



**HAL**  
open science

## On gestural variation and coarticulation effects in sound control

Tommaso Bianco, Vincent Freour, Nicolas Rasamimanana, Frédéric Bevilacqua, René Causse

► **To cite this version:**

Tommaso Bianco, Vincent Freour, Nicolas Rasamimanana, Frédéric Bevilacqua, René Causse. On gestural variation and coarticulation effects in sound control. Springer Verlag. Gesture in Embodied Communication and Human-Computer Interaction, Lecture Notes in Computer Science (LNCS) volume 5934, pp.134-145, 2010. hal-01161270

**HAL Id: hal-01161270**

**<https://hal.science/hal-01161270>**

Submitted on 21 Jan 2016

**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.

# On gestural variation and coarticulation effects in sound control

Tommaso Bianco, Vincent Freour, Nicolas Rasamimanana, Frederic Bevilacqua,  
and René Caussé

IRCAM UMR STMS,  
1 Place Igor Stravinsky Paris, France  
{tommaso.bianco, vincent.freour, nicolas.rasamimanana,  
frederic.bevilacqua, rene.causse}@ircam.fr

**Abstract.** In this paper we focus on the analysis of sound producing gestures in the musical domain. We investigate the behavior of intraoral pressure exerted by a trumpet performer in the production of single and concatenated notes. Investigation is carried out with functional data analysis techniques. Results show that different variation patterns occur for single note production, which depend on dynamic level, suggesting the hypothesis that two different motor control programs are available. Results from analysis on consecutive notes give evidence that the coarticulation between two gesture curves cannot be modelled by linear superposition, and that local coarticulation is affected by contiguous units.

**Key words:** coarticulation, music performance, gesture, anticipation, motor program

## 1 Introduction

In music performance, sound production is always linked to continuous gesture processes. Thus, the subtle variations of these processes must be taken into account to study how musician actually controls expressively their sound. Moreover, variation in the execution of a gesture depends on the musical context. When embedded into a continuous stream of units, each gesture segment undergoes influences from the surrounding segments, a contextual modification usually called, from phonetics literature, coarticulation.

In this paper we focus on the analysis of gestures for the production of sound in trumpet performance. Our analysis concentrates on two main types of variability: intrinsic, that is related to the gesture for a single note execution, and extrinsic, that is variability caused by coarticulation effects between adjacent gesture events. We apply functional data analysis techniques in order to identify specificities among different playing techniques.

Our aim is to give more evidence for the understanding of the underlying process related to the sound control in trumpet performance. The results of our analysis might be of help for recognition as well as for synthesis purposes in the general domain of control of sound synthesis.

## 2 Related works

The subject of coarticulation has been widely studied in phonetics [1]. Nevertheless, it is still a central and challenging theme both for automatic speech and sign recognition [2]. Among few theories elaborated for the modeling of this phenomenon (see Farnetani for a review [3]), the co-production theory of Fowler is of particular interest for the present account. The co-production theory explains the coarticulation between two segments as the result of their overlapping activity. The boundaries of the two segments thus extend onto each other, instead of a common “perpendicular to the time axis” frontier [4].

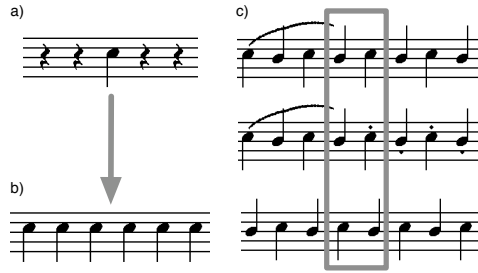
There is, however, a fundamental difference between natural speech and music performance. For the latter, in fact, the timing of gesture units is subjected to an exogenous temporal scheduling, usually dictated by the score tempo mark, or by the global cadence of the performers ensemble. Moreover, an expressive taxonomy - technically referred as “articulation” and explicitly formalized in the score notation - intervenes on the units transitions. Due to its intrinsic multiplicity and to the strict temporal scheduling of events, the coarticulation in the musical domain is also related to anticipation. This refers to the adjustments made to units of motor action in order to accommodate the instantiation of the next units in the sequence.

