Three-dimensional coarticulatory effects on spatial characteristics of labial targets for ‘VpV and ‘VfV Italian consonant in asymmetric bisyllables

Claudio Zmarich and Emanuela Magno Caldognetto

Istituto di Scienze e Tecnologie della Cognizione del C.N.R., Sez. di Fonetica e Dialettologia, Padova, Italy
E-mail: claudio.zmarich@csrf.pd.cnr.it, emanuela.magno@ csrf.pd.cnr.it

ABSTRACT

We describe the 3D spatial characteristics of upper and lower lip in the articulatory production of /p/ and /f/ consonantal targets in ‘VCV pseudo-words (V: /a i u/). Results will be relevant to the discussion of articulatory production and coarticulation theories, to the cross-linguistic comparisons, to the visual articulatory perception and to the bimodal speech synthesis.

1. INTRODUCTION

In Italian there are only two classes of consonants characterized by lips as active articulators: bilabial stops (oral or nasal: /p, b, m/) and labiodental fricatives (/f, v/). The aim of this research is to describe the 3D spatial characteristics of upper and lower lip movements in the articulatory production of /p/ and /f/ in ‘VCV pseudo-words. These phonetic sequences are stressed on the penultimate syllable as most of the Italian words, and are characterized by a vocalic symmetric and asymmetric context (V: /a i u/).

Specifically, we are interested in defining the spatial coordination of upper lip and lower lip movements for the bilabial closure in /p/ *vs* the relative independence of the same articulators in the production of labiodental constriction in /f/, under the constraints imposed by the coarticulatory influences of the surrounding vowels. For a recent review on labial coarticulation, see [1].

2. METHOD

In order to collect articulatory data, an automatic optotracking movement analyser for 3D kinematics data acquisition (ELITE) was used, which allows also a synchronous recording of the acoustic signal [2]. This system ensures high accuracy (100 Hz sampling rate, maximal error of 0.08 mm for a 20x20x20 cube) and minimum discomfort to the subject because it tracks the infrared light reflected by small (2 mm diameter), passive markers, which are attached, in this case, to the central point of the vermilion border of upper and lower lip and at the corners of the lips. The following 5 articulatory parameters were computed:

-Vertical movement of the upper lip (UL) and lower lip (LL), calculated as the distance between the markers placed on the central point of either the upper lip or lower lip and the transversal plane passing through the tip of the nose and the markers on the ear lobes (a virtual plane functioning as a reference). Here LL has to be interpreted as the combined lower lip plus jaw structure. No further differentiation will be made between the two articulators. These parameters correlate with the HIGH/LOW phonetic dimension.
-Horizontal variation of the labial width (LW, the distance between the left and right corners of the lips) which correlates with the ROUNDED/UNROUNDED phonetic dimension.
-Anterior/posterior movement (protrusion) of upper lip (UP) and lower lip (LP), calculated as the distance between the marker placed on the central points of either the upper and lower lip and the frontal plane containing the line crossing the markers placed on the lobes of the ears (a virtual plane functioning as a reference). These parameters correlates with the feature PROTRUDED/RETRACTED.

The 18 (3x2x3) different sequences /’VpV/ and /’VfV/ were preceded by the phone /b/, so as to form a series of pseudo-words, some of them not really attested, but all possible Italian words. These pseudo-words were inserted into a carrier phrase (“declama ‘bVpV/’bVfV magnificamente”) and randomly repeated four times each in four different recording sessions by one subject, university student and talker of northern Italian. At the beginning of each session, the position of the lips at rest was recorded (lips neither open, nor pressed).

The articulatory signal was segmented on the basis of the acoustic speech signal. The time course of the movement and the velocity characteristics of the articulatory parameters were displayed by means of a custom made, modified version of the software MAVIS 1.5, created by M. Tiede (ATR Human Information Processing Labs Inc.). The fig. 1 examplifies a visualization of the acoustic and kinematic signals considered in this study. Finding a single peak characterizing the vocalic and consonantal targets by means of segmentation of the articulatory signal for each articulatory parameter was not simple. The following is the procedure we followed: every single time the portion of the articulatory signal within the acoustic boundary of a target segment (vowel 1, consonant, vowel 2) had a single

ISBN 1-876346-48-5 © 2003 UAB

peak, that peak was taken as the target of that parameter for that segment. Finding a single pick was sometimes impossible, and this situation occurred mainly in two general cases: (1) when the movement evidenced an uninterrupted glide, so that no peak was discernible, or (2) when the movement presented more than a single peak.

