

Contextual effects on protrusion and lip opening for /i,y/

Anne Bonneau, Julie Busset, Brigitte Wrobel-Dautcourt

► **To cite this version:**

Anne Bonneau, Julie Busset, Brigitte Wrobel-Dautcourt. Contextual effects on protrusion and lip opening for /i,y/. 10th Annual Conference of the International Speech Communication Association - Interspeech 2009, Sep 2009, Brighton, United Kingdom. 2009. <inria-00433381>

HAL Id: inria-00433381

<https://hal.inria.fr/inria-00433381>

Submitted on 19 Nov 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Contextual effects on protrusion and lip opening for /i,y/

Anne Bonneau¹, Julie Busset¹, Brigitte Wrobel-Dautcourt¹

¹ Speech Group, LORIA UMR 7503 BP239 54506 Vandoeuvre-lès-Nancy

anne.bonneau@loria.fr, busset@loria.fr, wrobel@loria.fr

Abstract

This study investigates the effect of “adverse” contexts, especially that of the consonant /ʃ/, on labial parameters for French /i,y/. Five parameters were analysed: the height, width and area of lip opening, the distance between the corners of the mouth, as well as lip protrusion. Ten speakers uttered a corpus made up of isolated vowels, syllables and logatoms. A special procedure has been designed to evaluate lip opening contours. Results showed that the carry-over effect of the consonant /ʃ/ can impede the opposition between /i/ and /y/ in the protrusion dimension, depending upon speakers.

Index Terms: labial coarticulation, /i,y/, protrusion, lip opening, lip contour.

1. Introduction

The objective of this paper was to investigate the (labial) distinction between French vowels /i,y/, two vowels radically opposed with respect to lip positions, in “adverse” contexts. The opposition between /i/ and /y/ is phonological in French. We use a special method, using a semi-automatic tracking algorithm, to estimate the contours of lip opening.

Previous works on labiality often focused on anticipatory coarticulation, with the aim of determining the number of segments or the time interval affected by the anticipation of lip protrusion (see Farnetani and Recasens [1] for a review about coarticulation). Carry-over effects are generally considered to be less important with respect to their spatial effects or temporal expansion than anticipatory ones, and essentially due to articulatory constraints. Nevertheless, a study about the carry-over influence of sibilants on French vowels /i,y,e,ø/ [2] has shown that consonants /ʃ,z/ highly influence protrusion values for the vowel /i/ so that /i/ could be considered as “half-protruded”. In this vein, we chose to investigate the carry-over influence of the phonetic context on the (labial) opposition between /i/ and /y/. To that purpose, we chose a series of sequences, where /i/ and /y/ were preceded by /ʃ/ and three other alveo-dental consonants in monosyllables (Ci and Cy) and logatoms (iCy and yCi sequences). Five parameters were investigated, including lip protrusion and three parameters for lip opening. This investigation was preceded by an analysis of lip positions for four isolated vowels (/i,y,o,a/). We compared our results with the pioneering study of Fromkin [3] for American English vowels.

2. Data acquisition and measurements

2.1. Acquisition and corpus

Data were acquired with a 3D acquisition infrastructure which uses two cameras, a PC and painted markers which do not

change speech articulation [4]. The acquisition rate was 118, 4 frames/s.

A corpus, made up of speech sequences of various lengths, has been collected to study lip positions for French speech sounds. It was uttered by ten French speakers, five males and five females, in order to analyse inter-speaker variability and consistency. All speech sequences, with the exception of sentences, have been spoken in a carrier sentence.

We used a subset of this corpus: the four isolated vowels /i,y,a,o/, CV syllables (/i/ and /y/ preceded by /t,d,s,ʃ/) and dissymmetric VCV logatoms where V was either /i/ or /y/ and the consonant was one of /t,d,s,ʃ/, i.e. /ity, idy, isy, iʃy/ and /yti,ydi,yisi,yʃi/.

2.2. Labial parameters

We measured 5 parameters (see Figure 1):

- the width of horizontal lip opening, parameter A, i.e. the distance between the right and left points at the junction between both lips, measured from the inner lip contour
- the height of vertical lip opening, parameter B, measured from the inner lip contour
- the distance between mouth corners (outer contour), parameter C, measured between two fixed paint markers
- the area of lip opening, parameter S, delimited by the inner contour
- lip protrusion, parameter P, which evaluated the protrusion of both lips (profile view) with respect to a fixed point on the head (see [4] for details).

