# Strategies of labial coarticulation

Vincent Robert, Brigitte Wrobel-Dautcourt, Yves Laprie, Anne Bonneau

# Strategies of labial coarticulation

*Vincent ROBERT, Brigitte WROBEL-DAUTCOURT, Yves LAPRIE, Anne BONNEAU*

Speech Group, LORIA UMR 7503
BP 239 - 54506 Vandoeuvre
FRANCE

http://parole.loria.fr
email : vrobert@loria.fr

## Abstract

In this article, we present first conclusions about labial coarticulation strategies drawn from a corpus of speech audiovisual data. The general idea is to develop a talking head which would be understandable by lip readers especially deaf persons. With a stereovision system, we recorded a corpus with ten French native speakers (5 female and 5 male speakers). Visual and audio information was analysed to extract the labial parameters (opening of the lips, stretching and protrusion). Even if the analysis shows a great variability between speakers, we nevertheless found general tendencies which would help us to develop a reliable prediction algorithm of labial coarticulation.

## 1. Introduction

Our goal is to derive invariants and rules that should be verified by a labial coarticulation prediction model intended to pilot a talking head understandable by lip readers, especially deaf people.

Three main coarticulation models have been proposed: look-ahead, time-locked, hybrid. These models predict both the onset and the dynamics of labial coarticulation in $V_1CV_2$ and $V_1CCV_2$ sequences, where $V_1$ is unrounded, $V_2$ rounded and C is neutral with respect to labial coarticulation. In the Look-ahead model proposed by Henke[4] and Öhman [9] protrusion starts at the end of the unrounded vowel. On the other hand, in the Time-locked model proposed by Bell-Berti and Harris [2] (also called Coproduction by Boyce at al.) the temporal interval between the onset of the rounding movement and the onset of the rounded vowel is constant. Perkell and Chiang [6] proposed a hybrid model which predicts a two-stage gesture. The first low velocity stage begins at the offset of the first unrounded vowel and the onset of the higher velocity stage is linked to the rounded vowel as predicted by the time-locked model.

In the case of the anticipation of protrusion, these three models were not able to explained data acquired by Abry and Lalouache [1] who advocate an expansion model. According to their measures, the temporal interval of protrusion movements is easily expansive but cannot be easily compressed under the temporal observed for vowel to vowel transitions. However, it turns out that there is a great inter-speaker variability and they give anticipation expansion coefficients depending on speakers.

As mentioned in most of the works about labial coarticulation, data exhibit a large inter-speaker variability. We thus built a corpus for a larger number of speakers (five male and five female speakers) than previous comparable studies in order to enable a more precise evaluation of existing labial coarticulation models and draw speaker independent rules.

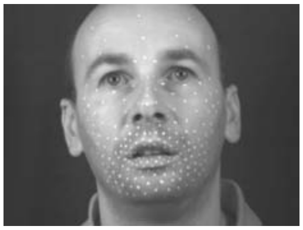

Figure 1: 210 white markers are painted on the speaker's face.

In addition, we are interested in studying the influence of the effect of labial consonant /p/ or consonants that imply a protrusion movement /ʃ/. After the description of the acquisition method and the corpus, we present the most salient conclusions about protrusion movements and labial coarticulation strategies.

## 2. Data acquisition

One of the main challenges to enable the investigation of labial coarticulation is the design of a low cost and easily available 3D acquisition infrastructure. The system we designed is more flexible than existing motion capture systems that usually use infrared cameras and glued markers.

Our system only uses two standard cameras, a PC and painted markers that do not change speech articulation and provide a sufficiently fast acquisition rate to enable an efficient temporal tracking of 3D points. We drawn for example 210 markers on the face (46 on lips) to enable a precise lip shape deformation information recovery (fig. 1) to build a precise face talking face ([8]).

For the study reported here, the data set was composed of the stereovision recording of 10 French native speakers (5 female et 5 male speakers), each speaker talked during about 120 seconds. After processing these sequences, our corpus is composed of 118.400 frames × 15 points × 3 coordinates.