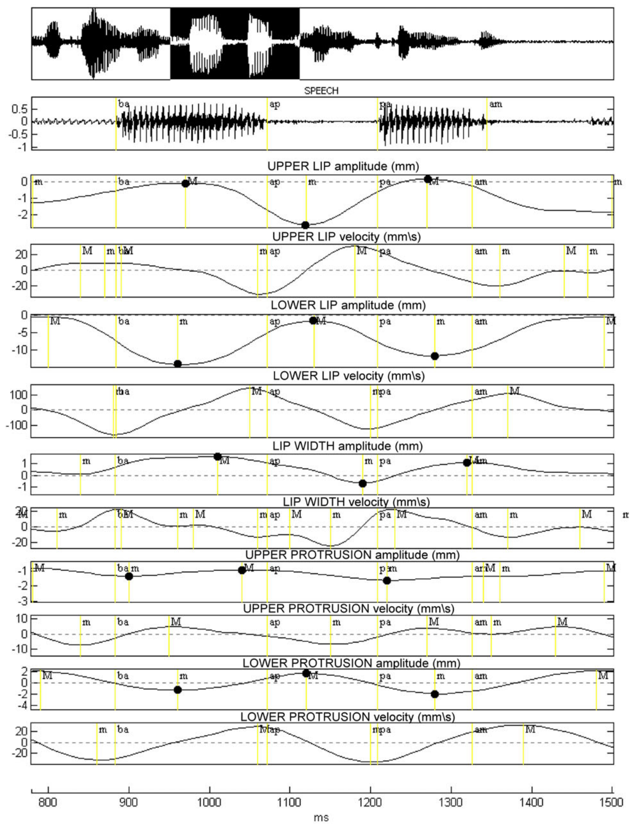

Fig.1. Time courses of the syncronous acoustic and kinematic (amplitude and velocity) signals for /'bapa/. The V1, C, V2 targets for each articulator are indicated by bold circles. The scales for UL, LL and LW are inverted.

In the first case, we tried to find a solution by looking at the velocity curve: when this curve evidenced a second peak having lower value and close to the zero velocity line, we considered the value of the amplitude curve at that temporal point as representing the target. In the second case, a number of velocity peaks aligned around the zero curve: most of the times, this situation characterizes an articulatory *plateau* (a stillness protracted for a certain amount of time). This *plateau* was demarcated at the first and the last crossing of the velocity curve on the zero-axis. Whenever both criteria failed to apply, it was agreed to identify the target point as the articulatory sample lying in the middle of the acoustic segment. All the values of the target points were finally normalized with respect to position of the lips at rest.

3. RESULTS

The number of consonantal targets are 355 (71 sequences X 5 articulatory parameters). The targets easily identified by a clear single peak are 315 (88.73%), those reconstructed on the basis of the velocity curve are 21 (4.61%) and those postulated by convention are 19 (4.17%).

Table 1 presents the spatial values assumed by each articulatory parameter at the target position for V1 and V2, in the 'VCV sequences. Since our aim is to describe the coarticulatory processes at the consonantal targets, the lack of space force us not to analyse them further on.

/p/	UL		LL		LW		UP		LP	
a	-0.1	0.2	13.8	10.4	-2.0	-0.7	-0.7	-0.9	-0.9	-1.9
i	-0.1	0.1	9.1	8.3	0.4	0.5	-0.8	-0.7	-1.0	-1.6
u	-0.1	-0.4	5.6	5.4	-1.6	-1.1	2.9	2.2	0.3	0.4
/f/	UL		LL		LW		UP		LP	
a	-0.1	-0.1	11.5	8.3	-1.9	-1.7	-0.7	-0.1	-1.0	-1.6
i	-0.3	-0.2	8.2	7.0	-0.1	-0.3	-0.5	-0.2	-1.8	-1.9
u	-0.4	0.1	4.9	5.6	-0.8	-1.9	1.9	2.0	-0.5	0.9

Table 1. Spatial values (mm), assumed by each articulatory parameter at the target position for V1 and V2, in the 'VpV (first three raws) and 'VfV (last three raws) sequences. The values are normalized with respect to the position of the lips at rest.

In order to ascertain the influence of V1 and V2 on the spatial position of the consonant, a series of separate one-way ANOVA analyses was carried out on the spatial values assumed by each articulatory parameter at the consonantal target (dependent measure, /p/ and /f/ considered separately), as a function of the identity of V2 (when V1 was equal to /a/ or /i/ or /u/), or as a function of the identity of V1 (when V2 was equal to /a/ or /i/ or /u/). More specifically, we selected sistematically a different V1 to test the effects of the V2 Factor, and a different V2 to test the effects of the V1 Factor. The following figures (2, 3), plot the values assumed by each articulatory parameter as a function of V1 and V2 (brackets represent ± 1 standard error). In interpreting the results, the reader must be aware that V1 was stressed, V2 was unstressed, and that some articulatory parameter could be affected by the different prosodic status of the vowels.

