may result primarily from the control of the position of the tongue and teeth).

Another noticeable inter-subject variation occur in the control of lip area between consonants uttered in 'labialized' contexts /y/ and /u/, and non-labial contexts /a/ and /i/. SF1 and SE seemingly do not differentiate jaw height between consonants uttered in labialized and

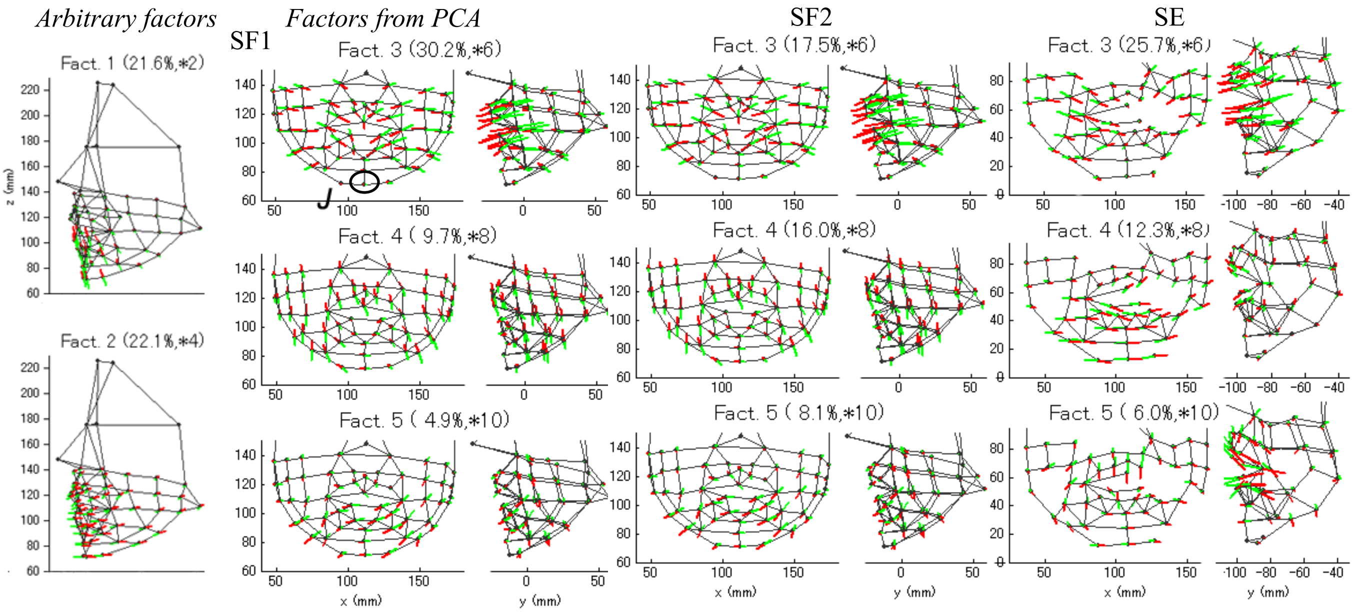

Figure 3: Factors from AFA (shown only for SF1, left) and PCA. Percentage of the total variance explained by each factor is indicated in parentheses (the value for factors 1 and 2 is SF2: 34.1% and 9.2%; SE: 30.8% and 11.4%, respectively). The length of vectors (tick marks) associated to each marker is for +/- the value of std. deviation indicated after the percentage.

 ISBN 1-876346-48-5

non-labialized contexts, whereas SF2 clearly distinguish two degrees of jaw height (higher for labialized contexts and /w/). Finally, all the subjects make a 'protrusion with rounding' gesture (factor 3) in labialized contexts.