Anticipation in music performance has been studied in the performance of piano [5] [6], drums [7], and violin [8], whereas coarticulation in relation to chunking has been tackled in [9]. To our knowledge, no work on the subject has been done for the trumpet.

The majority of literature on trumpet performance, and on wind-instrument in general, has focused on the analysis of steady values, such as sound dynamic level, sound spectrum, and blowing pressure thresholds [10], and connected to human respiratory mechanics [11]. Transient analysis has been investigated for the relationship between mouthpiece and instrument’s bell pressures [12], and for the tonguing and muscular synergy, with EMG [13][14] and cinefluoric techniques [15]. We are not aware of studies of temporal behavior of intraoral pressure, for single note as well as for coarticulation between notes.

## 3 Experimental Methods

**Protocol.** A professional trumpet player was asked to perform a combination of C5 and B4 quarter notes with specific dynamic level and articulatory indications, at a fixed 120 bpm tempo throughout all the recordings. An example of the score is shown in Fig.1. A first set of measurements focused on the execution of a single C5 quarter note with three different dynamical levels (*pp*, *mf* and *ff*), and with a staccato mark and *mf* dynamics. In a second set of measurements, the performer was asked to play a sequence of quarter notes (concatenation of C5 notes and of alternating B4-C5 notes) with *mf* dynamics. The last part of the experiment involved a sequence of alternating C5 and B4 notes with fixed *mf* dynamics. Two consecutive groups of four notes formed the entire sequence. Each group



**Fig. 1.** A subset of the tasks executed by the performer, for the first (a), second (b) and third (c) set of measurements.

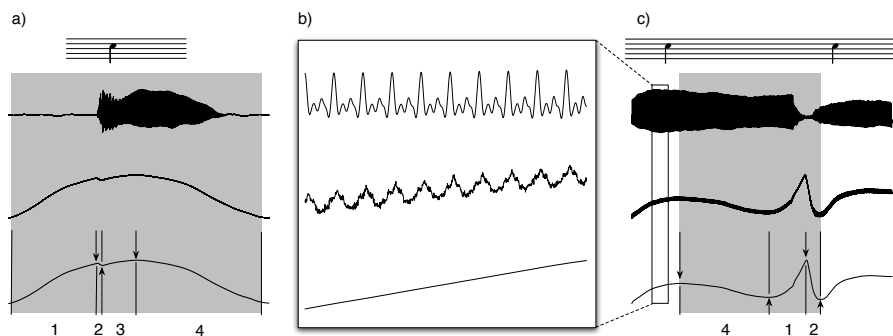
was assigned a type of articulation: *non-legato*, *legato*, and *staccato*. In all cases no measure bar was presented in the score, so as to avoid any possible dynamic accentuation on the first note of each measure, possible due to cognitive rhythmic grouping [16].

*Materials and recording* Intraoral pressure was measured with a catheter tube of approximately 15 cm. The pressure was sensed with a SCX Pressure Sensor, which can afford a 100 microseconds of response time. Pressure and sound were recorded with a NI acquisition card with sampling rate set to 48 kHz. The performer was asked to keep one extremity of the tube proximal to the soft palate, behind the lingua-alveolar place of occlusion for /t/ and /d/ consonants. Each measurement was checked after each take in order to detect possible artifacts in the measurements. The closing of the extremity of the tube by the contact with the internal mouth surfaces or tongue can lead to “locks” in pressure, which remains constant during the closing period. Thus, such cases were excluded from the analysis.

## 4 Analysis Methods

**Filtering.** The initial 2 seconds long sequences were filtered by a lowpass 6th order Butterworth filter with cutoff frequency equal to 300 Hz. This process removed possible sensor noise as well as the frequency of oscillation of the lips. An example of the resulting curves obtained after the filtering step is shown in Fig. 2 a.