3.1 CONSONANT /p/

Fig.2 shows the values assumed by UL as a function of the different identities of the contextual vowels. As a general observation, since all the targets exhibit positive values, this means that UL moved below its rest position. A series of one-way ANOVA analyses (see above for details) did not produce any statistically significant result.

The same figure presents the values assumed by LL as a function of the different identities of the contextual vowels.

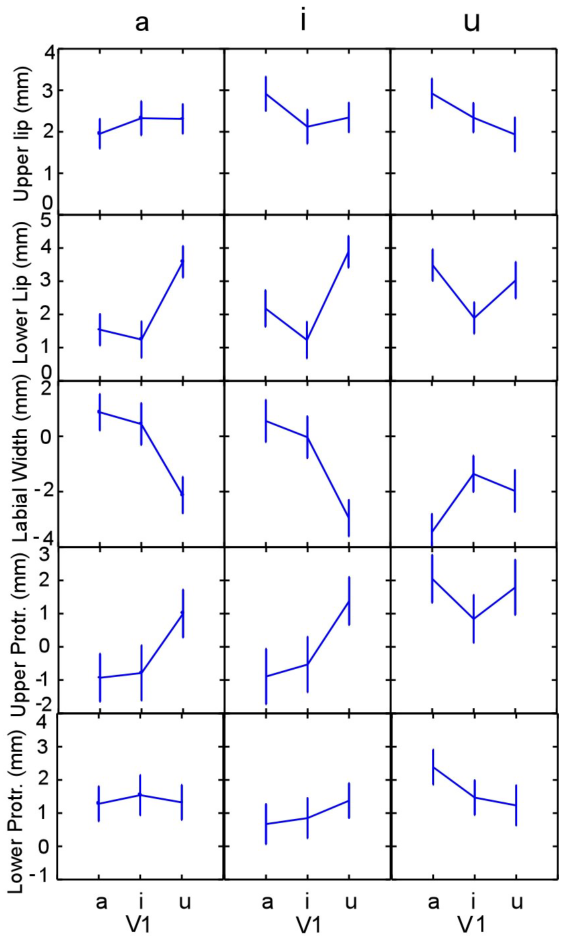

Fig. 2. Values for each articulatory parameter at the consonantal target position for /p/ as a function of V1 (bottom) and V2 (top)

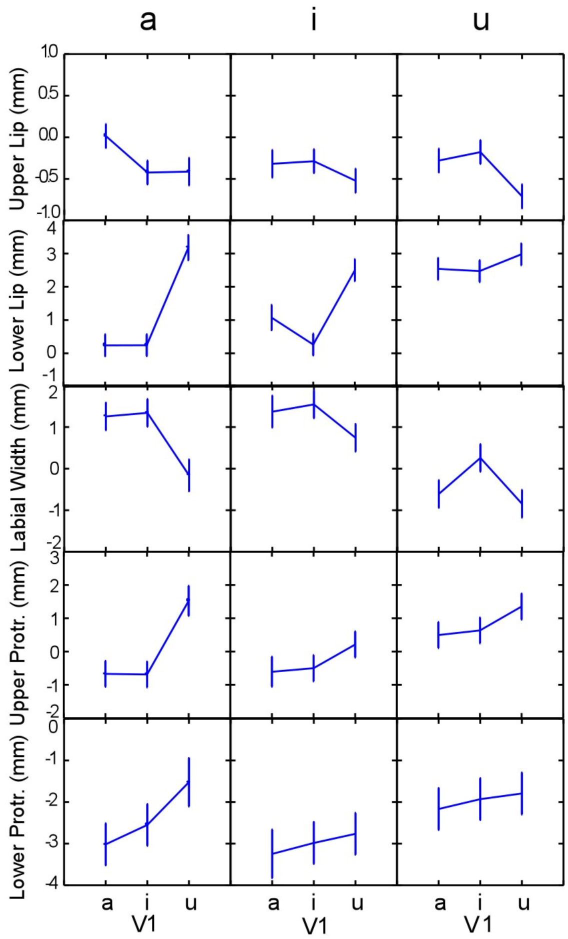

Fig. 3. Values for each articulatory parameter at the consonantal target position for /f/ as a function of V1 (bottom) and V2 (top)

A series of one-way ANOVA analyses (see above for details) allowed us to find two significant results: when V2 is /a/, the consonantal targets are affected by V1 ($p<.010$, $F_{2,8}$ 8.806). The *Bonferroni* post-hoc test reveals that /apa/ and /ipa/ contrast with /upa/. When V2 is /i/, the consonantal targets are affected by V1, ($p<.000$, $F_{2,8}$ 30.808). The *Bonferroni* post-hoc test reveals that /api/ and /ipi/ contrast with /upi/.

