A kinematic study of prosodic boundary effects on /i/ articulation in French.

Marija Tabain[†], Pascal Perrier[‡] and Christophe Savariaux[‡]
† Macquarie University, Sydney, Australia
‡ Institut de la Communication Parlée, Grenoble, France
E-mail: mtabain@maccs.mq.edu.au, perrier@icp.inpg.fr, savario@icp.inpg.fr

ABSTRACT

This study presents kinematic data for the vowel /i/ at several different prosodic boundaries in French. It follows on from previous work presented in Tabain (in press), where the vowel /a/ was shown to be highly affected by the strength of the prosodic boundary.

Consistent with previous data showing minimal contextual variability in the production of /i/, our results show minimal effects of prosodic boundary strength on tongue and jaw movement for /i/. However, durational data show the same pattern as is observed for all other speech sounds. We suggest that since a very precise production of /i/ is important for correct perception of this vowel, the speaker is not free to mark the prosodic hierarchy in terms of supralaryngeal articulation in the same way as it is marked for /a/. For this reason we suggest that duration is still the primary marker of prosodic boundaries in French.

1. INTRODUCTION

It has become clear in recent years that there are systematic strategies used by speakers to delineate prosodic boundaries at the supralaryngeal level (Fougeron & Keating, 1997). This is in addition to intonational, amplitude and durational cues to prosodic boundaries. The effects of the strength of the prosodic boundary on individual speech sounds can be summarized as follows: consonants are hyper-articulated at stronger prosodic boundaries (such as the Utterance or Intonational phrase), and hypo-articulated at weaker prosodic boundaries (such as the Word or Syllable). These results have been confirmed in a variety of articulatory studies (e.g. Fougeron & Keating, 1997; Byrd, 2000; Fougeron, 2001). However, not all speech sounds are equally affected by the strength of the prosodic boundary. Fougeron (2001) has shown that /s/, a segment which is typically highly resistant to coarticulation, is also much less variable across prosodic contexts than other consonants (at least in terms of tongue-palate contact).

Given Fougeron's results for /s/, one could hypothesize that speech sounds which are highly resistant to coarticulation are also less affected by different types of prosodic boundaries, since they exhibit less contextual variability in general. To test this hypothesis, we examine the vowel /i/ at different prosodic boundaries, since /i/ has been shown to be much more resistant to coarticulation than /a/ (Recasens 1999). In previous work using EMA (electromagnetic articulography - Tabain, in press) we showed that the vowel /a/ is highly affected by the strength of the prosodic boundary, with a lower Tongue and Jaw position at stronger prosodic boundaries, and a higher position at weaker boundaries. Duration of the vowel was also longer at stronger boundaries.

An interesting result from our previous study concerned the data for only one of the three speakers: for this speaker (GR), data for the Jaw at the Utterance boundary tended to pattern with the Word and Accentual phrase data, i.e. the /a/ was centralized rather than being hyperarticulated at the Utterance boundary. We interpreted this as articulatory declension, whereby supralaryngeal articulations at the end of the Utterance become progressively more "lax" (Vayra & Fowler 1992). Such a strategy is contradictory to the predictions made by the articulatory prosody model; we will therefore examine whether or not this pattern is repeated for the same speaker's /i/ data.

In the current study we use the same measures and techniques as were used in the Tabain (in press) study, in order to describe the effects of prosodic boundary on the vowel /i/ followed by the same consonants and uttered by the same speakers. However, our hypotheses for /i/ of course differ to those for /a/. Following Erickson (2002), who showed that Jaw position was lower and Tongue position higher and/or more forward in stressed as opposed to unstressed syllables containing /i/, we expect that Tongue position will be higher/more forward, and Jaw position lower, at stronger prosodic boundaries. The higher/more forward Tongue position reflects a more hyperarticulated front vowel, and the lower Jaw position is believed to indicate a more sonorous vowel. It might be noted that Harrington, Fletcher & Beckman (2000) found results similar to those of Erickson in terms of the apparently contradictory strategies of the tongue and jaw.

2. METHOD

Three native speakers of metropolitan French (two male and one female) were recorded in a sound-treated room at ICP, Grenoble. Articulatory (EMA) and acoustic data were

ISBN 1-876346-48-5

recorded simultaneously and time-synchronized, although only articulatory results are presented here. The EMA data were recorded at 200Hz. Transducers were placed on the Upper Lip, Tongue Tip, Tongue Body and Jaw. A reference transducer was placed on the gums above the upper teeth. Only the Tongue Body and Jaw data will be presented here.