**Segmentation.** For the analysis of single notes, the transition of the intraoral pressure over a background threshold level determined the start and the end of each sequence. The threshold level corresponded to the physical configuration in which the mouth is closed onto the instrument mouthpiece, but neither an airflow, nor a pressure potential are present. We partitioned the single note curves into four zones (as displayed in Fig. 2 a), each of them corresponding to a different phase in the sound production. In zone 1, the performer starts increasing the pressure in the oral cavity; its tongue closes onto the alveolar ridge, obstructing the airflow in the vocal tract. The beginning of zone 2 marks



**Fig. 2.** Segments of recorded data used in the analysis. Detail of sound, original pressure and filtered pressure (a); windows used for coarticulation (b), and single note (c) analysis

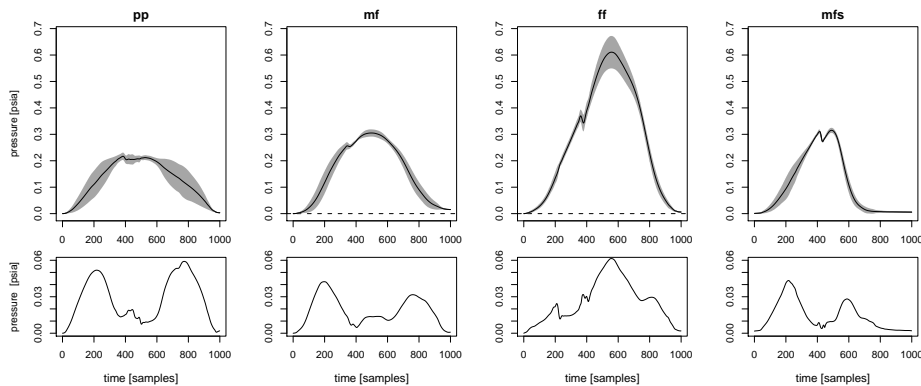
the opening - the detaching of the tongue from the alveolar ridge - of the vocal tract to the airflow. The pressure instantly drops, and the lips begin to oscillate, marking the onset of the note. This process is usually called alveolar stop in phonetics, and it characterizes voiceless plosive consonants such as /t/ and /d/, which are usually used as example in trumpet teaching. Zone 3 encloses the second pressure increase, during which the performer brings the note sound to the sustain phase. Finally, the pressure decreases all along zone 4 for the note closure.

This functional partition helped to identify similar zones in the recordings of concatenated notes. Hence, the coarticulation window built upon a combination of these four zones. The criteria based on the background level to distinguish the end of zone 4 and the beginning of zone 1 for the single note case, has here been substituted by a criteria based on the curve between the two notes. An example of the resulting segment is presented in Fig. 2 c. For the last part of the experiment the zone 3 of the incoming note has been excluded from the window of analysis. Indeed, as a staccato mark imposes the note to be played with shorter duration, a right boundary on the peak of sustain would have shortened the segment, hence allowing an external factor to interfere on the coarticulation study.

**Length reduction.** Last, in order to obtain sequences with an equal number of samples, and to reduce computation time in the analysis process, we extracted 1000 points from each segment, equally spaced in time from the start to the end of the segment.

**Functional data analysis.** The curves were processed with functional analysis techniques provided by the R package of Ramsay and colleagues<sup>1</sup> [17] [18]. Each time-series was converted into a linear combination of 200 equally spaced 6-order B-spline basis functions. Approximation to the original data points is obtained

<sup>1</sup> <http://r-forge.r-project.org/projects/fda/>



**Fig. 3.** Mean curves with confidence intervals (upper), and standard deviations (lower) for Pianissimo (pp), Mezzo forte (mf), Fortissimo (ff), and Mezzo forte staccato (mfs) articulations