The same fig.2 presents the values assumed by LW according to the different identities of the contextual vowels. A series of one-way ANOVA analyses produced four significant results: when V1 is /a/, the consonantal targets are affected by V2 ($p<.034$, $F_{2,8}$ 5.324). The *Bonferroni* post-hoc test reveals that /apa/ contrasts with /apu/. When V1 is /i/, the consonantal targets are affected by V2 ($p<.007$, $F_{2,8}$ 1.946). The post-hoc test reveals that /ipa/ and /ipi/ contrast with /ipu/. When V2 is /a/, the consonantal targets are affected by V1 ($p<.000$, $F_{2,8}$ 29.016). The post-hoc test reveals that /apa/ and /ipa/ contrast with /upa/. When V2 is /i/, the consonantal targets are affected by V1 ($p<.001$, $F_{2,8}$ 19.999). The post-hoc test reveals that /api/ and /ipi/ contrast with /upi/. Obviously, every time that V is /u/, regardless of its position, *all* the targets become rounded (i.e. LW decreases).

In the same fig.2, the values assumed by UP as a function of the different identities of the contextual vowels are shown. As a general observation, any time that V is /u/, regardless of its position, *all* the targets become protruded. A series of one-way ANOVA analyses (see above for details), did not produce any statistically significant result.

The same fig. 2 presents the values assumed by LP as a function of the different identities of the contextual vowels. A series of one-way ANOVA analyses (see above for details), did not produce any statistically significant results. All the items show a light degree of protrusion at the target.

3.2 CONSONANT /f/

The fig. 3 shows the values assumed by UL as a function of the different identities of the contextual vowels. A series of one-way ANOVA analyses (see above for details), did not produce a statistically significant result. Almost all the scores show a light degree of rising (above the rest position) at the consonantal target.

The same fig.3 presents the values assumed by LL as a function of the different identities of the contextual vowels. A series of one-way ANOVA analyses (see above for details) allowed us to find four significant results: when V1 is /a/, the consonantal targets are affected by V2 ($p<.005$, $F_{2,8}$ 11.296). The *Bonferroni* post-hoc

ISBN 1-876346-48-5 © 2003 UAB

test reveals that /afa/ and /afi/ contrast with /afu/. When V1 is /i/, the consonantal targets are affected by V2 ($p<.002$, $F_{2,8}$ 13.147). The *Bonferroni* post-hoc test reveals that /ifa/ and /ifi/ contrast with /ifu/. When V2 is /a/, the consonantal targets are affected by V1 ($p<.000$, $F_{2,8}$ 91.233). The *Bonferroni* post-hoc test reveals that /afa/ and /ifa/ contrast with /ufa/. When V2 is /i/, the consonantal targets are affected by V1 ($p<.002$, $F_{2,8}$ 15.989). The *Bonferroni* post-hoc test reveals that /afi/ and /ifi/ contrast with /ufi/. The contrasts obey to the same logic: every time V is /u/, independently from position, *all* the targets evidence lower values than at the rest position.

The same fig.3 illustrates the values assumed by LW as a function of the different identities of the contextual vowels. A series of one-way ANOVA analyses (see above for details) allowed us to find three significant results: when V1 is /a/, the consonantal targets are affected by V2 ($p<.001$, $F_{2,8}$ 20.798). The *Bonferroni* post-hoc test reveals that /afa/ and /afi/ contrast with /afu/. When V1 is /i/, the consonantal targets are affected by V2 ($p<.015$, $F_{2,8}$ 7.020). The *Bonferroni* post-hoc test reveals that /ifa/ and /ifi/ contrast with /ifu/. When V2 is /a/, the consonantal targets are affected by V1 ($p<.000$, $F_{2,8}$ 44.123). The *Bonferroni* post-hoc test reveals that /afa/ and /ifa/ contrast with /ufa/. The logic behind the contrasts is that any time that V is /u/, regardless of its position, *all* the targets become a little bit more rounded than at the rest position (i.e. labial width decreases).