Both acoustic and articulatory data were labelled using EMU (Cassidy & Harrington, 2001) and the R statistical package (Ihaka & Gentleman, 1996). All analyses of the data were carried out using these same tools.

The follow signal processing was carried out prior to labelling: (1) x- and y-data were smoothed using a Lowess filter (set to 1/3 the length of the analysis window); (2) mean values for the reference transducer were subtracted from values for the 4 movement transducers; and (3) the resulting data were rotated according to the measured occlusal plane of the speaker. Articulatory data were labelled automatically and hand-corrected. Velocity was calculated as the first differential of the smoothed displacement signal, and this first differential was smoothed using a median filter.

x- and y-targets for /i/ were located based on zero crossings in the velocity traces for the TB and Jaw data. For the movement duration data reported below, the following points were located based on zero crossings in the y-plane velocity trace: (1) the TB minimum for /a/ in the word "Papi", and (2) the TB maximum for /i/ in the word "Papi". Duration of the movement was taken as the difference between these two points in time.

In the present study, the prosodic hierarchy is assumed to be: Utterance > Intonational phrase > Accentual phrase > Word > Syllable. Stimuli consisted of 5 sentences, based on Fougeron (2001) each containing a prosodic boundary of interest between the 4^{th} and 5^{th} syllables (5^{th} and 6^{th} syllables in the case of sentence 5). These sentences were (with the type of prosodic boundary listed in brackets):

1. Paul aime Papi. **B**iba les protège en secret.

(Utterance)

"Paul loves Grandpa. Biba looks after them in secret"

2. Le pauv' Papi, **B**iba et Paul arriveront demain.

(Intonational phrase)

"Poor Grandpa, **B**iba and Paul are coming tomorrow"

3. Tonton, Papi, Biba et Paul arriveront demain.

(Accentual phrase)

"Uncle, Grandpa, Biba and Paul are coming tomorrow"

4. Paul et Papi **B**iba arriveront demain.

(Word)

"Paul and Grandpa Biba are coming tomorrow"

5. Tonton et Papi**b**i arriveront demain.

(Syllable)

"Uncle and Papibi are coming tomorrow"

The consonant in bold was varied to be one of /b d g f s ʃ/. There was thus a total of 30 different sentence stimuli (5 prosodic contexts * 6 consonants). Two of the speakers (the female and one male) produced 10 repetitions of the corpus, giving a total of approximately 300 utterances. The other male speaker produced 9 repetitions, giving a total of approximately 270 utterances. Speakers were encouraged to produce the Utterance boundary with a pause. The Intonational phrase tended to be read with a continuation contour (usually, though not always, without a pause), and the Accentual phrase tended to be read as a list.

2. RESULTS

Figure 1 shows plots of TB trajectories for the vowel /i/ at the end of the word "Papi", taken from the acoustic release of the second /p/ up to the acoustic endpoint of the vowel. Table I presents significance results for the x- and y-targets for these data, as well as for movement duration (for which the data are presented in Table II).

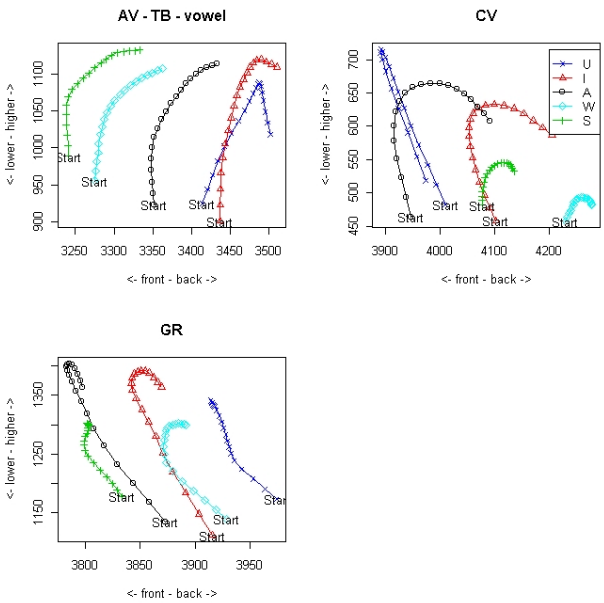

Figure 1: Plots of Tongue Body trajectories for the vowel /i/ at the end of the word "Papi". Data are presented separately for each speaker. Data are time-normalized and averaged across each prosodic context. The beginning of each trajectory, marked "Start", was taken at the acoustic release of the /p/ in /pi/, and the end of each trajectory was taken at the acoustic endpoint of the vowel. Each averaged, time-normalized trajectory is plotted with 20 points equidistant in time. Note that /i/ at the Utterance boundary is followed by a pause, whereas at the other boundaries it is followed by one of 6 different consonants.