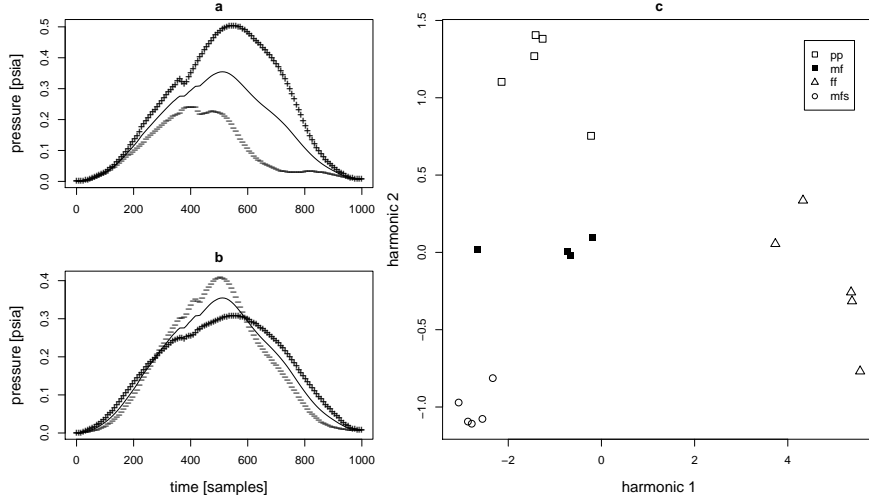
by a combined least squares and roughness penalty, with a  $\lambda$  weighing coefficient equal  $10^{-0.5}$  on the second derivative, as delivered by the generalized cross validation criterion [17]. The smoothing process with 200 spline basis functions for trajectories of 1000 points, in relation to the original lengths and sampling rates, assures - according to the relation formula in [19] - that variations up to 500 Hz are preserved in the functional smoothed representation.

Subsequently, a landmark registration synchronized the set of curves by performing nonlinear time stretching, ensuring the concurrent occurrence of an arbitrary number of key points among instances. A higher value for  $\lambda$  equal to  $10^5$  in the previous step was used to register velocities to guarantee equal number of valleys-peaks in the curves, required for the convergence of the registration process.

To investigate the primary modes of variation in the data, we recurred to functional principal component analysis. Roughly, the functional version of PCA replaces variable values with function values, converting the search for eigenvectors of the covariance matrix with the search for orthonormal eigenfunctions of the bivariate covariance function. Each eigenfunction  $\xi_j$  represents a form of variation around the mean curve  $\bar{x}(t)$ , participating in each instance curve  $x_i(t)$  with a score given by the integral  $\int \xi_i(t)[x_i(t) - \bar{x}(t)]dt$

## 5 Results

*Single note.* We registered the curves of each class separately, so as to obtain a representative mean curve for each articulation. The resulting means and standard deviations are presented in Fig. 3.



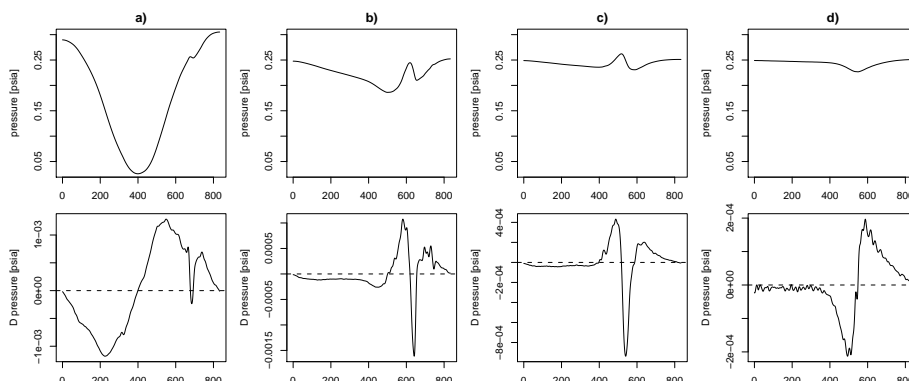
**Fig. 4.** Mean pressure curve with addition (++) and subtraction (--) of a suitable multiple of the first (a) and second (b) principal components. Scores of the first two principal components for each original instance

A close examination of the plots show that the variability distribution among conditions can be classified into two groups according to shape. In the first case there is an intensification for the main transitions, for *pp*, *mf*, and *mfs* cases, where pressure is raising before the initiation of the air flow, and is subsequently decreasing for the note closure. In the second case, for *ff* dynamics, the variability increase in the sustain stage of the note.

