AUDIBILITY AND STABILITY OF ARTICULATORY MOVEMENTS
Deciphering two experiments on anticipatory rounding in French

C. Abry
M. T. Lallouache

Institut de la Communication Parlée, CNRS URA 368
Grenoble, France

ABSTRACT

The proposal that the search for economy principles in movement control could take advantage of the potentiality that some phases and/or components of articulatory movements are *poorly* or *non-audible*, is illustrated by two experiments on anticipatory coarticulation for French rounding.

At a first guess, within speaker variability between sessions, and within session variability for the same speaker, could simply point out that poorly audible protrusion movements through five consonants [kstsk] complexes are «free» to vary, provided they reach their audible goal, in our case the vowel [y].

However, a trend to proportional stability in one session, where the eliciting technique allowed to improve prosodic pausing control, indicated that, in complex clustering tasks, prosodic mastery can counteract effectively the variability induced by complexity, ultimately preventing poorly audible movements – in spite of their increased sluggishness – to become «bumpy».

0.FOREWORD

The topics of the present paper were not directly oriented towards issues currently held in relating speech production and acoustics : they were initially dedicated to movement control, namely *anticipatory rounding* behavior. Of course, the fact that unavoidable and invaluable theories – proponents of different *control spaces* for speech production – were encountered «by-the-way» was not completely beyond our control, since we are currently interacting with colleagues on quantal experience [1] and experiments in learning gestures-from-sounds [2, 9]. So the proposal we are holding that, in search of economy principles for movement control, one could take advantage of the potentiality that some phases and/or components of articulatory movements are *poorly* or *non-audible*, this proposal will not receive here a thorough treatment, but just preliminary support from some evidences of such loose links between production and acoustics.

1.PARADIGM & EXPERIMENTS

On testing competing models of anticipatory rounding – the so-called *look-ahead* (LA), *time-locked* (TL), and *hybrid* (H) models, following the procedure set up by Perkell [11] – with French data, we had the opportunity to observe a lot of variational behavior, for the *same* speaker, between and *within* sessions (2 sessions spaced by about 6 months). Since we will focus our attention on the variability *within the same task*, examples presented here will not include those obtained by manipulating the number of consonants and the position of the juncture within [i–>y] transitions. Thus we chose deliberately the most complex case : the «mirror» sequence [...ikstsky...] in *Ces deux Sixte sculptèrent* «These two Sixte [popes] sculptured» (the classical French *sinistre structure* [4] appeared to be unpronounceable without schwa, even by Northern speakers). A mean representative of [iky] tokens was selected only as a «control» reference.

Four illustrations are shown on Figs 1 & 2, displaying upper lip protrusion time functions, with kinematic events, which were detected manually on instantaneous velocity and acceleration functions (derived from cubic spline functions fitted to raw measurements on each 20 ms video field; for more details see [10]). An *obstruence interval* was determined on the synchronized audio signal (sampled at 16KHz) by detecting [i] offset and [y] onset, corresponding respectively to the disappearance and appearance of a clear vocalic formant structure. Among parameters other than upper lip protrusion (this one is chosen here mainly for comparison with [11]), image processing enabled us to track between-lips area from front views. Additional cepstral formant tracking and sections were checked when needed.

Events of ten samples for each session are overlaid on Fig. 3. Perkell's conventions have been essentially adopted, though this presentation suffers from statistical artefact, i.e. part-whole correlations ([5] : this has been taken into account, at least in part, in further work, pers. comm.). Protrusion kinematic events are referred to [y] onset (upper plot) or [i] offset (lower plot). This mirror image is adopted simply to bring out all the possible correlations (regression slopes and intercepts being entirely redundant). Regressions traced here are only those which reached significance at $p<0.01$. Symbols are of course aligned vertically for each token. The minimum protrusion events in the *1st session* are linked by the vertical lines.

Results will be first discussed qualitatively (movement *profiles*); then quantitatively (*dates* of events).