A special procedure has been designed to determine the inner lip contour, from which the three parameters characterizing lip opening (A, B and S) are derived. This procedure is explained below (section 2.4).

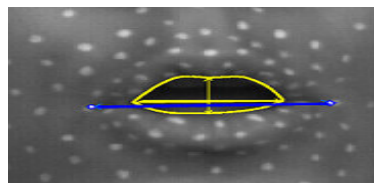


Figure 1. Labial parameters. The inner lip contour, the horizontal and vertical lip opening (parameters A and B) are represented in yellow. Parameter S is the area delimited by the inner contour. The distance between mouth corners (par. C) is represented in blue.

Parameters A and C could be seen as redundant since they are both dependent upon lip stretching. Nevertheless, although there are complex relations between most parameters, we can consider that the distance between the corners is more directly linked to lip stretching/retraction, whereas the width of opening depends also on lip height (see section 3). Since it has been shown that the width of lip opening allows a good distinction between rounded and unrounded isolated vowels even without speaker normalisation [3], whereas the distance

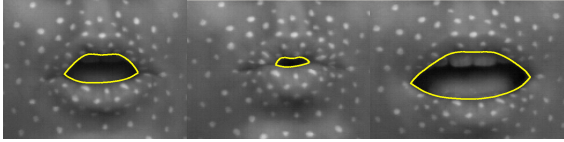


Figure 2. *Examples of contours*

between the corners exhibits relatively small fluctuations, and cannot be analysed without speaker normalisation, we decided to investigate both parameters.

Normalization. The data have been analyzed with and without speaker normalization. Indeed, labial parameters, especially the protrusion parameter and the distance between the corners, depend upon speakers' anatomic characteristics. To normalise the data, we used the formula: $(X-\mu)/\sigma$; where X was the parameter value to be normalised, μ the mean and σ the standard deviation of the parameter calculated for each speaker on all the speech segments of two phonetically balanced sentences.

2.3. Evaluation of lip opening contours

Our objective was to evaluate lip opening contours from sequences of images recorded for different speakers. There are mainly two approaches. In the first case, tracking is based on snakes [5] without any shape constraint. But conflicts appear when there are several contours. One solution used to go around this problem is to paint lips with a colour which increases the contrast with skin. The second approach is deformation models like AAM (Active Appearance Model) [6] which adapt themselves to images. However an additional stage is required to derive contours.

In our case teeth and paint markers used for the 3D stereovision create conflicts. The snake technique is not efficient and we decided to use a semi-automatic tracking algorithm proposed by Fontecave and Berthommier [7].

Beforehand, the head movement is removed for all the images, the first image of the sequence being chosen as a reference. This tracking algorithm consists in outlining by hand the inner contour of the lips for a number of key images. Technically, contours are represented in the form of B-splines curves with a fix number of control points. Then, the idea is to find the three closest key images for other images so as to interpolate the lip contour from those of key images. The new contour, i.e. the control points of the B-spline, is the weighted average (using distances) of the contours of the three closest images. The distance between images is calculated by using DCT (Discrete Cosine Transform) coefficients over a rectangular region of interest containing lips. Two contours are used: the inner contour of the upper lip and that of the lower lip.

A first visual examination is carried out to find out images which cannot be indexed correctly because they are too far from key images. These images are thus added as key images. The ratio of key images used is about 3% for each speaker. Figure 2 displays examples of some contours.

3. Lip positions for isolated vowels

We examine here the values of the five labial parameters for four isolated vowels /i,a,y,o/ pronounced by ten speakers (Table 1, Figures 3 and 4). Note that the protrusion values correspond to a distance between a fixed point on the head. We took into consideration the values at the centre of each

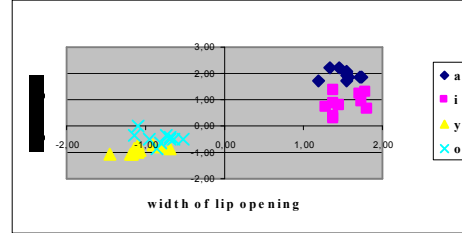


Figure 3. *Relation between the width of horizontal lip opening and lip height for isolated vowels. Normalized data.*

vowel. Let us first consider the characteristics of each class of vowels. All the unrounded vowels /i,a/ were clearly distinguished from all the rounded ones /y,o/ in each of the five parameters, without any normalisation for the parameters A, B (width and height of lip opening) as well as S –lip opening area- and with speaker normalisation for the protrusion parameter and the distance between the corners (C). The only exception was found for the vowels /i/ and /o/ uttered by one speaker, which exhibited very similar values for lip height (they were separated by less than one mm).