## 3. Measurements

We extracted three labial parameters: Opening, Stretching and Protrusion (see Fig. 2). There are several ways of measuring lip protrusion: higher lip only, both lips together, both lips plus lip commissures. . . We thus conducted a preliminary experiment with a larger number of markers (210 on lips, jaw and cheeks) for one speaker and applied principal component analysis to find the most important factors and markers the most tightly associated to these factors. It turned out that lip commissures are also closely related to the protrusion movement. We thus designed the protrusion measure by taking into account the four markers $A$, $B$, $C$ and $D$ (see Fig.2). Protrusion is evaluated as follows. Firstly, we determine the average (on all the frames acquired for one speaker) vector normal to the plane formed by vectors $\vec{AB}$ and $\vec{CD}$, and a reference point $F$ after the overall head movements have been compensated for. Protrusion is given by the distance between the center of gravity of the four reference markers ($A$, $B$, $C$, $D$) and $F$. In addition to its relevancy with respect to protrusion movements, this measure turns out to give a smaller noise than isolated markers. Despite, the very low level of noise (see Fig.3) we applied a slight smoothing through regularized splines to allow a relevant computation of protrusion velocity.

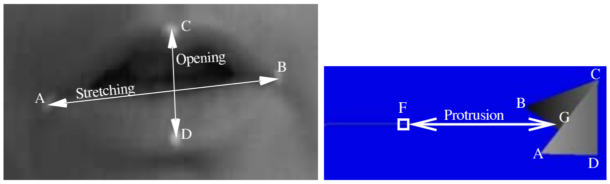

Figure 2: Mesuring Opening, protrusion and stretching

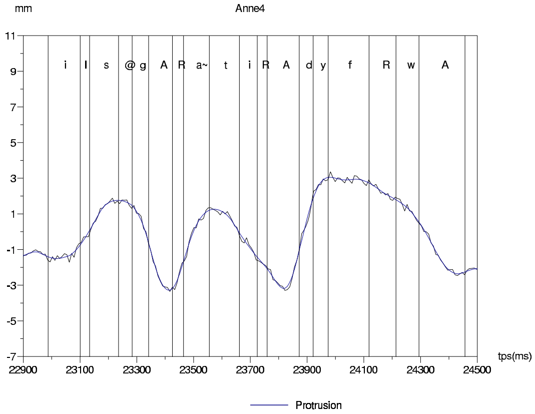

Figure 3: Protrusion values before and after smoothing

We did not use these measurements in millimeters to perform comparisons between speakers because the measures are directly influenced by the anatomic characteristics of speaker. We thus use centered and normalized values. In all the coarticulation figures, the $y$ axis represents $\frac{X-\mu}{\sigma}$, where $X$ is the parameter considered, $\mu$ the average value of this parameter over all the speech segments uttered by a subject and $\sigma$ its standard deviation.

## 4. Corpus and first results

### 4.1. Corpus

Our corpus was made up of 4 isolated vowels (/i, y, a, o/), 6 consonants (/p, t, d ,s, ʃ, f/) followed by schwa, 8 CV, 20 VCV, 18 VCCV and 2 phonetically balanced sentences. Unlike most of the previous studies we also included consonants with a primary labial articulation (/p, f/) and a secondary (but not mandatory) one (/ʃ/) because we are also interested in the general process of labial coarticulation. A carrier sentence as neutral as possible with respect to labial coarticulation and prosody was chosen. With measurements from ten French native speakers (5 male and 5 female speakers), our results take more account of inter-speaker variability than studies which include, in the best of cases, four (Abry & Lallouache[1], Parkell & Matthies [7]) or seven (Matthies & Perrier [5]) speakers.

Finally, we extracted 380 curves from our corpus which give us a large panel of inter speaker and intra speaker variability.

