HAL
open science

Speech timing organization for the phonological length contrast in Italian consonants

Claudio Zmarich, Barbara Gili-Fivela, Pascal Perrier, Christophe Savariaux, Graziano Tisato

To cite this version:

Claudio Zmarich, Barbara Gili-Fivela, Pascal Perrier, Christophe Savariaux, Graziano Tisato. Speech timing organization for the phonological length contrast in Italian consonants. Interspeech 2011 - 12th Annual Conference of the International Speech Communication Association, Aug 2011, Florence, Italy. pp.401-404. hal-00619254

HAL Id: hal-00619254

https://hal.science/hal-00619254v1

Submitted on 5 Sep 2011

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.

Speech timing organization for the phonological length contrast in Italian consonants

Claudio Zmarich [1], Barbara Gili Fivela [2], Pascal Perrier [3], Christophe Savariaux [3], Graziano Tisato[1]

[1] CNR-Institute of Cognitive Sciences and Technologies, Padova, Italy
[2] Salento University, Lecce, Italy
[3] DPC/GIPSA-lab, Grenoble-INP & CNRS, Grenoble, France

claudio.zmarich@pd.istc.cnr.it, barbara.gili@unisalento.it, Pascal.Perrier@gipsa-lab.grenoble-inp.fr, christophe.savariaux@gipsa-lab.grenoble-inp.fr, graziano.tisato@pd.istc.cnr.it

Abstract

In Italian, length contrast is exploited in the consonant system. Previous articulatory studies have focused on the temporal organization of gestures in Italian geminates and on the lower lip kinematics of the singleton/geminate distinction, and have showed that the time interval between the nuclei of two successive syllables does not depend on the number of intervening consonants (Öhman's Vowel-to-Vowel model) . In this paper, data on lip and tongue gestures from four Italian subjects saying "mima" and "mimma" at fast and comfortable rate of delivery are discussed in order to directly test the validity of the Öhman's model for the gestural organization of Italian geminate consonants.

Index Terms: speech timing, consonant gemination, kinematic analysis

1. Introduction

Geminate consonants are at the core of debates that focus on two related issues, namely their underlying phonological structure and their timing in relation with the surrounding vowels. Regarding their phonological structure, there is a wide agreement that in Italian they correspond to a combination of two shorter, strongly connected, segments rather than to a unique and long segment. Consistently, we have found kinematic evidence [1,2] supporting that acoustic time patterns in VCV sequence, in which C is a geminate, are much more similar to corresponding VCV patterns in which C is an heterosyllabic cluster than to VCV patterns in which C is a singleton. This result supports the view that Italian geminates correspond to two identical segments belonging to different syllables [3].

As far as the processes underlying their timing is concerned, there are two main classes of phonetic hypotheses: (1) their longer duration (as compared to singleton, see [4]) is intrinsically specified and it is associated with a global reorganization of the timing of the surrounding vowels; (2) their longer duration results from more macroscopic motor control strategies, and their impact on the surrounding vowels is a secondary effect of this control. Browman & Goldstein's c-center model for syllable production [5] belongs to the first class of hypotheses, and it suggests that the initial consonant and the vowel are produced in phase, while the consonant in coda and the vowel are in anti-phase. Öhman's model [6] suggests that VCV sequences are produced on a V-V basis, which determines the global timing of the sequence, perturbed locally by the production of the consonant. This model is at the basis of Fowler's model [7], reframed by Smith [8] in gestural terms. However, Smith leaves unaffected the timing implications of the Öhman's model for the VCV sequence. This model would induce that, as compared to singletons, the lengthening of the geminates would be realized within the VCV sequence by an anticipation in the first vowel of the consonant closing gesture and a postponing in the second vowel of the consonant opening gesture. These observations refer also to the first class of hypotheses. On the other hand, the virtual target hypothesis proposed by Löfqvist [9] suggests that the main factor inducing the lengthening of the geminates is a change in the motor commands which specifies a higher virtual target for the main articulator associated with the production of the consonant. In some cases this motor control change is combined with changes in the specified consonant duration [10]. Thus, this hypothesis belongs to the second class. In a former study [2, 11], based on the analysis of acoustic and articulatory data, we found for the Italian bilabial geminates (in "mima" vs. "mimma") little support for the virtual target hypotheses (the velocity timing of the closing gesture did not match the predictions of this theory) and some trends in favour of Smith's hypothesis in the fact that the closing gesture was clearly anticipated in the preceding vowel. However, we found no clear evidence in support of the hypothesis of a constant duration of the V-V sequence across singletons and geminates. Moreover, only data from two subjects were analyzed in that paper, and only a limited number of measures considered to be critical for the evaluation of Smith's hypothesis were performed.