Given our hypothesis that the Tongue Body for /i/ should be higher and more forward at stronger prosodic boundaries, the following observations can be made:

(1) speaker AV has a strong effect of the prosodic hierarchy on the x-dimension (front-back), but not in the expected direction (i.e. speaker AV's stronger boundaries are more back and the weaker boundaries more front). There is no significant effect on the y-dimension for this speaker's data (where reference is made to statistical significance results, these are based on a one-way ANOVA with posthoc LSD tests – space limitations preclude the presentation of these results).

(2) speakers CV and GR group their data into two sets: data for {U, I, A} are higher than data for {W, S}. This is in line with the predictions.

(3) there is an ordering within the classes {U, I, A} and {W, S} for speakers CV and GR. Syllable is more forward than Word for both speakers; Accentual is more forward than Intonational for both speakers (although this is not significant for speaker GR); Utterance is highest and furthest forward for speaker CV (in line with predictions); and Utterance is furthest back (and intermediate in height between {I, A} and {W, S}) for speaker GR. This "unusual" patterning for speaker GR's Utterance data may be the realization of articulatory declension for /i/ at the level of the Utterance for this speaker.

AV	d.f.=4,295		
y-target	F = 2.05	n.s.	-
x-target	F = 32.08	p < 0.001	U=I>A>W =S
Duration	F = 107.31	p < 0.001	U>I=A>W =S
CV	**d.f.=4,274**		
y-target	F = 19.19	p < 0.001	U>I=A>W=S
x-target	F = 14.66	p < 0.001	U<I>A<W>S
Duration	F = 191.64	p < 0.001	U=I<A>W=S
GR	**d.f.=4,305**		
y-target	F = 18.18	p < 0.001	U<I=A>W=S
x-target	F = 8.29	p < 0.001	U>I=A<W=S
Duration	F = 102.92	p < 0.001	U=I>A>W=S

Table I: Statistical significance results for measures of the Tongue Body x- and y-targets for /i/ in "Papi", as well as duration of the movement from the /a/ into the /i/. Results are based on a one-way ANOVA with posthoc Least Significant Difference tests for adjacent pairs in the prosodic hierarchy. Alpha has been set at 0.05.

	AV		CV		GR	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
U	308	30.0	311	44.6	319	62.3
I	270	51.6	302	42.6	338	53.8
A	254	51.3	377	57.2	316	30.0
W	175	33.3	192	26.9	224	27.8
S	188	39.1	208	30.0	217	38.7

Table II: Duration of movement from the Tongue Body minimum for /a/ to the TB maximum for /i/ in "Papi". Data for three speakers of metropolitan French.

Table II presents data for the duration of TB movement from /a/ to /i/ in "Papi". With the exception of speaker CV's Accentual boundary data (for which we suspect measurement error due to difficulty in finding the TB minimum for /a/), there is a clear pattern of greater duration of movement at stronger prosodic boundaries. Word and Syllable boundaries pattern together, and the higher boundaries also tend to pattern together.

Figure 2 presents Jaw trajectory data parallel to the TB trajectory data in Figure 1, and Table III presents statistical significance results for these data.

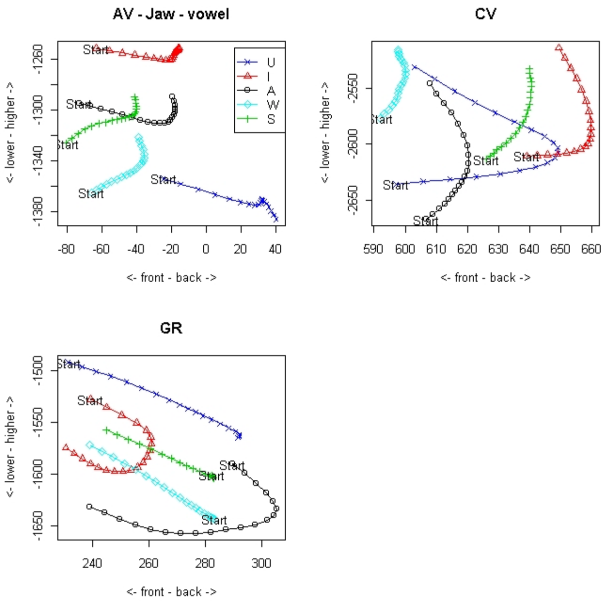

Figure 2: Plots of Jaw trajectories for the vowel /i/ at the end of the word "Papi". Details as for Figure 1.