	A	B	S (area)	P	P Norm	C	C Norm
a	41.0	14.9	492	99.2	1.55	51.0	0.57
	3.4	2.9	118	5.5	0.32	3.2	0.31
i	41.1	10.4	326	99.5	1.46	53.6	1.42
	4.7	2.1	86	5.3	0.22	3.3	0.40
o	15.2	5.1	58	105.8	0.54	44.9	-1.47
	2.7	1.8	24	5.8	0.26	3.7	0.38
y	12.3	3.3	33	106.5	0.76	46.4	-1.01
	2.0	0.9	13	5.9	0.13	2.9	0.35

Table 1. *Means and standard deviations for each labial parameter, in mm, or mm² for S (see section 2.3). Isolated vowels. Normalized values are also given for parameters particularly dependent upon speakers' anatomy (P and C).*

Let us consider now the differences between vowels of a same class. For each speaker, lip height and lip area were more important for /a/ than /i/ whereas the distance between the corners was slightly larger for /i/ than /a/. All these differences were highly significant (Student's t-test, paired samples). These observations can be generalized to all the exemplars of /i/ and /a/ after speaker normalization.

Concerning the rounded vowels /y,o/, we observed that, for each speaker, /y/ had the lower lip height (only one exception) and the lower lip area. These differences between /y/ and /o/ were statistically significant (Student's t-test, paired samples).

Interestingly, the width of lip opening has approximately the same values for /i/ and /a/ whereas, for each speaker, the

distance between the corners was always higher for /i/ than for /a/. This larger distance for /i/ corresponds well to the classical view of /i/ as being the most stretched vowel. The distance between the corners is highly dependent on lip stretching (for unrounded vowels) whereas the width of lip opening depends not only on lip stretching but also on the height of lip opening. Indeed, when lip height decreases, the surface of the mouth closed by the junction of the upper and lower lips at its right and left hand sides tends to increase. This probably explains why the vowels /a/, with higher values for lip height, and the vowels /i/, with higher values for the distance between the corners, exhibited approximately the same values for parameter A.

As expected, we found important correlations between protrusion and the distance between the corners (parameter A).

Our results are in good agreement with those of Fromkin [3], as far as we can compare our sets of vowels, and those of Descout *et al.* [2]. In particular, Fromkin found that the width of lip opening (parameter A) was a very efficient parameter to distinguish (back) rounded from (front) unrounded vowels, and did not differ significantly for front unrounded vowels. Her data showed that lip height was generally higher for /ae/ (a vowel close to the French /a/) than for /i/. In this dimension, the values of many exemplars of /i/ and /o/ overlap, whereas this phenomenon appeared for only one pair of vowels in our study. Nevertheless, the distribution of vowels in the lip height dimension show similar trends: in both studies, lip height exhibits the lowest values for close rounded (/y/ or /u/) and the highest for (front) open vowels (/a/ or /ae/).

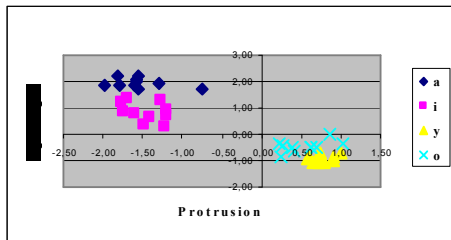


Figure 4. Relation between protrusion (x-axis) and lip height (y-axis). Isolated vowels. Normalized data (section 2.2).

Finally, this study confirms that the isolated rounded vowels /y,o/, when compared to the isolated unrounded vowels /i,a/, have the smaller values for all parameters concerning lip opening (A, B, S), and the higher values for lip protrusion. Most combinations of two of these parameters provide a very clear separation between both groups of vowels (see Figure 3 and 4). The object of the following experiment is to verify whether the opposition between /i/ and /y/ is affected in “adverse” contexts.

4. Vowels in context

4.1. Protrusion

Let us first analyse the influence of the phonetic context on vowels. For each vowel exemplar, the value taken into account for each parameter was averaged across all the frames of this

exemplar. Table 2 shows that the protrusion values for /i/ were higher when the vowel was preceded by /f/ than by the other consonants (/s,t,d/). This effect was observed for each speaker for monosyllables (Figure 5).