In this paper, a more extensive and accurate evaluation of Smith's hypothesis is provided, thanks to the analysis of data from four Italian subjects recorded at both normal and fast speaking rate. In fact, higher speech rate often reduces the phonetic contrast of two structures in phonological opposition, and reduces the number of alternative articulatory strategies

2. Corpus and method

2.1. Corpus

The experimental corpus was composed by the words 'mima' (from the verb "to mime") and 'mimma' (proper name). Target words were inserted in the carrier sentence *richiama mim(m)a malamente* "s/he calls back mim(m)a badly"). Speakers were asked to read each sentence aloud at least ten times both at a natural speech rate and, immediately afterwards, at a faster speech rate. The sentences were randomly distributed among a wider corpus of stimuli, involving other consonant opposition for gemination (see [10]). All the subjects were PhD students. A female speaker of a north-eastern variety of Italian (AG), a female speaker (BG) and a male speaker (MP) of a north-western variety, and a

female speaker of a variety from mid Italy (FC) were recorded. The audio signal was acquired by means of a DAT recorder, while the kinematic data were collected using the 2D EMA system at GIPSA-lab (formerly ICP) in Grenoble. In this system sensors are all glued in the mid-sagittal plane of the head. Two reference sensors were glued on the nose bridge and the upper incisors, two were placed on the upper and lower lips, and four were glued of the tongue in the range of about 1 cm to 5 cm from the tongue tip. The data discussed here only relate to measurements of Lip Aperture (i.e., the vertical distance between the two lips) to study the consonant gesture and to measurements of tongue dorsum (sensor located at around 3.5 cm from the tongue tip) for studying the vowel-to-vowel gesture.

2.2. Auditory test

Times Consonant duration is distinctive in (standard) Italian, but various dialects spoken in northern Italy do not use geminates and do not exploit the geminate/singleton contrast, especially at fast rate and/or informal speech. First of all then, we decided to verify, by means of a perception test, whether our speakers had adequately produced geminate consonants. Secondly, we sought to look at acoustic/articulatory correlates of clear geminate *vs.* singleton contrasts.

A perception test was run in order to verify whether our speakers had adequately produced geminate consonants, especially at fast speech rate, and to identify and exclude (from statistics analysis) stimuli that were ambiguously realized with respect to the singleton vs. geminate contrast (i.e. we want to look at correlates of clearly produced and perceived geminate and singleton consonants).

A selection of the stimuli produced by the four subjects was used for the perception test. For each subject, two repetitions realized at a normal speech rate and all repetitions realized at fast rate were selected (for 160 “mima/mimma” stimuli, 96 were selected for the perception test). Five subjects from the northern part of Italy (Torino) and five subjects from the southern one (Lecce-Taranto) took part in the perception test. They listened to audio files containing target words in carrier sentences and judged whether the target words included singleton or geminate consonants. Test results showed that the selection of normal rate “mima/mimma” stimuli was basically correctly recognized (error rate is around 4%), while fast rate stimuli were misidentified in around 22% of cases. Stimuli that were wrongly recognized by at least 4 subjects were considered ambiguous and not taken into account for acoustic and kinematic analysis. The comparison of these “no contrast“ cases (mainly observable at a faster speaking rate) with the “clear contrast” was left to a follow-up study.

2.3. Measurements

The duration of the consonant and both the preceding and the following vowels was measured by manually segmenting and labelling the acoustic signal in PRAAT. Boundaries within the VC(C)V sequence were inserted thanks to spectrogram inspection, looking for formant and intensity changes (increase towards the vowel and drop towards the consonant). That is, the boundaries between the vowels and the target singleton/geminate consonant were identified and, in case of gemination, no acoustic boundary was placed within the target consonant sequence. Kinematic measurements were performed after semiautomatic segmentation and labelling of the signal [11]. Figure n. 1 illustrates the acoustic signals (waveforms and sonograms) together with the kinematic signals (Lip Aperture, Tongue Body vertical) relative to the sentences “(richiam)a ‘mima/mimma ma(rcatamente)”. The singleton consonant (up) and geminated consonant (bottom) are aligned to the beginning of the first vowel “i” of “’mi(m)ma) (vertical line on the left), based on Tongue Dorsum (vertical) signal.