Absolute variability also changes quantitatively among cases. Values of global coefficients of variation (average values over the curves) for *pp*, *mf*, *mfs* and *ff* are respectively 0.23, 0.12, 0.13 and 0.09, whereas maxima in standard deviation fall at roughly 10% of the pressure maxima for *mf*, *ff*, and *mfs* case and at 20% for *pp*. Assuming that variability may increase with task demands [20], it seems that *pp* performance represents a higher demanding task compared to the other dynamics, a principle already established for vocal performance [21].

In order to better discriminate the differences of each case, we explored the major modes of variability in the data by means of functional PCA. Fig. 4 reports the results for the first two components, which account for 97.4% of the overall variability among the curves. In Fig. 4 a) and b), the mean curve is presented, in conjunction with the addition and subtraction of a suitable multiple of the principal components. Plot c) reports the scores of the two principal components for each instance, which distinctly cluster into four groups.

The results of the analysis indicate that the variation in the data, with this set of components, is roughly two-dimensional. The first component, which dominates with 90.6% of the variation, incorporates different typologies of variation: the amplitude of the curve, which operates on the loudness of the note, the curve



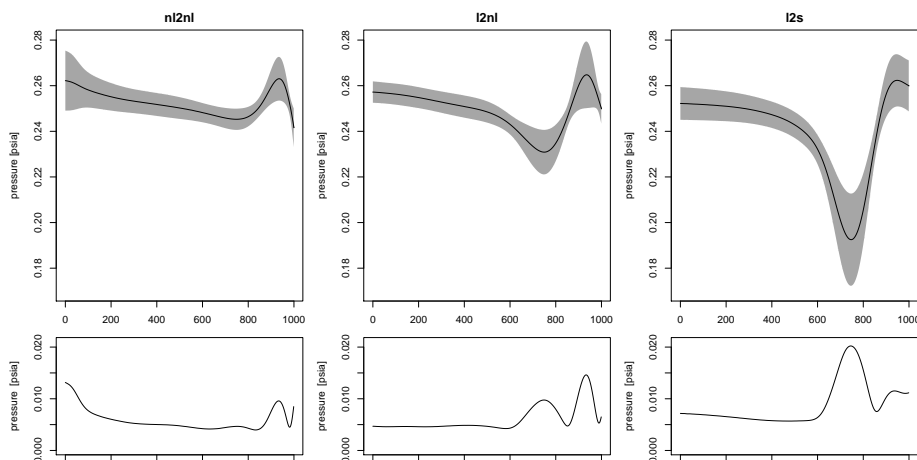
**Fig. 5.** Pressure curves (upper) and first derivatives (lower) for coarticulation of synthesized case (a), non-legato transition between two C5 notes, non-legato transition between C5 and B4 notes, legato transition between C5 and B4 notes. The sound onset of the incoming note occurs just after the pressure peak for b and c, while in pressure minimum for c.

span, the dynamics of the pressure decrease. However, this component only helps to distinguish the *ff* instances from the other articulations. In order to differentiate the remaining classes, one needs to consider the second component, which accounts for 6.8% of the variation, and which resumes variation on the curve narrowness. A positive contribution of the second component on the mean compresses the curve peak, widening and lowering its shape, and spans uniformly the *pp*, *mf*, and *mfs* examples.

**Coarticulation.** The concurrent activation of primitives in the muscular force field has been shown to be the result of roughly linear vectorial superposition [22][23]. Under the assumption that intraoral pressure approximates to a uniform scalar field inside the oral cavity, the transposition of this principle to our case would resolve into a scalar superposition. We would expect that, in zones of pressure overlap, the overall curve should be approximated by a direct sum of two single note curves.

We therefore compared the recorded coarticulation segments - which comprised zones 4,1,2,3 as of Fig. 2 - to a simulated coarticulation, obtained by linearly superposing the real recordings of two single notes. The superposition was built upon the mean curve of *mf* C5 single note as follows: the curve was first resized, in order for the corresponding sound envelope length to match between the single and concatenated occurrences; it was then segmented into two pieces at the pressure maximum, where the decrease in pressure begins - start of zone 4. A shifted version of the first segment has been linearly superposed to the second one, so as to build a max-to-max window of the same length as