### 4.2. Analysis of the results

Our aim is to study labial coarticulation, its spanning and strength, in $V_1CV_2$ or $V_1CCV_2$ sequences. We will try to explain coarticulation as a function of the specific labial characteristics of each sound. Our corpus will also allow us to determine the values of each articulatory parameter (opening, stretching, and protrusion) for each phoneme of the database.

#### *4.2.1. Labial parameters for each phoneme*

Before recording our corpus, we have proposed an estimation of the three labial parameters for each phoneme (fig 4). As there is no systematic description of these features in the literature, then the data will provide a good opportunity to establish it by refining partial existing descriptions.

| Phoneme | Opening | Stretching | Protrusion |
|---|---|---|---|
| i | $O_1$ | $E_4$ | $P_1$ |
| a | $O_4$ | $E_1$ | $P_1$ |
| y | $O_1$ | $E_1$ | $P_4$ |
| o | $O_2$ | $E_1$ | $P_3$ |
| p | $O_0$ | | |
| t | | | |
| k | | | |
| f | $0_{0.5}$ | | |
| s | | | |
| ʃ | | | $P_3$ |
| r | | | |
| ɹ | | | |

Figure 4: Extract of our phonetic classification

Figure 4 shows the average degree of stretching, opening, and protrusion (rated from 0 to 4) for each phoneme of our corpus. This description is independent of the phonetic context. Note that the relative values of each parameter are easier to estimate for sounds belonging to a same class than for sounds from different classes. As an example, if we know that the stretching of /i/ is larger than that of /a/, the comparison between the stretching of /a/ (non labialised vowel) and that of /y/ (labialised vowel) is less evident. Whereas all vowels have typical values for the three labial parameters, we made the hypothesis that only four classes of consonants strongly influence lip shape: the bilabial consonants /p,b,m/ articulated with closed lips, the labial consonants /f,v/, articulated with slightly opened lips -the lower lip is very close to the upper front teeth- , the fricatives /ʃ, ʒ/, for which protrusion enhances their acoustic specificities, as well as the semi-vowels /j, w, ɥ/. One member of each of the first three classes is present in our corpus /p,f,ʃ/.

Our results confirm the phonetic description proposed (see figures 5). These results represent the variability of protrusion, opening and stretching for all speakers and all contexts of our corpus. Data are normalized as described before to reduce anatomic effects of speakers. We will briefly comment some of the results. We observe that the vowels /y, o/ are always more protruded than any other vowel of the corpus, and their values are always positive, even when we consider their minimum. We can make the same remark for the stretching of /i/. The protrusion of /ʃ/ is the most important with respect to that of other consonants. On the other hand, the opening values are always negative for /p/ and /f/.

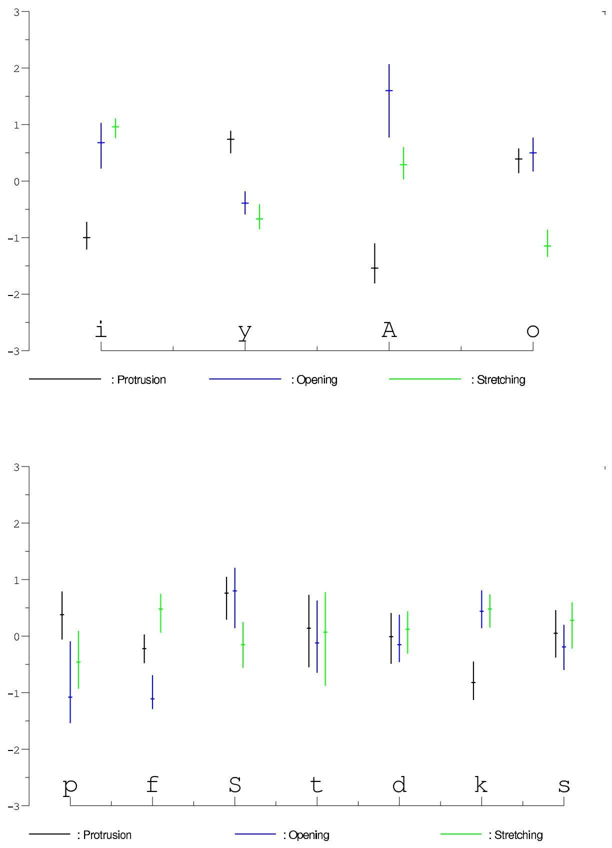

Figure 5: Variability of Protrusion, Opening and Stretching for vowels and consonants