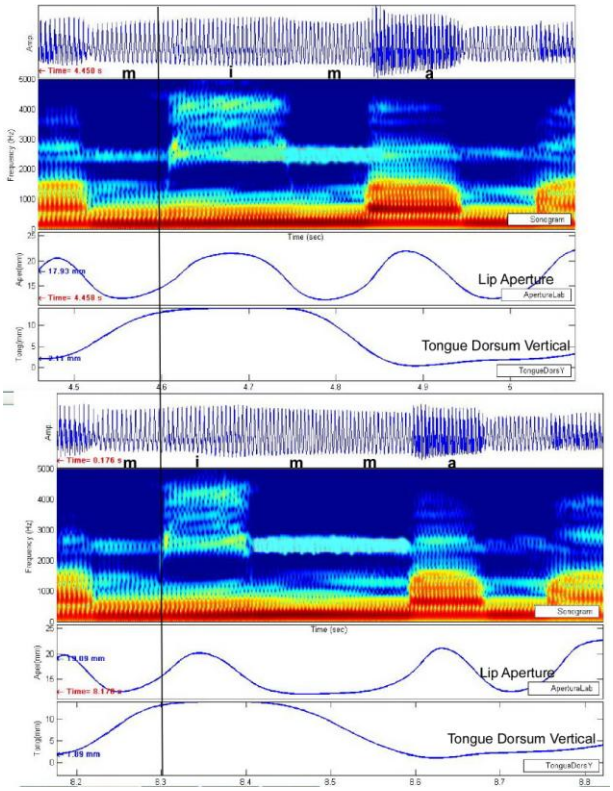

Figure 1: Acoustic (waveforms and sonograms) and kinematic (Lip Aperture, Tongue Body vertical) signals relative to “’mima” (up) and “’mimma” sentences (bottom). The two panels are aligned to the beginning of the first vowel “i” in the Tongue Dorsum signal (vertical line on the left)

On the basis of the time variations of the Lip Aperture parameter, of the vertical displacement of the tongue dorsum sensor, and of the derivatives of these two variables, the onsets and offsets of the closing and opening consonant gesture and of the [i-a] gesture were determined. The offset of the gesture toward the first vowel [i] and the onset of the gesture from the second vowel [a] were also determined on the same basis. Thus, the following measures could be provided (see fig. 2):

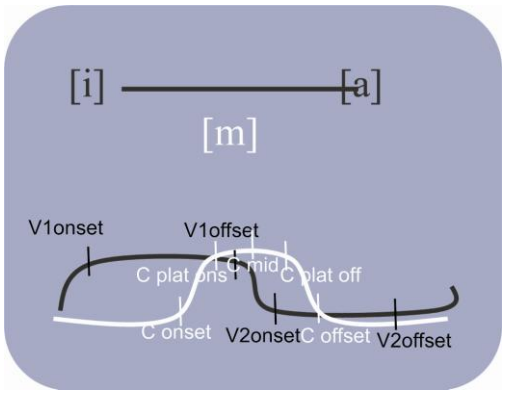

Figure 2: *Blueprint of the reference points on the kinematic trajectories of Tongue Dorsum (vertical) for first and second vowels (black) and Lip Aperture for the consonant (white) in “mim(m)a”*

