Effort and coordination in the production of bilabial consonants

Maëva Garnier[1], Sarah Bouhake[1], Christophe Jeannin[1,2]

[1] *GIPSA-lab, Speech and Cognition Department, CNRS and Grenoble Universities, Grenoble, France*
[2] *Dental School, Claude Bernard Lyon I University, Lyon, France*

maeva.garnier@gipsa-lab.grenoble-inp.fr

Abstract

The first goal of this study was to compare two methods of measurement of lip articulation force. Bilabial consonants (/p/, /b/, /m/) were produced by a single French female speaker in modal and whispered speech, with varying levels of vocal effort. Lip compression force measured with a pressure sensor glued on the lower lip, correlates significantly, though only partially, with lip muscle activity, measured with surface EMG.
The second goal was to better understand the variations of lip articulation force in the production of stop consonant. Variations in lip compression correlate significantly, but not with a high level of correlation, with the variations in intra-oral pressure. This does not support the idea that the primary goal of lip compression is to oppose to the pressure built up in the oral cavity. Lip compression force and Pio were found to correlate significantly, though moderately, with the intensity of the burst created at occlusion release, but not with its spectrum mean.

Keywords: stop consonants, articulation force, EMG

1. Introduction

Speech production requires the precise coordination of breathing, laryngeal and articulatory gestures. Learnt by children during their speech development, this coordination is then automatized. However, it still remains a complex motor action that can become destabilized in non standard speaking styles (whisper, shouted or hyper-articulated speech) and that can dysfunction in several voice or speech disorders (vocal straining, stuttering, dyspraxia, …).
Stop consonants are of great interest for a better understanding of this coordination as they require a coordination not only in the timing of breathing, laryngeal and articulatory gestures, but also in their force. They are the first consonants to be produced at the babbling stage, and the most problematic segments for stutterers. They are also of particular interest for the measurement and modeling of production efforts, as the complete occlusion of the vocal tract requires a contact force between the lips, or between the tongue and the palate. It also implies vertical displacements of the larynx and adduction forces during the occlusion phase, perturbed vibration of the vocal folds (for voiced consonants) and a more or less abrupt glottal onset at the occlusion release (for unvoiced consonants).

The acoustical result of this coordination has been much studied and described in the phonetic literature: The spectrum of the bursts created at the occlusion release, the voicing intensity, the duration between the occlusion release and the voicing onset (Voice Onset Time (VOT)), or the direction of the first three formants transitions after the occlusion release, are as many audible cues to (un)voicing and place of articulation features that distinguish the different stop consonants (Forrest et al. 1988; Blumstein and Stevens 1979). The aerodynamics (intra-oral pressure, oral air flow) of the stop consonants production has also been extensively documented (Koenig et al. 2011). The articulatory targets of vocal tract constriction have been characterized using ultrasound imaging, electromagnetic articulography or (eletro)palatography, depending on preceding and following speech segments, and depending on the phonation mode (whisper/modal, soft/loud) (Lofqvist and Gracco 1997). So far, articulatory forces have mainly been estimated from cinematic criteria such as velocity peaks of speech articulators (Nelson 1983). No method has been consensually adopted to measure static forces yet. Some first attempts have been made to measure contact forces using pressure sensors between the lips (Hinton et al. 1992), inserted in dental prosthesis or artificial palates (Jeannin et al. 2008). Other studies have tried to estimate lip compression force using surface electromyography (Lubker and Parris 1970).
Most of the phonetic work on stop consonants aimed at describing significant differences between phonological *categories* (voiced vs. unvoiced, different places of articulation, comparing stops in different languages, …). However, it is not fully understood yet (1) which parameters and production gestures control the variation of phonetic features, and (2) the physical interactions that exist between the different control parameters. The present study aims at exploring the relationship between different acoustic, articulatory and aerodynamic descriptors of the bilabial stop production, with a particular interest in the variation of the lip articulation force.