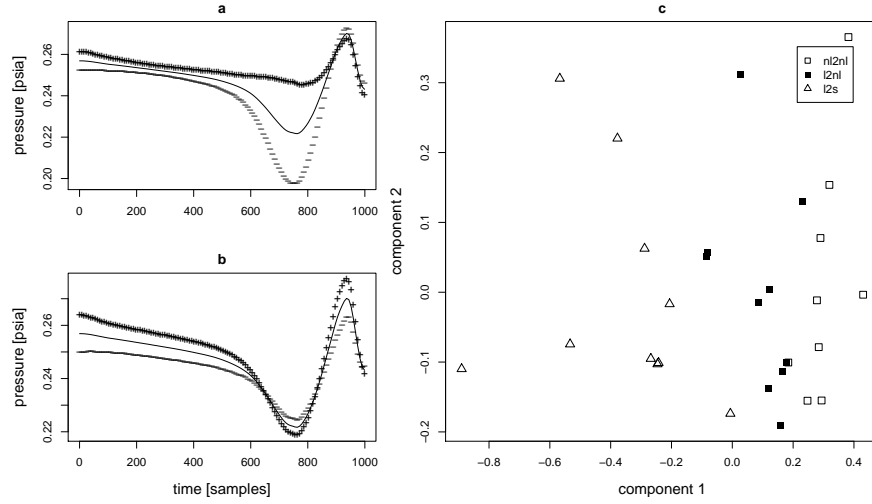




**Fig. 6.** Mean curves with confidence intervals (upper), and standard deviations (lower) for non-legato to non-legato (n|2n|), non-legato to legato (l2n|), and legato to staccato (l2s) coarticulations

the one segmented from the real recordings. The cut-and-paste operation has been performed by adjusting the spline coefficients level, to maintain continuity in derivatives for the resulting curve, and therefore to assure a maximal smooth transition. The resulting curve and the real coarticulation recordings are shown in Fig. 5. From a visual inspection of the plots, one can remark that the synthesized curve does not fit any of the real curves. Neither the concatenation of equal notes (b), nor the transition between different notes (c). In the simulated coarticulation, indeed, the main pressure valley falls far beyond the minima values experimentally recorded. For pressure first derivative, the shape of the synthesized version resembles the real cases: negative values during the closure of the outgoing note, to which three alternating positive-negative peaks follow. However, for experimental data the velocity in the first part is approximately constant all along the closure of the outgoing note. In a dynamical perspective, the musician appears to hold further the previous note, delaying to the very last instants the instantiation of the new note. Last, the intermediate pressure peak is almost absent in the simulated segment.

In what follows we will show that coarticulation between two units is also affected by preceding or following units. The segments analyzed correspond to the grey window of Fig. 2 c (sequential union of zones 4,1 and 2), extrapolated from the transition between the fourth and fifth notes during the performances of the scores in Fig. 1 c. According to the notation in the scores, the transition in the three occurrences should be performed in a similar fashion. Indeed, the legato mark enclosing the outgoing note only regulates the transition with its predecessor, and such transition is not considered in our segment. And the



**Fig. 7.** Mean pressure curve with addition (++) and subtraction (--) of a suitable multiple of the first (a) and second (b) principal components, for coarticulation records. Scores of the first two principal components for each original instance

staccato mark on the incoming note only behaves on its last part, which also is omitted from our segment.

The registered mean curves and their standard deviation, synchronized by a collective landmark registration process, are reported in Fig. 6. The three cases manifest a clear individual profile, with an increasing (relatively for *nl2nl*, *l2nl*, and *l2s*) negative spurt just before the instantiation of the incoming note. The peak's amplitudes correlate with the peaks in standard deviation, which attain between 20% and 30% of the peak amplitude. A minimum level in standard deviation equal to 2% of the average pressure level, and common to all records, suggests inconsistency in the level when playing *mf* among executions.

In order to analytically locate the principal modes of variation, we again turned to functional PCA, whose results are presented in Fig. 7. The first component (a), which accounts for 74.5% of the variation, represents the amount of pressure valley, while the second (b) resumes variability in the steady level and slightly in spurt over/undershoot. The plot for components' scores (c) tells us that the first component suffices for classification purposes, giving evidence that the amount in pressure valley can explain the different coarticulation behaviors.