2.VARIABILITY & STABILITY OF MOVEMENT PROFILES

During the 2nd session, all three movement *profiles* – characterized by Perkell [11] – were observed : (i) a unique ramp, with a nearly constant slope, i.e. a *one-phase* gesture (a variant of it, corresponding to a *bell shape velocity profile*, is shown on Fig. 2, upper part); (ii) no (or a weak) movement phase, followed by a rather steep start of protrusion, i.e. a *two-phase* protrusion (a preretracted variant is seen in Fig. 2, middle part); (iii) an initial ramp-like phase, followed by a steeper phase, i.e. *two phases* again (the example shown on Fig. 2, lower part, is an extreme case, since it shows *temporal overshoot* beyond the end of the vowel, *into the following [l]*; it will be discussed later).

In contrast to this large variety of profiles, the 1st session displayed almost exclusively movements of the type (ii), as shown on Fig. 1.

3.VARIABILITY OF KINEMATIC EVENTS

Inspecting Fig. 3, we must agree with Perkell [11 : 280] in rejecting all three "strong versions" of LA, TL and H models. The protrusion «beginning» (conventionally : minimum value) was not locked at the offset of the unrounded first vowel (LA; contrary to [4] for French), nor fixed relative to the onset of the rounded second one (TL; its peak acceleration neither, thus rejecting H). The only consistent fact through both sessions was that peak protrusion was locked *about* the onset of [y] (with one notable exception, see Fig. 2, lower part; its events are marked by + on Fig. 3). This means simply that no plateau-like and/or spatially overshooted anticipations were observed.

So our data exemplify all the three main types of *profiles*, but they violate all three models with respect to their predicted *dates*.

4.ELICITING TWO STRATEGIES

Such negative results have puzzled students in coarticulation for years. And for our part, we were about to give up and to come to a conclusion about variability *per se*, when we suddenly remembered (*post hoc*!) that we had used innocently two different eliciting techniques to make produce such complex consonant chains as [...kstsk...]. In the 2nd session, instruction was to repeat the sentence, prompted with a long pause : «Ces deux Sixte...sculptèrent», *as a whole*; whereas in the 1st one the subject had to repeat, when prompted, the noun phrase : «Ces deux Sixte», linking up with : «Ces deux Sixte sculptèrent». This possibility to «prime» the action could be compared with a trial approach before jumping the hurdle, allowing to size it up. In our case the effect was a better movement «chunking» (corresponding to prosodic parsing) in this 1st performance, which is visible looking at converging cues, such as overall longer obstruence intervals (Fig. 3) and less elisions of the closure phase for [t] (Fig. 1 vs Fig. 2).

Since it has been recently reemphasized that any lengthening effect on the obstruence interval – such as slow rate, stress and number of consonants – would lead to more complex patterns of protrusion movement, allowing «individual gestures [vocalic *and consonantal* protrusions] to emerge as distinct entities» [6 : 186], it is interesting to note that the greatest stability in movement profiles was obtained in the 1rst session, which brought about the longest and more carefully pronounced tokens. Testing the *proportionality* (for this procedure, cf. [8]) for the different kinematic events, it was found that, within the obstruence interval, maximum velocity was relatively stable in the 1rst session (33% ≤ protrusion lead ≤ 47%), whereas it drifted towards [y] onset in the 2nd one, as obstruence duration increased (from 70% to 4% lead).This last trend was followed by acceleration events, a behavior which corresponds, in the part-whole presentation, to the steepest regression lines on Fig. 3 (lower part).

An interpretation of the proportional behavior in the 1st session could be that *more pausing lengthens rounding anticipation* (see here [7], for evidence in silent pauses, for French and for the same speaker), hence increasing *both* phases : the first one that corresponds to a rather clear realization of [kst≠] (without a full silent pause, of course), with no, slow or just starting protrusion; and the second to the deceleration phase towards peak protrusion.

The noteworthy findings concerning with the 2nd session are that – due to the eliciting technique – pausing was not as easily controlled, which led to unsteadily junctured products. But we have no suggestion to explain the fact that this instability caused the maximum velocity event to draw nearer to [y] onset... which – as noted above – led peak protrusion to occur *right in the middle of the following [l]*, in one extreme case of temporal overshoot.