2. Material and method

One French female speaker was recorded in laboratory conditions while producing the logatoms /apa/, /aba/ and /ama/ in modal and whispered voice, with 3 levels of vocal effort in the modal mode (murmured, conversational and shouted speech) and 2 levels of vocal effort for the whispered mode (normal and shouted whisper). 60 repetitions of the logatoms were recorded in each mode and level of vocal effort. A red light flashing every second gave her the speech rate.
Several signals were simultaneously recorded:
- The audio signal was recorded with a 1/4-in pressure microphone (Bruël and Kjær 4944-A) placed 30 cm away from the lips. The sound pressure level was calibrated using the 1 kHz internal reference signal of the conditioning amplifier (Bruël and Kjær Nexus).
- The intra-oral pressure (Pio) was recorded with a small capillary tube and the EVA acquisition system.
- Lip muscle activity was measured with two pairs of surface electrodes placed on the superior orbicularis (EMG1) and the depressor anguli oris (EMG2) (see Figure 1).

- The compression force between the upper and lower lips was estimated using a force sensor similar to that inserted in the PRESLA device (Jeannin et al. 2008), glued on the lower lip and calibrated in cm H20, using a water column and a latex container (see Figure 1).

The different segment phases (occlusion, burst, formant transitions) were segmented manually from the audio signal, using Praat software. The voice onset time (VOT) was measured as the time between the burst of energy, at the occlusion release, and the second formant onset point of the vowel that follows (Jovičić and Šarić 2008), which enabled us to measure a VOT for whispered speech. With this definition, the VOT of the voiced stop /b/ is consequently positive.
The 4 following descriptors were then extracted for all the logatoms: the mean intensity of the logatom (in dB SPL), the maximum Pio (in cm H2O), the maximum energy of the two EMG signals, and the maximum force of lip compression (in kPa).
The 3 additional descriptors were extracted for the stop consonants /p/ and /b/ only: the VOT (in ms), the intensity of the burst in energy (in dB SPL) and the center of gravity of its spectrum (in kHz). The burst spectrum was analyzed following the method described in Forrest et al. 1988, but up to 8kHz only (instead of 10kHz). In French, the burst of the voiced stop /b/ is produced, in modal speech, with voicing. In that specific case, we first modeled the burst signal as the sum of a harmonic component and a noisy component (HNN decomposition (Stylianou 1990). Thus we were able to subtract the estimated voicing component to the burst signal before analyzing it.

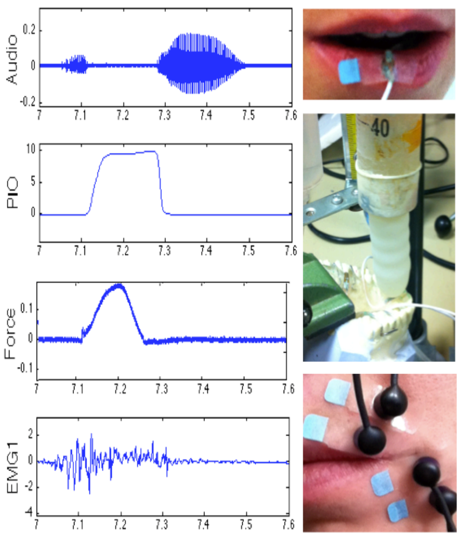


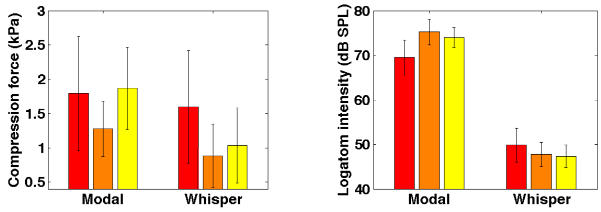
Figure 1: *Illustration of the force sensor and its calibration system, of the position of the EMG surface electrodes, and of the signals acquired simultaneously.*

Statistical analysis was conducted with the software R, considering, with a linear model, the variation of each descriptor as a function of the effects MODE (2 levels: modal and whispered speech at conversational levels) and SEGMENT (3 levels: pa, ba, ma for some parameters, and only 2 levels, pa, ba for other parameters).
The degree of correlation between several descriptors, taken by pairs, was also tested, as well as the effect of the MODE and the SEGMENT on this correlation. In that case, the productions were considered with different levels of vocal effort (murmured, conversational, shouted for the modal speech; normal and shouted for the whispered mode).