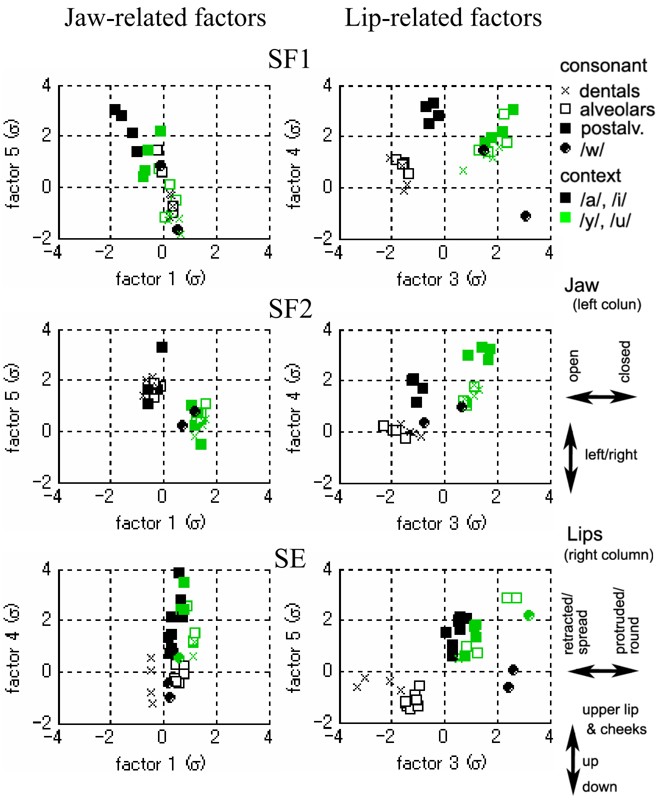

Figure 4: Scatter plot of jaw-related factors: high-low jaw factor (f1: left column, x-axis) and left-right jaw factor (f5 or f4: y-axis); and lip-related factors: PR factor (f3: right column, x axis) and lip vertical factor (f4 or f5: y-axis) for the three subjects. Values of the lip vertical factor of SE are +/- inverted to correspond to the same directions as for SF1 and SF2. Consonant groups are identified by shape and vocalic contexts by color (/a and i/: black; /y and u/: green).

In summary, lip-related factors extracted from PCA fit well the phonetic categories of the consonants. In addition, the combination of lip-related factors, f3 and f4 or f5 depending on subjects, seem optimal to characterize consonants in the articulatory space, at least for the subjects and/or languages actually studied.

DISCUSSION AND CONCLUSION

By decomposing labial gestures in the factor analysis, we observed that 'labialization' could appropriately be specified by two lip components. Direct measurements revealed that both postalveolars and /w/ are protruded, for all the subjects we studied, but the two consonant groups are completely opposed concerning the estimate of lip area, /w/ being closed by rounding and postalveolars 'open'.

We assume this accurate control of lip shape to be related to the acoustic process. We demonstrated in the previous study [1] the importance of protrusion in the noise spectrum of postalveolar sibilants. The labial-velar approximant /w/ is usually modeled as a double Helmholtz' resonator [10], with the lip tube being the nozzle of the front cavity. The lips should, therefore, remain rounded to insure the Helmholtz configuration; their protrusion theoretically lowers the related resonance (resulting in a low F2). No definitive explanation, however, can be given to the difference of lip area observed between labialized consonants. Nevertheless, we can assume the fundamental difference between /w/ and (postalveolar-) sibilants to be as follows: the source at the glottis of the former is relatively independent from the shape of the vocal tract, while it is less likely for the latter. The noise source of sibilants is well modeled by a turbulence resulting from an air jet (created by a narrow constriction) striking an obstacle (teeth) [11]. This may create conditions in which source-resonator interactions occur: the properties of the vocal tract would have an influence both on source generation and resonance. Trade-off relations resulting in such cases would differ from those of vowel-like sounds.

We thank Elise Bedaton of Synthétique (Bondues, France) and Rémi Brun of Attitude Studio (Pantin, France) for their expertise in the data acquisition. This work was supported in part by the project SAALSA No. 7215/17/00 in the program PRIAMM of CNC, France, and by Telecommunications Advancement Organization of Japan.

REFERENCES

[1] M. Toda, S. Maeda, A. J. Carlen and L. Meftahi, "Lip gestures in English sibilants: articulatory-acoustic relationship", *proc. 7th ICSLP*, 2165-2168, 2002.

[2] P. Ladefoged and I. Maddieson, *The Sounds of the World's Languages*, Blackwell, Mass., 1996 (1999).

[3] S. Maeda, M. Toda, A. J. Carlen and L. Meftahi, "Functional modeling of face movements during speech", *proc. 7th ICSLP*, 1529-1532, 2002.

[4] S. S. Narayanan, A. Alwan and K. Haker, "An articulatory study of fricative consonants using magnetic resonance imaging", *J. Acous. Soc. Am.* 98(3), p. 1325-47, 1995.

[5] M. Stone and A. Lundberg, "Three-dimensional tongue surface shapes of English consonants and vowels", *J. Acous. Soc. Am.* 99(6), p. 3728-3737, 1996.

[6] H. Yehia, P. Rubin, and E. Vatikiotis-Bateson, "Quantitative association of vocal-tract and facial behavior", *Speech Com.* 26(1-2), p. 23-43., 1998.

[7] http://www.vicon.com

[8] C. Abry and L.-J. Boë, "Laws for lips", *Speech Com.* 5, 97-104, 1986.

[9] M. Toda, S. Maeda, A. J. Carlen, and L. Meftahi, "Dissociation de la protrusion et de l'arrondissement dans la production des consonnes labialisées de l'anglais", *24èmes Journées d'Étude sur la Parole*, Nancy, 2002.

[10] G. Fant, *Acoustic theory of speech production*, Mouton, The Hague, 1960.

[11] C. Shadle, *The Acoustics of Fricative Consonants*, PhD thesis, M.I.T., 1985.