5.AUDIBILITY & STABILITY

For this acoustically critical case (Fig. 2, lower part) – where, after all, the peak protrusion stays in the domain of the syllable – we checked formant values and found, as early as the first periods of the vowel, maximum energy at about 2000Hz (this energy concentration is normally at about 3000Hz for this and other French speakers for [i] [12]). Looking at between-lips area it appeared that, about at the beginning of the vowel, it was well on the way to reach its minimum value. On the other hand, since this second impulse in narrowing began with the release of the [k], the result was that its burst kept an [i]-like coarticulation. Having not performed, at present time, systematic identification tests, we can only refer to a gating experiment done earlier on French [kstR], etc. [3], where it was found that listeners could hardly identify the following vowel until they were delivered at least half of the last [R] consonant. So whatever the perceptual effects of a possible conflict between the cues of the burst and the vowel may be, it is likely that rounding anticipation is rather *lately* perceived in such cases, at least *auditorily*, if not *visually* [7].

From a *motoric* point of view, the problem remains to explain why during certain poorly audible phases of articulatory movements (under *don't care conditions* in learning gestures-from-sounds) velocity profiles are classically bell-shaped (i) & (ii), whereas during other such phases, they are double peaked (iii) (or even resemble those of a bow (i) !).

We hope to have shown that rate is not the only factor capable of smoothing articulatory trajectories – including poorly audible ones – as it increases. In the case of a complex clustering task (hence «bumpy» in nature in its execution, and this not necessarily because it would prevent an aggregation of idiosyncratic gestures [6]), an improvement in prosodic pausing control can counteract effectively the variability induced by complexity, ultimately preventing poorly audible movements – in spite of their increased sluggishness – to become «bumpy».

• Thanks to S. Maeda for trying to improve our English... and ideas.

6.REFERENCES

[1]ABRY, C., BOË, L.-J. & SCHWARTZ, J.-L. (1990), "Plateaus, catastrophes and the structuring of vowel systems", *Journal of Phonetics*, 17, 47-54.

[2]BAILLY, G., LABOISSIÈRE, R. & SCHWARTZ, J.-L. (1991), "Formant trajectories as audible gestures : an alternative for speech synthesis", *Journal of Phonetics*, 19, 9-23.

[3]BENGUÉREL, A.-P. & ADELMAN, S. (1976), "Perception of coarticulated lip rounding", *Phonetica*, 33, 113-126.

[4]BENGUÉREL, A.-P. & COWAN, H.A. (1974), "Coarticulation of upper lip protrusion in French", *Phonetica*, 30, 41-55.

[5]BENOIT, C. (1986), "Note on the use of correlations in speech timing", *J.A.S.A.*, 80, 1846-1849.

[6]BOYCE, S.E., KRAKOW, R.A., BELL-BERTI, F. & GELFER, C. E. (1990), "Converging sources of evidence for dissecting articulatory movements into core gestures", *Journal of Phonetics*, 18, 173-188.

[7]CATHIARD, M.-A., TIBERGHIEN, G., CIROT-TSEVA, A., LALLOUACHE, M.T. & ESCUDIER, P. (1991), "Visual perception of anticipatory rounding during acoustic pauses : a cross-language study", *Proc. of the XIIth Congress of Phonetic Sciences*, Aix-en-Provence.

[8]GENTNER, D.R. (1987), "Timing of skilled motor performance : tests of the proportional duration model", *Psychological review*, 94, 255-276.

[9]LABOISSIÈRE, R. , SCHWARTZ, J.-L. & BAILLY, G. (1990), "Motor control for speech skills : a connectionist approach", in D.S. Touretzky, J.L. Elman, T.E. Sejnowski & G.E. Hinton (Eds), *Connectionist models, Proc. of the 1990 Summer School*, Morgan Kaufmann Publishers, Palo Alto, CA, 319-327.

[10]LALLOUACHE, M.T. (1990), "Un poste 'visage-parole'. Acquisition et traitement de contours labiaux", *Actes des XIIIèmes Journées d'Études sur la Parole*, Montréal, 282-291.