3. Measurement of lip articulation force

3.1. Lip compression force measured with a pressure sensor

Lip compression force, measured with the pressure sensor, was found to vary significantly with the phonation mode and the segment type (significant interaction $F(2,366)=8.20$, $p<.001$ ***): whispered stops tended to be produced with reduced compression of lips, compared to modal logatoms (see Figure 2. Test LRT, df=3, $F=22.96$, $p<.001$ ***), although that difference is significant for /b/ and /m/ only. Lip compression also depends significantly on the stop consonant (test LRT, df=16, $F=19.68$, $p<.001$ ***): always greater for /p/ than /b/.



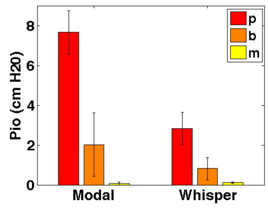
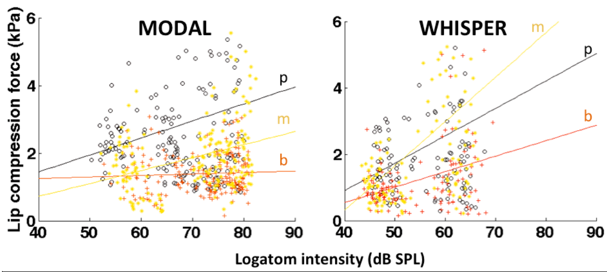


Figure 2: *Lip compression force, mean intensity and maximum intra-oral pressure (Pio) measured for the logatoms /apa/, /aba/ and /ama/ in comfortable modal and whispered voices.*

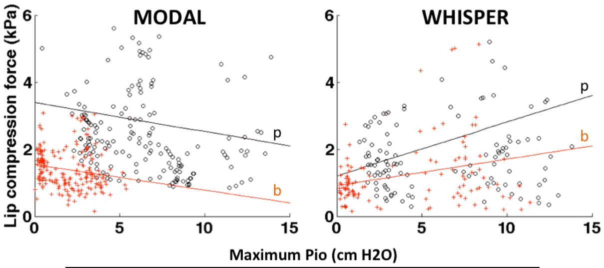
The greater lip compression force observed in modal speech compared to whispered speech can simply come from the fact that modal speech is globally more intense (by about 20dB, test LRT, df = 3, $F=2039.4$, $p< .001$ ***) and also produced with higher intra-oral pressures (Pio) (test LRT, df=3, $F=294.94$, $p<.001$ ***). A significant positive correlation is indeed observed between the lip compression force and the average intensity of the logatom (test LRT, df=4, $F=25.67$, $p<.001$ ***). So is also a significant positive correlation observed between lip compression and Pio (test LRT, df=2, $F=13.34$, $p<.001$ ***).



	apa	aba	ama
Modal	R=0.21, ** a=0.0020	R=0.05, NS	R=0.33, *** a= 0.0015
Whisper	R=0.27, ** a=0.0032	R=0.34, *** a=0.0018	R=0.48, *** a=0.0052

Figure 3: *Correlation between lip compression force and average intensity of the logatoms /apa, /aba/ and /ama/ produced in modal or whispered speech, with varying levels of vocal effort.*