V1 onset – C Mid: between the articulatory onset of the first vowel (/i/) and the center of the consonantal cycle /m(m)/. The articulatory onset for the vowel is defined as the point where the vertical component of the tongue dorsum reaches the highest point coming from a lower position for the last /a/ of "*richiama*" (i.e. call back). The exact point is calculated by making reference to the point where the istantaneous velocity reduces to 1/10 of maximal velocity for the same gesture. It is also the onset of the articulatory plateau for /i/, identified as the stable trajectory lying between the 2 crossings of the 1/10 of the maximal velocity. The articulatory target of the consonant is defined as the point where the vertical component of Lip Aperture has a unique maximum (that is, the end of the closing gesture coincides with the beginning of the opening gesture). In the case of an articulatory plateau, identified as the stable trajectory lying between the 2 crossings of the 1/10 of the maximal velocity, the mid point of the plateau is taken.
Öhman's prediction: VCCV should be similar to VCV
C Mid – V2 offset: between the center of the consonantal cycle /m(m)/ and the articulatory offset of the second vowel (/a/) (see definitions above).
Öhman's prediction: VCCV should be similar to VCV
V1 onset – V2 offset: between the articulatory onset of the 1st vowel (/i/) and the articulatory offset of the 2nd vowel /a/.
Öhman's prediction: VCCV should be similar to VCV
V1 onset – C onset: Between the onset of the articulatory plateau for /i/ and the beginning of the closure gesture for /m(m)/, identified as the point where the instantaneous velocity is 1/10 of the maximal velocity.
Öhman's prediction: VCCV should be less than VCV
C offset – V2 offset: Between the final moment of the opening gesture for the consonant /m(m)/, identified as the point where the velocity drops to 1/10 of the maximal velocity, and the end of the articulatory plateau for /a/, identified as the stable trajectory lying between the 2 crossings of the 1/10 of the maximal velocity.
Öhman's prediction: VCCV should be less than VCV
C onset – C offset: between the beginning of the closure gesture and the end of the opening gesture.
Öhman's prediction: VCCV should be more than VCV
V1 onset –V2 onset: between the articulatory onset of the 1st vowel (/i/) and the articulatory onset of the 2nd vowel /a/.
Öhman's prediction: VCCV should be be similar to VCV
V1 onset –V1 offset: the duration of the articulatory plateau for the first vowel.
Öhman's prediction: VCCV should be similar to VCV
V2 onset – V2 offset: the duration of the articulatory plateau for the second vowel.
Öhman's prediction: VCCV should be similar to VCV
C plateau onset – C plateau offset: duration of the articulatory plateau for /m(m)/, identified as the stable trajectory lying between the 2 crossings of the 1/10 of the maximal velocity.
Öhman's prediction: VCCV should be more than VCV.

3. Results

Since we excluded a significant number of repetitions on the basis of the perceptual assessment of the singleton/geminates contrast, we did not use the repeated measures ANOVA method and preferred to carry out Univariate ANOVAs separated for rates and subjects, with geminate/singleton as the between-subject factors, and the durations of the first vowel (V1) and the intervocalic consonant (C) as dependent factors for the acoustic analyses, and the timing intervals from articulatory measures as the dependent factors for the kinematic analyses.

3.1. Acoustic analyses

For all subjects, C duration in normal rate productions is significantly higher and V1 duration is significantly lower when a geminate rather than a singleton occurs (for C, speaker AG: [F(19,1)= 98.297; p= 0.000]; BG: [F(19,1)= 366.008; p= 0.000]; FC: [F(23,1)= 118.862; p= 0.000]; MP: [F(21,1)= 82.102; p= 0.000]; for V1, speaker AG: [F(19,1)= 64.981; p= 0.000]; BG: [F(19,1)= 32.111; p= 0.000]; FC: [F(23,1)= 14.655; p= 0.001]; MP: [F(21,1)= 37.590; p= 0.000]). As far as stimuli produced at fast speech rate are concerned, C duration is still significantly higher when a geminate rather than a singleton occurs for three out of four speakers; (speaker AG: [F(19,1)= 32.119; p= 0.000]; FC: [F(23,1)= 41.424; p= 0.000]; MP: [F(22,1)= 58.362; p= 0.000]. For the other speaker, the difference goes in the same direction although it is not significant (BG: [F(19,1)= 0.016; p= 0.901]). This is considered a speaker dependent feature, due to a less accurate articulation of the contrast at fast rate. On the other hand, V1 duration is still lower before geminates, although the difference is significant for just one out of four speakers; (speaker AG: [F(18,1)= 4.819; p= 0.042]; BG: [F(19,1)= 0.378; p= 0.546]; FC: [F(23,1)= 0.020; p= 0.889];MP: [F(21,1)= 1.408; p= 0.249]).

3.2. Kinematic analyses

The results are summarized in Table 1 and 3 (for normal speaking rate) and Table 2 and 4 (for fast speaking rate). In tables 3 and 4 the first column presents the predictions that would be made for the impact of the gemination contrast on the considered duration in the context of Öhman's model.