Concerning /ʃ/ the measured opening could seem surprisingly large. Actually, this measure does not exactly correspond to lip opening but to the distance between the two points $C$ and $D$ (see fig 2). Since /ʃ/ is fairly protruded, the "unfolding" of lips due to protrusion generates an increase of lip opening as it can be seen in fig 6.

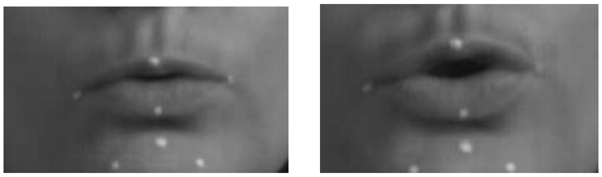

Figure 6: /s/ and /ʃ/ phonemes in /isy/ and /iʃy/ sequence

We also noted that great tendences are respected from one speaker to the other. In the case of /i/ for instance (fig 7), the stretching is always prominent whatever the speaker.

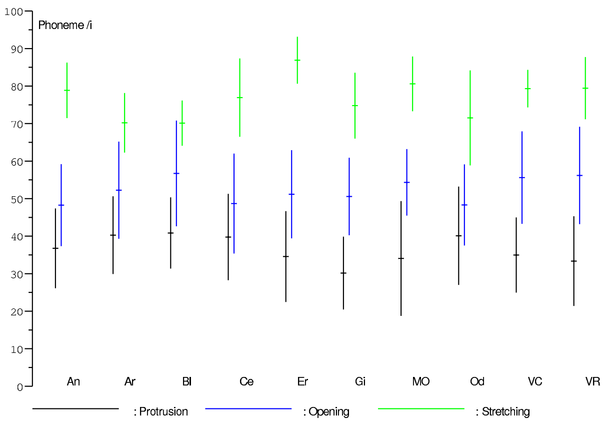

Figure 7: Variability of values for protrusion, opening and stretching for all the speakers and for the phonemes /i/

### 4.2.2. *Anticipation of labial articulation*

We give here a first series of visible results concerning the anticipation of $V_2$ labial gestures in $VCV$ and $VCCV$ sequences. A more thorough analysis of all the results will be undertaken. When the consonants have no specific characteristics (see fig 4), we make the hypothesis that the movement can begin as soon as the end of the first vowel, and we will try to determine whether the beginning of labial anticipation can be explained by a coarticulation model. We assume that the anticipatory movement is blocked to a certain extent when a consonant has specific labial characteristics which differs from that of $V_2$. We will also verify whether labial coarticulation affects the main characteristics of each vowel.

A first analysis of our data allow us to verify whether our results were in agreement with one of the three main models, when the consonant is neutral with respect to lip protrusion: Look-Ahead ([4], [9]), time-locked ([2]) or hybrid ([3],[6]). For that purpose, we tried to find the beginning of the protrusion movement for $V_1CV_2$ sequences when $V_1$ is unprotruded, $C$ is a neutral consonant and $V_2$ is protruded. We measured the beginning of the protrusion movement (maximum acceleration) and the maximum of protrusion. Protrusion always begin in the first half of the consonant, but not necessarily at the beginning of the neutral consonant, so we did not observe a look-ahead model. The time between the beginning of the protrusion and the maximum of protrusion is also very variable, which is in contradiction with the time-locked and the hybrid model (fig 8).