Contexts	P	H	Contexts	P	H
i	99.5	10.4	y	106.5	3.3
Ci	101	10.7	Cy	106	3.4
fi	102.9	13.6	fy	106.1	3.7
si ti di	100.3	9.7	sy ty dy	105.8	3.2
yCi	101.9	10.3	iCy	106.4	3.3
yfi	102.8	11.9	ify	106.8	3.7
ysiyti ydi	101.7	9.8	isy ity idy	106.3	3.3

Table 2. Protrusion (P) and lip height (H). Values (mm) for isolated vowels, and for /i,y/ preceded by all the consonants (Ci,Cy,yCi,iCy), by /f/ and by /s,t,d/, in monosyllables and logatoms. All the values have been averaged across speakers and contexts.

For these monosyllables, the differences between the values of /i/ preceded by /f/ and those of /i/ preceded by each of the other consonants (/f/ vs. /s/, /f/ vs. /t/ and /f/ vs. /d/) were all highly significant (p being always lower than 10^{-4} , Student’s t-test, paired samples, associating the data of each speaker), as was the difference between the context /f/ and all the other consonantal contexts considered together ($p < 0.01$). For this last test, the protrusion values were normalized since the numbers of data to be compared were not equal so we could not use the “paired samples” method. In /yCi/ context (logatoms), the average value of /i/ increased significantly with respect to monosyllables (103 mm for logatoms vs. 101 mm for monosyllables Ci, $p < 0.01$). The effect of /f/ was less systematic than in monosyllables but nevertheless significant although to a lesser level (p between 0.02 and 0.04, comparisons between /f/ and each other consonantal context).

As expected, /y/ being already protruded, the influence of /f/ on this vowel was weaker than on the vowel /i/. Indeed, we remark that the mean values estimated for /y/ in various contexts (in isolation, in logatoms or in monosyllables) were all relatively similar, whether the vowel was preceded or not by /f/ (table 2). We note only two slightly significant effects (we accepted a level of 0.05), for /ifjy/ vs. /ity/ and /fy/ vs. /sy/. We do not observe any significant effect due to one of the three other consonants on the vowels /i/ and /y/.

As shown by table 2, the protrusion value for /i/ preceded by /f/ (around 103 mm for monosyllables and logatoms) was approximately at equal distance between those estimated for /i/ in isolation or in the monosyllables /si,ti,di/ (99.5 mm and 100 mm, respectively) and those observed for the vowel /y/, in any context (from 106 up to 106.8 mm). This is quite in agreement with the study by Descout *et al.* [2] which concerned the realizations of one speaker. We note also the increasing effect of the context /yCV/ on the protrusion values for /i/ (102 mm). But the influence of the phonetic context can be more drastic. Table 2 gives data which have been averaged across speakers. If we study the data of each speaker separately (Figure 5), we see that the effect of the phonetic context, especially the effect of the consonant /f/ on the vowel /i/, could drastically reduce and, for some speakers, even impede the separation between the rounded vowel /y/ and the unrounded /i/ in the protrusion dimension. For at least two speakers (2 and 6), protrusion values for /i/ and /y/ exhibited slight fluctuations as a function

of contexts, which did not impede a clear separation between both categories of vowels. As expected for these speakers, the values for the vowels in context were similar to those observed for isolated vowels. But, for most speakers, some of the values obtained for /i/ were very close to (and even higher than, in two cases) those observed for the vowel /y/. To take a concrete case: the percent of data for which the protrusion value for one exemplar of one vowel was separated from less than one mm (a value which is probably close to the precision of the estimation) to the value estimated for at least one exemplar of the other vowel reached 20% for /y/ and 10% for /i/ (respectively 40% and 20% for the five speakers concerned by this effect). As expected, the lack of differentiation between /i/ and /y/ values affected mainly the contexts /ʃi/ and in less cases /yʃi/ as well as the contexts /sy,ty,dy/.

Then results clearly showed that, for non isolated vowels and for about half of speakers, the protrusion parameter was not an efficient cue for characterizing /i/ vs. /y/.

4.2. Lip height.

For both series of contexts, monosyllables /Ci/ and logatoms /yCi/, lip height values were higher when /i/ was preceded by /ʃ/ than by any other consonant. Indeed, due to the effect of /ʃ/, the mean values rose from 9.7 to 13.6 mm (monosyllables) and from 9.8 to 11.9 mm (logatoms). This influence was observed for almost all speakers and was sometimes very impressive (see Figure 5). We reproduced the same statistical tests that those designed for the discussion about protrusion, and found that the increasing effect of /ʃ/ on lip height values with respect to values estimated in each other consonantal context was always highly significant, for monosyllables and logatoms ($p. < 0,001$, in all the cases).