Principal curve analysis for higher order derivatives did not improve the clustering results.

## 6 Conclusions

Functional data modeling represents a promising tool for synthesis purposes. First, the spline basis representation allows to easily smoothly join consecutive

curves. An arbitrary long sequence of gesture units can then be built by connecting single note and coarticulation segments. The presence of end-point level differences between segments could be eliminated by using a dual representation as in [24], so to preserve the same pressure (and therefore sound output) level among notes repetitions. Second, as shown by the low dimensional set of eigenfunctions, functional PCA allows for the storage of a reduced set of curves (means and dominant eigenfunctions), but with the possibility to span the set of original real recordings, or unrecorded but plausible instances, via their linear combination. Further investigation have to be done on the verification of original curves reconstruction trough principal functions, and on the relation between components' scores and task practice [25].

In the analysis of single notes, we showed how a single gesture unit exhibits systematically different behaviors, depending on the dynamic level required for the sound outcome. On the basis of a twofold shape for variability distribution, we hypothesize that two motor control programs may be available for the completion of the task: a continuous feed-back guided control for the production of *pp*, *mf*, and *mfs* notes, and a pre-planned movement control for the *ff* performance. The first movement strategy is employed to target a value with a maximal accuracy in the endpoint, that is with minimum steady-state error and overshoot issues. In practice, the oral and respiratory systems have to raise the pressure at the minimum level necessary to instantiate the oscillation of the lips. The performer's skill resides in the organization of the muscles synergy, in order to reach the target level at a specific time instant, and starting from different initial configurations. He continuously adjusts the forces applied to the muscles - this is confirmed as well by a higher number of acceleration zero crossings - transferring variability from the endpoint trajectory outcome to the whole movement pattern. Variability as a result of compensation has been defined as "adaptive variability" [26], and it has been found in more general motor tasks [27]. The second control strategy - the one which underlies the *ff* performance - draws similarities with ballistic movement theory: a fast, preprogrammed and impulsive movement, which lacks of on-line control, and whose purpose is to reach a target in the shortest time. This motor program would explain the higher variability in the sustain stage, for the occurrence of target undershoot/overshoot. In order to validate these hypothesis, however, one should also consider the relation between intraoral pressure variability and sound variability on a perceptive scale, for the performer regulates expiratory airflow rate according to auditory feedback [28].

In the second part of this work, we provided evidence that further complexity is added when individual gestures are interconnected or overlap each other in a continuous sequence of units. At a local level, we gave evidence of the inappropriateness of linear superposition for the simulation of coarticulation between two notes. The transition revealed a compound interaction between the two units, by lengthening the end of the first and adding extra movement (the increase for the pressure peak) to the beginning of the second. At a contextual level, results establish the hypothesis that the surrounding units exert influence on the local coarticulation. Anticipatory behavior and memory effects are the most plausible

cause for regulating differently the three conditions. On one hand, anticipation in the preparation of a staccato note before its instantiation, as shown by the difference between *l2nl* and *l2s* curves. On the other hand, memory effects for the use of previous legato when switching to non-legato articulations, as evident from the comparison between *nl2nl* and *l2nl* curves. Performance memory has proved to transfer between consecutive tasks [29]. Here, however, the performer seems to “overreact” to a possible transfer of performance skill, emphasizing the abruptness in transition when departing from a legato context, and performing what would behave as non-legato transition with increased overshoot.

*Acknowledgments* We would like to thank Fabien Norbert, Gérard Bertrand and “Instrumental Acoustics” team for their help in this study.