However, the variations in lip compression force between the 3 bilabial consonants can hardly been explained by variations in Pio and global intensity. Figure 2 shows, in particular, how the segment /m/ is produced with comparable articulation force to the two other segments, although no pressure builds up in the oral cavity for that nasal stop. Consequently, the correlation between lip compression force and global intensity or Pio, is not always underlined by the same relationship but is significantly affected by the factors MODE and SEGMENT. Thus, the slope of the linear regression between lip compression force and global intensity is always greater in whisper than in modal speech (test LRT, df=1, F=20.49, $p=7.10^{-6}$ ***) and different between the 3 segments (test LRT, df=2, F= 6.56, p=0.001 **). Likewise, the slope of the linear regression between lip compression force and Pio is greater in whisper than in modal speech (test LRT, df=1, F=21.38, p<.001 ***). However, it does not depend significantly on the segment (see Figure 4).



	apa	aba
Modal	R=-12, NS	R=-0.17, * a=-0.0030
Whisper	R=0.29, ** a=0.0063	R=0.26, ** a=0.0031

Figure 4: *Correlation between lip compression force and maximum Pio, for the two bilabial /p/ and /b/ produced in modal or whispered speech, with varying levels of vocal effort.*

3.2. Articulation force estimated from lip muscle activity

Force /EMG1	apa	aba	ama
Modal	R=0.23, ** a=2.29	R=0.25, *** a=3.76	R=0.44, *** a=4.66
Whisper	R=0.42, *** a=3.88	R=0.39, *** a=4.75	R=0.47, *** a=6.73

Force /EMG2	apa	aba	ama
Modal	R=0.41, *** a=4.55	R=0.30, *** a=8.29	R=0.50, *** a=8.75
Whisper	R=0.32, *** a=4.53	R=0.20, * a=4.71	R=0.62, *** a=6.50

EMG2 /EMG1	apa	aba	ama
Modal	R=0.23, ** a=0.25	R=0.41, *** a=0.75	R=0.38, *** a=0.63
Whisper	R=0.67, *** a=1.02	R=0.51, *** a=0.97	R=0.23, ** a=0.17

Table 1: *Correlation between the lip compression forces measured with a pressure sensor, and the activity of two lip muscles (EMG1: superior orbicularis; EMG2: depressor anguli oris), for the three logatoms /apa/, /aba/ and /ama/.*

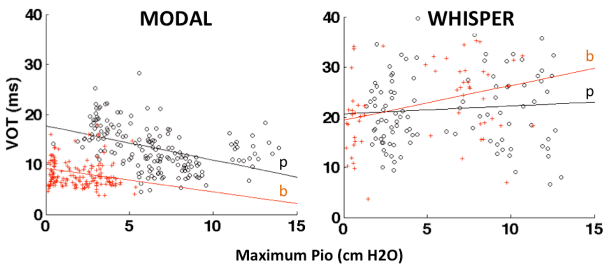
A significant and positive correlation was found between the lip compression force, measured with the pressure sensor, and the level of lip muscle activity, for the superior orbicularis (EMG1: test LRT, df=4, F=44.6, p<.001 ***) and the depressor anguli oris (EMG2: test LRT, df=3, F=59.2, p<.001 ***). However, this relationship is not invariant but depends significantly on the segment: the slope of the linear regression (coefficient “a” in the tables) tends to be greater for the segment /m/, compared to /b/, and to /p/ (EMG1*Segment: test LRT, df=4, p<.001 ***; EMG2*Segment: Test LRT, df=2, F=4.0, p=0.019, *). This correlation also depends significantly on the mode of phonation for the superior orbicularis muscle (EMG1*Mode: test LRT, df=2, p<.001 ***), with always a greater slope of regression in whisper than in modal speech. This is not the case for depressor anguli oris muscle (EMG2*Mode, test LRT, df=1, F=1.2 p=0.26 NS).
These results indicate that the measures of lip compression force and lip muscle activity are not redundant or equivalent, and that the level of muscle activity cannot entirely predict the amount of lip compression force.