[11]PERKELL, J.S. (1990), "Testing theories of speech production : implications of some detailed analyses of variable articulatory data", in W.J. Hardcastle & A. Marchal (Eds), *Speech production and speech modelling*, Kluwer Academic Publishers, Dordrecht, Boston, London, 263-288.

[12]SCHWARTZ, J.-L., ABRY, C., BEAUTEMPS, D., ESCUDIER, P., MANTAKAS, M. & PERRIER, P. (1989), "Stratégies interindividuelles de production d'équivalences perceptives", in H. Méloni (Ed.), *Speaker variability and specificity*, Marseille-Luminy, 78-86.

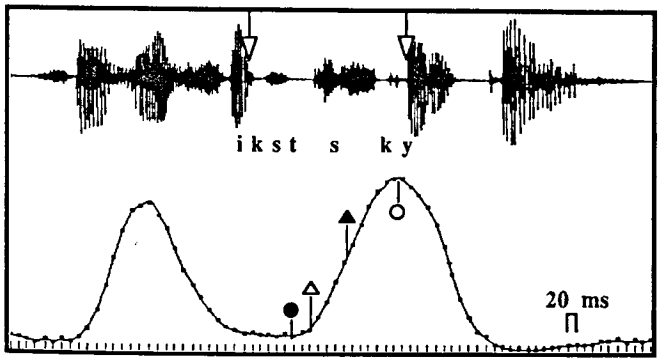

Fig. 1 .– Typical movement profile for upper lip protrusion obtained during the 1st session (cubic spline functions fitted to raw data points; filled circle: min. protrusion; open triangle: max. acceleration; filled triangle: max. velocity; open circle: peak protrusion; white arrows indicate obstruence interval).