We made the hypothesis that the anticipation is blocked when the consonant has specific values which differs from that of $V_2$. This appears very clearly with the bilabial /p/, for which lips must be closed, and which always blocked the anticipation of lip opening, as well as for the labio-dental /f/, even if, as expected, the effect of this consonant on the opening gesture is less drastic.

### 4.2.3. *Inter- and intra-speaker strategies*

**Important inter-speaker variability.**

If we except the main trends reported above which are speaker independent, the data exhibit an important degree of speaker variability.

In the case /ykʃi/ for instance (see Fig.9), inter-speaker cor-

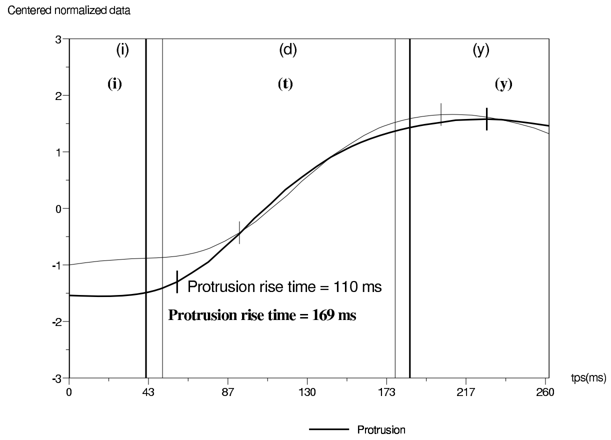

Figure 8: Protrusion evolution for /ity/ and /idy/ sequences for one speaker

relation range from 0.46 to 0.97. We thus tried to make some groups of similar speakers emerge. In the case of /ykʃi/ it can be seen that a first group of 6 speakers out of the 10 present very similar strategies. A second group of two speakers (MO and VR) present similar strategies and two other speakers intermediary strategies (Er and Od). It seems that these last two speakers strengthen more the protrusion of /ʃ/ than others. However, these groups of speakers do not emerge in other situations. Further investigations will be necessary to find out whether more relevant and constant groups of speakers exist.

| | An | Ar | Bl | Ce | Gi | VC | *Er* | *Od* | **MO** | **VR** |
|---|---|---|---|---|---|---|---|---|---|---|
| An | 1 | 0.97 | 0.88 | 0.90 | 0.97 | 0.96 | *0.86* | *0.85* | **0.73** | **0.73** |
| Ar | 0.97 | 1 | 0.90 | 0.95 | 0.96 | 0.95 | *0.89* | *0.79* | **0.75** | **0.81** |
| Bl | 0.88 | 0.90 | 1 | 0.94 | 0.88 | 0.91 | *0.92* | *0.68* | **0.78** | **0.79** |
| Ce | 0.90 | 0.95 | 0.94 | 1 | 0.89 | 0.89 | *0.89* | *0.64* | **0.69** | **0.79** |
| Gi | 0.97 | 0.96 | 0.88 | 0.89 | 1 | 0.98 | *0.85* | *0.88* | **0.74** | **0.75** |
| VC | 0.96 | 0.95 | 0.91 | 0.89 | 0.98 | 1 | *0.88* | *0.85* | **0.78** | **0.77** |
| *Er* | *0.86* | *0.89* | *0.92* | *0.89* | *0.85* | *0.88* | *1* | *0.73* | **0.86** | **0.80** |
| *O* | *0.85* | *0.79* | *0.68* | *0.64* | *0.88* | *0.85* | *0.73* | *1* | **0.71** | **0.60** |
| **MO** | **0.73** | **0.75** | **0.78** | **0.69** | **0.74** | **0.78** | **0.86** | **0.71** | **1** | **0.90** |
| **VR** | **0.73** | **0.81** | **0.79** | **0.79** | **0.75** | **0.77** | **0.80** | **0.60** | **0.90** | **1** |

Figure 9: Correlations of Protrusion, Opening and Stretching between all speakers for the /ykʃi/ sequence