The same fig.3 presents the values assumed by UP as a function of the different identities of the contextual vowels. A series of one-way ANOVA analyses (see above for details) allowed us to find two significant results: when V1 is /i/, the consonantal targets are affected by V2 ($p<.026$, $F_{2,8}$ 5.651). The *Bonferroni* post-hoc test reveals that /ifa/ contrasts with /ufa/. When V2 is /a/, the consonantal targets are affected by V1 ($p<.004$, $F_{2,8}$ 12.079). The *Bonferroni* post-hoc test reveals that /afi/ and /ifi/ contrast with /ufi/. Taking all of the results into consideration, we can say that any time V is /u/, the targets tend to become more protruded than at the rest position, although at different degrees.

The same fig.3 presents the values assumed by LP as a function of the different identities of the contextual vowels. A series of one-way ANOVA analyses (see above for details) did not produce any significant result. Lower protrusion is always reduced with respect to the rest position.

Fig 4 shows the pairwise correlations of the values assumed by each articulator for /p/ target at the occlusion place. The statistically significant correlations are as follows: LL and LW ($r=-.866$, $p<.000$), LL and UP ($r=.762$, $p<.000$), UP and LW ($r=-.822$, $p<.000$), and UP and LP ($r=.630$, $p<.001$). Fig. 5 presents the pairwise correlations of the values assumed by each articulator for the /f/ target at the constriction place. LL and LW ($r=.830$, $p<.000$), LL and UP ($r=.720$, $p<.000$), UP and LW ($r=-.676$, $p<.000$), and UP and LP ($r=.601$, $p<.002$) are the statistically significant correlations.

Considering all of the results of the correlational analyses, we can say that, whenever LL realizes its consonantal targets for /p/ and /f/ at progressively lower positions, then LW reduces, while UP increases. This last parameter is also correlated directly to LP, and inversely with LW.

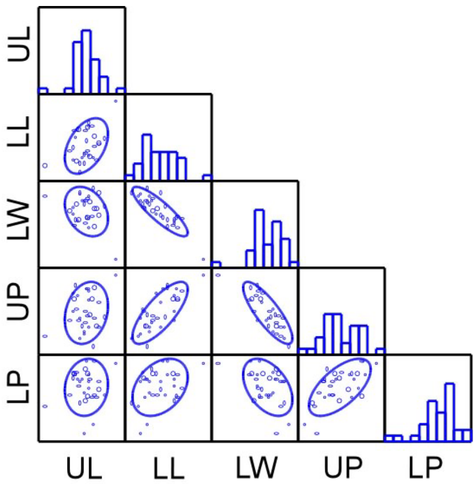

Fig. 4. Correlations among all the articulators for /p/

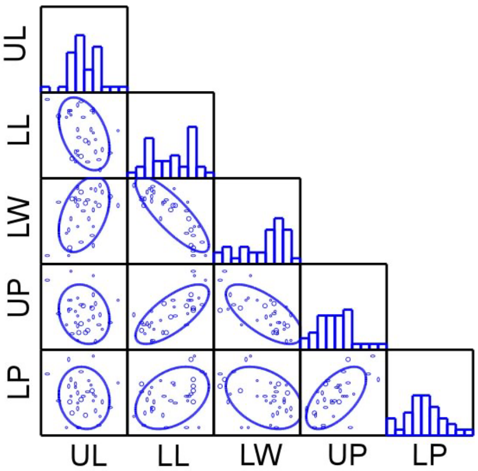

Fig. 5. Correlations among all the articulators for /f/

4. CONCLUSIONS

This research describes the 3D spatial positions assumed by the upper and lower lips at the consonantal targets for /p/ and /f/, in the context of /a i u/ surrounding vowels. Among others, the following results are of some values: the /u/ context exerts almost always a strong coarticulatory influence on the consonantal targets, by making the lips more rounded and more protruded (expecially the upper lip) than in the /a i/ contexts. The differences between the consonants are implemented mainly by LP (retracted for /f/) and UL (raised for /f/).

REFERENCES

[1] E. Farnetani, "Labial coarticulation", in *Coarticulation. Theory, Data and Techniques*, W. J. Hardcastle and N. Hewlett, Eds., pp. 144-163. Cambridge: Cambridge University Press, 1999.

[2] E. Magno Caldognetto, C. Zmarich and P. Cosi, "Statistical definition of visual information for Italian vowels and consonants", in *Proceedings of AVSP '98*, D. Burnham, J. Robert-Ribes and E. Vatikiotis-Bateson, Eds., Terrigal (AUS): pp. 135-140, 1998.