Likewise, the intensity of the two EMG signals are significantly and positively correlated (test LRT, df=4, F= 58.1 p<.001 ***. See also Table 1). However, the correlation is only partial (R=0.67 at its best for /p/ in modal speech) and still depends on the mode (test LRT, df=1, F= 23.2, p= 2.10^{-6} ***) and the segment (test LRT, df=2, F=4.0, p=0.018, *). Consequently, we can consider that these two signals are not redundant and bring complementary information about the articulation of stop consonants.

4. Variation of phonetic cues

4.1. Voice Onset Time (VOT)

In agreement with previous studies, the VOT was found to be shorter for voiced consonant /b/ compared to the unvoiced consonants /p/ (F(1,206)= 31.59, p<.001 ***). A significant increase of the VOT was also observed from modal speech to whispered speech (F(1,206)=180.16, p<.001 ***). .



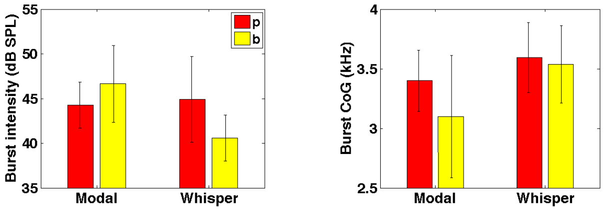
	apa	aba
Modal	R=-0.45, *** a=-0.68	R=-0.15, * a=-0.46
Whisper	R=0.08, NS	R=0.32, * a=0.69

Figure 5: *Correlation between the voice onset time and the maximum Pio, for the two bilabial /p/ and /b/ produced in modal or whispered speech, with varying levels of vocal effort. The VOT was measured as the time difference between the burst and the onset of the second formant.*

The variations of Pio partly explain these variations of VOT (Correlation VOT ~ Pio: test LRT, df=3, F=12.44, p<.001 ***), although such a correlation was not observed for /p/ in whispered speech (see Figure 5). Again, this relationship is not always the same and depends significantly on the segment (test LRT, df=1, F= 4.89, p=0.027 *) and the mode (test LRT, df=2, F=18.34, p<.001 ***): with positive correlations in modal speech and negative ones in whisper

4.2. Burst intensity and spectrum

The burst intensity varies significantly as a function of both the mode of phonation and the segment (mode*segment: F(1,246)= 50.59, p<.001 ***). Although no general tendency can be drawn, the variations of the burst intensity can partly be explained by the variations in intra-oral pressure (Correlation: Test LRT, df=3, F=46.45, p<.001 ***) and lip compression force (Test LRT, df=3, F=16.86, p<.001 ***), excepted for the segment /b/ produced in modal voice. This relationship depends significantly on the mode of phonation, with positive correlations in whisper and negative ones in modal speech.



	I_{Burst} ~ Pio		I_{Burst} ~ Lip force	
	apa	**aba**	**apa**	**aba**
Modal	R=-0.39, *** a=-0.63	R=-0.02, NS	R=-0.22, ** a=-12,2	R=-0.02, NS
Whisper	R=0.46, *** a=0.71	R=0.65, *** a=1.08	R=0.29, ** a=20.4	R=0.41, *** a=56.6

	CoG_{Burst} ~ Pio		CoG_{Burst} ~ Lip force	
	apa	**aba**	**apa**	**aba**
Modal	R=0.14, NS	R=-0.01, NS	R=0.23, ** a=867	R=0.10, NS
Whisper	R=0.13, NS	R=0.20, NS	R=-0.05, NS	R=0.07, NS

Figure 6: *The top figures represent the intensity and the spectrum mean (CoG) of the burst produced at the occlusion release for the logatoms /apa/ and /aba/ in conversational modal and whispered voices. The bottom tables summarize the results of the correlation analysis between the burst intensity or CoG, and two control parameters: the intra-oral pressure (Pio) and the lip compression force, measured wit the pressure sensor.*

The center of gravity of the burst spectrum (CoG) was found to be slightly higher in whispered speech, compared to modal speech (test LRT df=2, F=26.83, p<.001 ***). On the other hand, no significant correlation was found between the variation of CoG and the Pio or the lip compression force, except for the segment /p/ produced in modal voice.