**Small intra-speaker variability.**

In order to investigate the intra-speaker variability we examined a series of pairs of utterances that should present similar or very similar coarticulation profiles. VCV or VCCV logatoms of each pair only differ from one consonant, for instance /ity/ and /idy/. In this first case /t,d/ share the same features but one (the voicing feature). In other cases, the consonants vary with respect to the place of articulation (/t/ vs. /k/ like /ikʃy/ and /itʃy/, /ykʃi/ and /ytʃi). This deeper modification should not involve any modification in the labial coarticulation strategy. For each speaker and each pair of logatoms, we calculated the correlation between the two coarticulation profiles. Fig. 10 give results averaged over these pairs. The inter-speaker correlation given in Tab. 10 is the average correlation between the profile of the speaker considered and those of other speakers.

## 5. Conclusions

Strong relations between protrusion, stretching and opening are exhibited by our data. The most important one is the relation between protrusion and stretching (|correlation| > 0.95), which move in opposite directions. We also noted that, especially for vowels, the same relation between protrusion and opening.

| Speaker | An | Ar | Bl | Ce | Er | Gi | MO | Od | VC | VR |
|---|---|---|---|---|---|---|---|---|---|---|
| Intra Cor | 0.94 | 0.90 | 0.95 | 0.97 | 0.84 | 0.91 | 0.88 | 0.96 | 0.96 | 0.86 |
| Inter Cor | 0.89 | 0.89 | 0.87 | 0.89 | 0.86 | 0.84 | 0.78 | 0.86 | 0.87 | 0.77 |

Figure 10: Correlations between close sequences, each speaker with itself (intra correlation) and each speakers with the others (inter correlation).

The second conclusion is that there is a large degree of freedom in the labial realizations exhibited by speakers. However, some strong constraints on labial distinctive features of vowels and some consonants must be respected. This large degree of freedom is in contradiction with existing labial coarticulation models.

This means that a general prediction model should incorporate two components: a limited set of coarticulation constraints complemented by speaker-dependent strategies. In order to choose an efficient speaker dependent strategy we will search for the strategy which gives the best lipreading results by deaf people.

## 6. References

[1] C. Abry and T. Lallouache. Le mem: un modèle d'anticipation paramétrable par locuteur: Données sur l'arrondissement en français. *Bulletin de la communication parlée, 3*, pages 85–99, 1995.

[2] F Bell-Berti and K.S Harris. A temporal model of speech production. *Phonetica*, 38:9–20, 1981.

[3] R. A. Bladon and A. Al-Bamerni. One stage and two-stage temporal patterns of velar coarticulation. *The Journal of the Acoustical Society of America*, 72, 1982.

[4] W. Henke. Preliminaries to speech synthesis based on an articulatory model. pages 170–171, 1967.

[5] Mathies M., Perrier P., Perkell J.S., and Zandipour M. Variation in anticipatory coarticulation with changes in clarity an rate. *Journal of Speech,Language and Hearing Research*, 44:340–353, 2001.

[6] J.S Perkell and C.M Chiang. Preliminary support for a 'hybrid model' of anticipatory coarticulation. *Proceeding of the XIIth International Congress of Acoustic*, 1986.

[7] J.S Perkell and M.L Matthies. Temporal measures of anticipatory labial coarticulation for the vowel /u/: Within-and cross-subject variability. *The Journal of the Acoustical Society of America*, 91:2911–2925, 1992.

[8] B. Wrobel-Dautcourt, M.-O. Berger, B. Potard, Y. Laprie, and S. Ouni. A low-cost stereovision based system for acquisition of visible articulatory data. In *Proceedings of the 5th Conference on Auditory-Visual Speech Processing, Vancouver Island, BC, Canada*, 2005. submitted.

[9] Sven E. G. Öhman. Numerical model of coarticulation. *The Journal of the Acoustical Society of America*, 39:310–320, 1967.