The consonant /ʃ/ also increased lip height values of the vowel /y/. As expected, this influence was weaker for /y/ than it was for /i/. Indeed, the values rose from 3.2 mm (monosyllables) or 3.3 mm (logatoms) to 3.7 mm (both series) due to the effect of this consonant. Nevertheless, for monosyllables, this effect was found to be significant for all the tests undergone (/ʃ/ vs each other context and, after a normalisation of data, /ʃ/ vs all the other contexts). For the logatoms /iCy/, we note only one significant effect (/ʃ/ vs. /s/).

Unlike the effect of /ʃ/ on protrusion values, the increasing effect of this consonant on lip height values did not impede the separation between /i/ and /y/ in lip height dimension. This was quite the contrary, since the average difference between the values of these vowels increased. We can see that there was no overlap between /i/ and /y/ values, even when we considered all the speakers simultaneously and did not normalise the data (Figure 5).

Results for other parameters could not be reported in detail in this paper. But a preliminary analysis showed that parameters characterizing lip opening (the width of lip opening, and the area) ensured a clear distinction between /i/ and /y/, whereas values for rounded and unrounded vowels overlapped if we consider the distance between the corners.

5. Conclusions

This study showed that carry-over coarticulatory effects, in particular the influence of the consonant /ʃ/, can have drastic

consequences on lip protrusion for /i/, impeding, for about half of the speakers involved in this study, the distinction between /i/ and /y/ in this dimension. The (labial) opposition between these vowels was nevertheless ensured by other labial parameters, such as the height of lip opening. Results also put in light the existence of large variations among speakers' coarticulatory habits. More experiments appeared necessary to find out the various influences of consonantal contexts and transconsonantal vowels on visual cues for vowels.

6. References

- [1] Farnetani, E. and Recasens, R., "Coarticulation models in recent speech production theories" In Hardcastle and Hewlett [Ed], Coarticulation theory: data and techniques, 31-65, 1999.
- [2] Descout, R., Boë, J.-L., Abry, C., "Labialité vocalique et labialité consonantique en français" In Abry *et al.* [Ed], Labialité et Phonétique, Publications de l'Université des Langues et Lettres de Grenoble, 112-126, 1980
- [3] Fromkin, V., "Lip positions in American English vowels" Language and Speech, (7):215-225, 1964.
- [4] Robert, V., Wrobel-Dautcourt, B., Laprie, Y., Bonneau, A., "Strategies of labial coarticulation" Interspeech, 2005.
- [5] Kass, M., Witkin, A. and Terzopoulos X., "Snakes: Active contour models", International Journal of Computer Vision, 1(4): 321-331, 1987.
- [6] Cootes, T.F., Edwards, G.J. and Taylor, C.J., "Active Appearance Models", IEEE, Transactions on Pattern Analysis and Machine Intelligence, 23(6):681-685, 2001
- [7] Fontecave J. and Berthommier F., " Semi-Automatic Extraction of Vocal Tract Movements from Cineradiographic Data" In Interspeech, Pittsburgh, 2006.

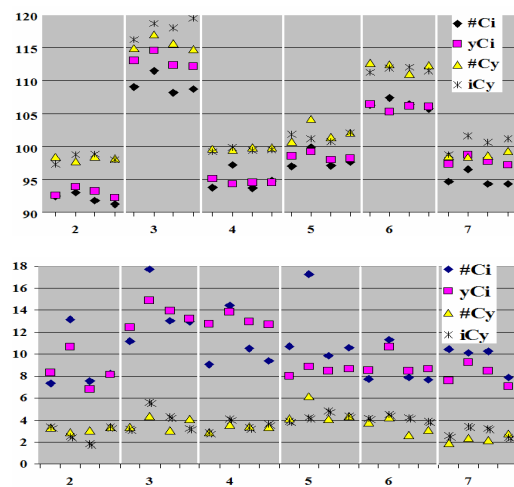


Figure 5 Protrusion (top) and lip height (bottom) values for /i,y/ preceded by /s,f,t,d/ (in that order in the figure) in monosyllables (Ci and Cy) and logatoms (yCi and iCy), for six representative speakers. The y-axis represented the values in mm for each parameter (no normalisation). Data of each speaker have been isolated by a vertical line.