## References

1. Hardcastle, W.J., Hewlett, N.: Coarticulation: theory, data, and techniques. Cambridge University Press (1999)
2. Gineke, t.H., Hendriks, P., Andringa, T.: Why dont you see what i mean? prospects and limitations of current automatic sign recognition research. *Sign Language Studies* **6**(4) (2006) 416–437
3. Farnetani, E.: Coarticulation and connected speech process. (1999)
4. Fowler, C.: Coarticulation and theories of extrinsic timing control. *Journal of Phonetics* **8** (1980) 113–133
5. Engel, K., Flanders, M., Soechting, J.: Anticipatory and sequential motor control in piano playing. *Experimental brain research* **113**(2) (1997) 189–199
6. Loehr, J., Palmer, C.: Cognitive and biomechanical influences in pianists’ finger tapping. *Experimental brain research* **178**(4) (2007) 518–28
7. Dahl, S.: The playing of an accent - preliminary observations from temporal and kinematic analysis of percussionists. *Journal of New Music Research* **29**(3) (2000) 225–234
8. Rasamimanana, N.H., Bevilacqua, F.: Effort-based analysis of bowing movements: evidence of anticipation effects. *The Journal of New Music Research* (in press) (2009)
9. Gody, R., Jensenius, A., Nymoen, K.: Production and perception of goal-points and coarticulations in music. In: *Proceedings of Acoustics '08*. (2008)
10. Fletcher, N.H., Tarnopolsky, A.: Blowing pressure, power, and spectrum in trumpet playing. *Journal of Acoustical Society of America* **105**(2) (1999) 874–881
11. Bouhuys, A.: Lung volumes and breathing patterns in wind-instrument players. *Journal of Applied Physiology* **19**(1) (1964) 967–975
12. Campbell, M., Bromage, S., Chick, J.: Attack transients on lip reed instruments. *Journal of Acoustical Society of America* **117**(4) (2005) 2477
13. White, E., Basmajian, J.: Electromyography of lip muscles and their role in trumpet playing. *Journal of Applied Physiology* **35**(6) (1973) 892–7
14. Berger, K.: Electromyographic recording during wind instrument performance. *Annals of the New York Academy of Sciences* **155**(1) (1968) 297–302
15. Merriman, L.C., Meidt, J.A.: A cinefluorographic investigation of brass instrument performance. *Journal of Research in Music Education* **16**(1) (1968) 31–38
16. Cooper, G., Meyer, L.B.: *The Rhythmic Structure of Music*. University of Chicago Press, Chicago, Illinois (1960)

17. Ramsay, J.O., Silverman, B.W.: *Functional Data Analysis*. Springer (2005)
18. Ramsay, J., Hooker, G., Graves, S.: *Functional Data Analysis with R and MATLAB*. Springer (2009)
19. Koenig LL, Lucero JC, P.E.: Speech production variability in fricatives of children and adults: results of functional data analysis. *The Journal of the Acoustical Society of America* **124**(5) (2008) 3158–3170
20. Yan, J.H., Thomas, J.R., Stelmach, G.E., Thomas, K.T.: Developmental features of rapid aiming arm movements across the lifespan. *Journal of Motor Behavior* **32**(2) (2000) 121–140
21. Coward, H.: *Choral Technique and Interpretation*. H.W. Gray (1914)
22. Bizzi, E., Giszterb, S.F., Loeba, E., Mussa Ivaldi, F.A., Saltiela, P.: Modular organization of motor behavior in the frog’s spinal cord. *Trends in Neurosciences* **18**(10) (1995) 442–446
23. Bizzi, E., d Avella, A., Saltiel, P., Tresch, M.: Modular organization of spinal motor systems. *The Neuroscientist* **8**(5) (2002) 437–442
24. Ramsay, J.O., Munhall, K.G., Gracco, V.L., Ostry, D.J.: Functional data analyses of lip motion. *Journal of Acoustical Society of America* **99**(6) (1996) 3718–3727
25. Sanger, T.D.: Human arm movements described by a low-dimensional superposition of principal component. *The Journal of Neuroscience* **20** (2000) 1066–1072
26. K., K., O., T.: Adaptive variability in skilled human movements. *Information and Media Technologies* **3**(2) (2008) 409–420
27. Darling, W., Cooke, W.G.: Movement related emgs become more variable during learning of fast accurate movements. *Journal of Motor Behavior* **19**(3) (1987) 311–331
28. Bouhuys, A.: Airflow control by auditory feedback: Respiratory mechanics and wind instruments. *Science* **154**(750) (1966) 797–9
29. Palmer, C.: Nature of memory for music performance skills. (2006) 39–53