	AG		MP		FC		BG	
Time Interv.	G	S	G	S	G	S	G	S
V1ons-Cmid	136	114	163	154	195	187	196	179
Cmid_V2off	158	140	113	92	139	121	140	96
V1ons-V2off	294	254	276	246	333	308	336	275
V1ons-Cons	3	16	47	65	70	90	45	81
Coff-V2off	47	64	1	10	15	38	11	11
Cons-Coff	245	174	228	171	249	180	281	183
V1ons-V2ons	259	225	264	235	327	292	324	266
V1ons-V1off	52	66	97	116	119	142	117	134
V2ons-V2off	31	29	12	11	7	16	12	9
CplOns-Cploff	42	17	51	17	46	15	60	14

Table 1. *Subjects' means for time intervals at normal rate (rounded to ms; G = geminates; S = Singletons)*

	AG		MP		FC		BG	
Time Interv.	G	S	G	S	G	S	G	S
V1ons-Cmid	81	57	91	82	105	96	80	117
Cmid_V2off	138	110	108	85	124	95	104	100
V1ons-V2off	219	167	199	167	229	191	184	217
V1ons-Cons	-3	-9	23	23	34	38	14	41
Coff-V2off	66	51	32	26	43	22	32	23
Cons-Coff	157	124	145	119	152	131	139	153
V1ons-V2ons	189	147	172	152	201	175	163	199
V1ons-V1off	34	15	45	36	74	66	32	71
V2ons-V2off	30	19	27	15	28	16	22	18
CplOns-Cploff	14	6	11	8	13	7	10	12

Table 2. *Subjects' means for time intervals at fast rate (rounded to ms; G = geminates; S = Singletons)*

Time Intervals	Öhman	AG	MP	FC	BG
V1ons-Cmid	=	≠	=	=	=
Cmid_V2off	=	=	≠	≠	≠
V1ons-V2off	=	≠	≠	≠	≠
V1ons-Cons	≠	=	≠	≠	≠
Coff-V2off	≠	=	=	≠	=
Cons-Coff	≠	≠	≠	≠	≠
V1ons-V2ons	=	≠	≠	≠	≠
V1ons-V1off	=	=	≠	=	=
V2ons-V2off	=	=	=	≠	=
CplOns-Cploff	≠	≠	≠	≠	≠

Table 3. *Subjects response to Öhmans' predictions at normal rate (=: VCV not different from VCCV; ≠: VCV different from VCCV; p<.05)*

Time Intervals	Öhman	AG	MP	FC	BG
V1ons-Cmid	=	≠	=	=	=
Cmid_V2off	=	≠	=	≠	=
V1ons-V2off	=	≠	=	≠	=
V1ons-Cons	≠	=	=	=	=
Coff-V2off	≠	=	=	=	=
Cons-Coff	≠	≠	≠	≠	=
V2ons-V2off	=	≠	=	≠	=
V1ons-V1off	=	≠	=	=	=
V2ons-V2off	=	=	=	=	=
CplOns-Cploff	≠	=	=	≠	=

Table 4. *Subjects response to Öhmans' predictions at fast rate (=: VCV not different from VCCV; ≠: VCV different from VCCV; p<.05)*

These tables show a clear inter-speaker variability. Since only sequences where the singleton/geminates contrast was perceptually clearly noticeable, we don't think that this variability is due to dialectal differences. In addition the results obtained for the two speakers of the north-western variety of Italian are not fully compatible. It suggests that speakers of Italian could have different ways to implement the length contrast. In general, our results confirm the classical finding that geminates are significantly longer than singletons (Cons-Coff). It is important to observe that this contrast is well preserved at fast speaking rate (3 among the 4 speakers) at least in the selected sequences in which the contrast was perceptually considered to be maintained. However, our results are in clear disagreement with the predictions of Öhman's model: only one subject at normal speaking rate and two subjects at fast speaking rate confirm the hypothesis of a constant [i-a] duration (V1ons-V2off); only one subject at fast speaking rate is compatible with the hypothesis of a postponing of the consonant opening gesture in the second vowel (Coff-V2off). As concerns the third basic assumption underlying the use of Öhman's model for geminates (V1ons-Cons), i.e. the anticipation of the closing gesture in the preceding vowel, results are twofold: this hypothesis is essentially supported at a normal speaking rate, while all subjects behave in opposition to this prediction at fast speaking rate. Hence, contrary to Smith's [7] conclusion, it can be assumed that Italian geminates are not produced in a way compatible with Öhman's hypothesis.