5. Conclusion

Two methods of measurement or evaluation of lip articulation force have been compared here on a single female French speaker, for a large variety of productions (modal and whispered speech, from murmured to shouted levels). Lip compression forces measured with a pressure sensor glued on the lower lip were observed with comparable values to the literature (Hinton et al. 1996) and were found to increase with the global level of vocal effort. The activity of two lip muscles controlling the movement of the upper and the lower lips correlates significantly with this lip compression force, as well as with each other, without providing redundant information.
The lip compression force correlates significantly, though not with a high level of correlation, with the maximum Pio of the bilabial plosives /p/ and /b/. Furthermore, the nasal stop /m/ demonstrates comparable articulation force to the two other segments, although it is produced with weak or null Pio. These results do not support the idea that the primary aim of lip compression is to maintain occlusion of the vocal tract by opposing to variations of Pio.
Lip compression force was also found to correlate significantly, though not with a high level of correlation, with the burst intensity. No such relationship was found with the spectrum CoG of the burst. This results support the idea that lip articulation force may partly control the variations of burst intensity.
Further investigations will explore the combined contribution of different control parameters to the variation of the burst characteristics, but also to the transient characteristics that follow the occlusion release and that are crucial to the perception of stop categories (Blumstein and Stevens 1979).

6. Acknowledgements

This work was supported by the pole CSVSB of the University Joseph Fourier (Grenoble). We also thank Thomas Hueber for his help with the HNN decomposition.

7. References

Blumstein, S. E., & Stevens, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. The Journal of the Acoustical Society of America, 66(4), 1001-1017.

Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word initial voiceless obstruents: Preliminary data. The Journal of the Acoustical Society of America, 84(1), 115-123.

Jeannin, C, Perrier, P, Payan, Y, Grosgogeat, B, Dittmar, A, Géhin, C (2008) « PRESLA: An Original Device to Measure the Mechanical Interaction between Tongue and Teeth or Palate during Speech Production » . Proceedings of the 8th International Seminar on Speech Production, Strasbourg, France.

Hinton, V. A., & Luschei, E. S. (1992). Validation of a modern miniature transducer for measurement of interlabial contact pressure during speech. Journal of speech and hearing research, 35(2), 245.

Hinton, V. A. (1996). Interlabial pressure during production of bilabial phones. Journal of Phonetics, 24(3), 337-349.

Jovičić, S. T., & Šarić, Z. (2008). Acoustic analysis of consonants in whispered speech. Journal of voice, 22(3), 263-274.

Koenig, L. L., Fuchs, S., & Lucero, J. C. (2011). Effects of consonant manner and vowel height on intraoral pressure and articulatory contact at voicing offset and onset for voiceless obstruentsa). The Journal of the Acoustical Society of America, 129(5), 3233-3244.

Lofqvist, A., & Gracco, V. L. (1997). Lip and jaw kinematics in bilabial stop consonant production. Journal of Speech, Language, and Hearing Research, 40(4), 877.

Lubker, J. F., & Parris, P. J. (2005). Simultaneous measurements of intraoral pressure, force of labial contact, and labial electromyographic activity during production of the stop consonant cognates/p/and/b. The Journal of the Acoustical Society of America, 47(2B), 625-633.

Nelson, W. L. (1983). Physical principles for economies of skilled movements. Biological cybernetics 46.2: 135-147

Stylianou, I. (1990). Harmonic plus Noise Models for Speech, combined with Statistical Methods, for Speech and Speaker Modification (PhD, Signal et Image, ENST Paris,Paris).