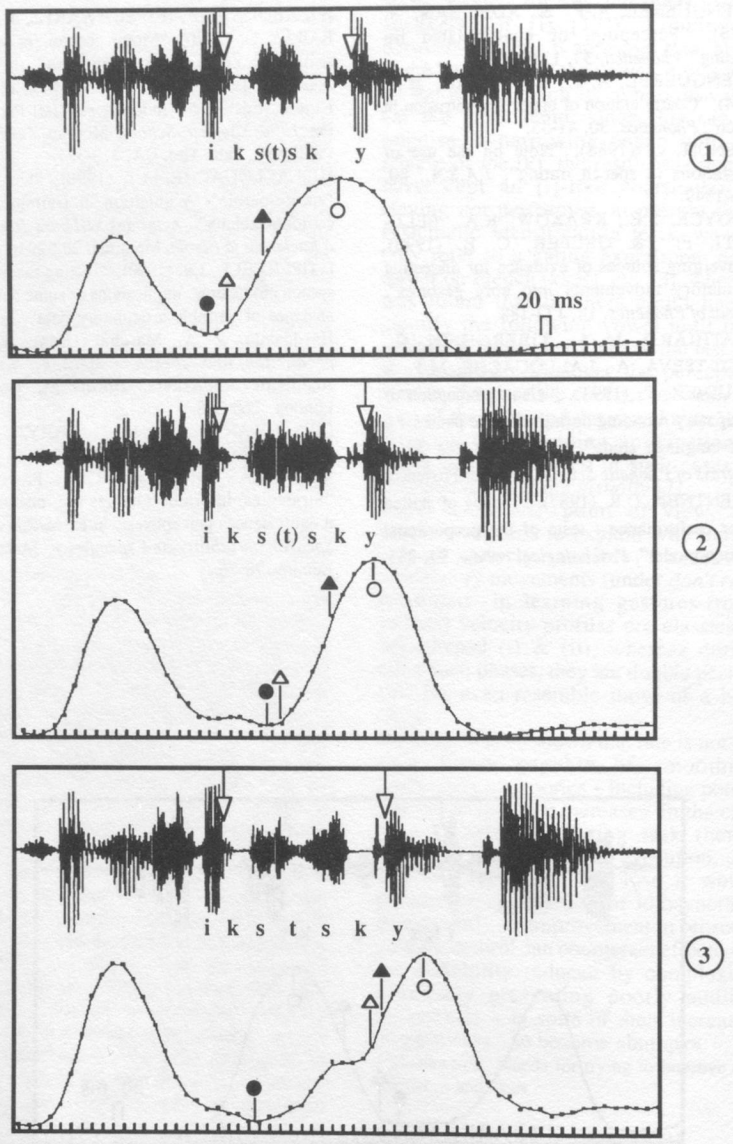

Fig. 2 .– Three types of movement profiles for upper lip protrusion obtained during the 2nd session (cubic spline functions fitted to raw data points; filled circles: min. protrusion; open triangles: max. acceleration; filled triangles: max. velocity; open circles: peak protrusion; white arrows indicate obstruence interval).

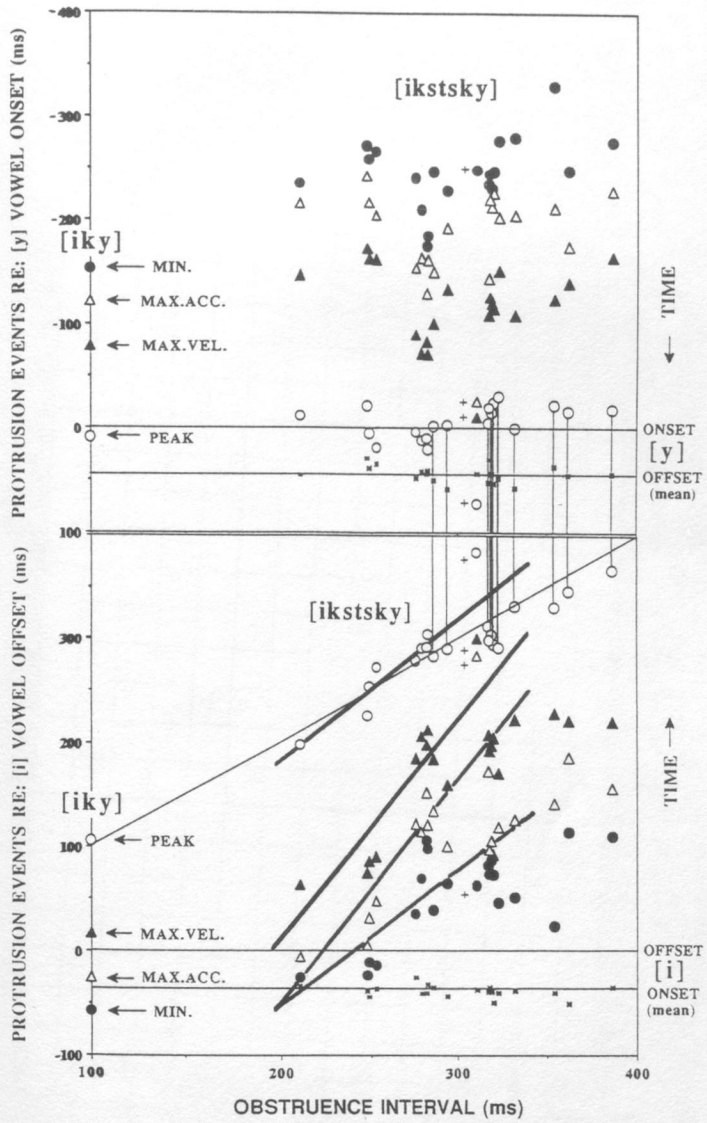

Fig. 3 .– Overlaid plot of kinematic events (same symbols as in Figs 1-2, plus cross-stars indicating [y] offsets and [i] onsets) vs obstruence interval duration. Events are referenced to [y] onset (upper part) or to [i] offset (lower part). Peak protrusion symbols of the 1st session are linked by the vertical lines. Only regression lines significant at $p < 0.01$ are displayed (lower part, thick lines; thin oblique line : $y = x$) : they correspond to the 2nd session. Symbols prefixed by + are events belonging to the temporally overshooted sample edited on Fig. 2 (lower part). [iky] mean representative "control" events (for both sessions) are actually slightly outside the plot (mean obstruence interval = 96 ms). See text.