A constant synchronization of the consonant target with the preceding vowel (V1ons-Cmid) is generally observed, while the duration (Cmid_V2off) is generally depending on the consonant contrast. This is an interesting result supporting the idea of a planning of the gemination within the VCV sequence or the VC sequence rather than within the CV sequence. This result provides also evidence against the c-center model, that would suggest a constant phasing between the consonant and the following vowel.

4. Discussion and Conclusions

In general, our results confirm the classical finding that geminates are significantly longer than singletons, even at fast speaking rate (3 among the 4 speakers). Our results suggest further that speakers of Italian could have different ways to implement the length contrast independently of the variety of Italian they speak. Smith's results [8] appear not to be confirmed both at the normal and fast speech rate. Our data show that the timing of maximum constriction with respect to tongue movement was not always constant across geminate\singleton consonants in the consonant-to-vowel gesture, and the vowel-to-vowel interval varied significantly. On the other hand, Smith's hypothesis concerning the anticipation of the geminate consonantal gesture in the preceding vowel is confirmed (consistent with Öhman's model); however, no significant delay in the following vowel is observed. At the fast speech rate, geminates and singletons show fewer significant differences than at normal rate.

Acknowledgements

To CNR for assigning to the first and the third author a Short Mobility Grant (in the years 2005 and 2010, respectively).

5. References

[1] Gili Fivela, B., Zmarich, C., "Italian Geminates under Speech Rate and Focalization Changes: Kinematics, Acoustic, and Perception Data", *InterSpeech 2005*, Lisbon, 2897-2900, 2005.

[2] Zmarich C., Gili-Fivela B., Perrier P., Savariaux C., Tisato G., "Consonanti scempie e geminate in Italiano: studio acustico e cinematico dell'articolazione linguale e bilabiale", in V. Giordani, V. Bruseghini, P. Cosi (a cura di), *Atti del III Convegno Nazionale dell'Associazione Italiana di Scienze della Voce (AISV)*, Trento, 29-30/11-1/12/2006, EDK Editore srl, Torriana (RN), 151-163, 2006.

[3] Romanu, C. & Calabrese, A., "The representation of geminate consonants: Evidence from the phonological errors of an aphasic patient", *J. Neurolinguistics, 9(3)*, 219-235, 1996.

[4] Bertinetto, P.M. , *Strutture prosodiche dell'italiano*, Accademia della Crusca, Firenze, 1981.

[5] Browman C. P. & Goldstein L. M., "Towards an articulatory phonology", *Phonology Yearbook, 3*, 219-252, 1986.

[6] Öhman S. E. G., "Numerical Model of Coarticulation", *JASA.*, 41, 310-320, 1967.

[7] Fowler, C., "Coarticulation and theories of extrinsic timing", *Journal of Phonetics*, 8, 113-133, 1980.

[8] Smith, C. L., "Prosodic patterns in the coordination of vowel and consonant gestures". In: B. Connell & A. Arvaniti (eds) *Papers in Laboratory Phonology IV, Phonology and phonetic evidence*. CUP, 205-222, 1995.

[9] Löfqvist, A., "Lip kinematics in long and short stop and fricative consonants", *JASA*, 117, 2, 858-878, 2005.

[10] Löfqvist, A., "Interarticulator programming: Effects of closure duration on lip and tongue coordination in Japanese", *JASA*, 120, 5, 2872-2883, 2006.

[11] Gili-Fivela, B., Zmarich, C., Perrier, P., Savariaux, C., & Tisato, G., "Acoustic and kinematic correlates of phonological length contrast in Italian consonants", *Actes du 16th International Congress of Phonetic Sciences, ICPhS 2007*, Sarrebruck, Allemagne, 469 – 472, 2007.

[12] Tisato G., Cosi P., Sommavilla G., Zmarich C., "Nuovi strumenti di Interface per l'elaborazione delle facce parlanti", in V. Giordani, V. Bruseghini, P. Cosi (a cura di), *Atti del III Convegno Nazionale dell'Associazione Italiana di Scienze della Voce (AISV)*, Trento, 29-30/11-1/12/2006, EDK Editore srl, Torriana (RN), 353-382, 2006.