AV	d.f.=4,295		
/i/ y-target	F = 8.88	p < 0.001	U<I=A=W=S
/i/ x-target	F = 5.67	p < 0.001	U>I=A=W=S
CV	**d.f.=4,274**		
/i/ y-target	F = 1.83	n.s.	-
/i/ x-target	F = 2.28	n.s.	-
GR	**d.f.=4,305**		
/i/ y-target	F = 12.31	p < 0.001	U>I>A=W=S
/i/ x-target	F = 2.077	n.s.	-

Table III: Statistical significance results for measures of the Jaw x- and y-targets for /i/ in "Papi". Details as for Table I.

It can immediately be seen that results for the Jaw are not as clear as those for the Tongue Body. For speaker AV the Utterance boundary data are lower and more back, in accordance with our predictions. However, for speaker GR, the Utterance boundary data are higher than the Intonational boundary data, which are in turn higher than the Accentual boundary data. This is counter to our

ISBN 1-876346-48-5 © 2003 UAB

predictions (it might be noted that the remainder of speaker AV's data, which were not statistically significant, follow a similar pattern to speaker GR's). Speaker CV shows no statistically significant effects on Jaw targets.

4. DISCUSSION

Perhaps contrary to our hypothesis, we did observe some effects of the prosodic hierarchy on the TB data for /i/, and almost no effects for the Jaw (and certainly no effects for the movement into the consonant). These results are different from those for the vowel /a/ reported previously, where there were very clear and consistent effects of the prosodic hierarchy for both the Jaw and TB data. These results are in line with Fougeron (2001), who reported fewer significant differences for /i/ as well as /s/ compared to other speech sounds.

It is interesting, however, that the articulatory data presented here appear to show a larger number of significant effects than do the acoustic data (presented in Tabain, Perrier & Savariaux 2002) from the same recordings. This may be due to quantal effects of the articulatory-to-acoustic relationship (Stevens 1989) or equally to "saturation effects" of the muscular-activity-to-articulation relationship. According to Perkell (1996), such saturation effects should be reflected in greater variability in constriction location, but not in constriction degree. This is due to the fact that the tongue body is stiffened in production of a vowel, so that as increased muscle activity pushes the tongue against the palate, the lateral edges of the tongue brace against the sides of the palate. As a result, cross-sectional area of the constriction (effectively, the area of the palatal vault) does not increase beyond a certain point, and formant values remain relatively stable. The fact that we observed some differences in the x-plane as well as in the y-plane suggests that both quantal effects and saturation effects are relevant to the present data.

Perhaps the main conclusion from the present study is that duration is still the main cue to prosodic structure – at least in French - as evidenced by the much clearer patterns for the duration data than for the TB or Jaw trajectory data. Any supralaryngeal hypo- or hyper-articulation is simply a consequence of the underlying durational specification. We might note that Fougeron (2001) also stated her belief that duration was the main cue to prosodic boundary. It may be suggested that, instead of describing the effects of prosodic boundary on the supralaryngeal articulation of individual speech sounds, we can work from two well-known and well-described principles: (1) coarticulatory resistance, and (2) the encoding of prosodic boundaries via duration.

ACKNOWLEDGEMENTS

We would like to thank our speakers for their time. This work was funded by the Australian Research Council.

REFERENCES

Byrd, D. (2000) "Articulatory vowel lengthening and coordination at phrasal junctures," Phonetica **57** 3-16.

Cassidy, S. & Harrington, J. (2001) "Multi-level annotation in the EMU speech database management system," Speech Communication **33** 61-77.

Erickson, D. (2002) "Articulation of extreme formant patterns for emphasized vowels," Phonetica **59** 134-149.

Fougeron, C. (2001) "Articulatory properties of initial segments in several prosodic constituents in French," Journal of Phonetics **29** 109-135.

Fougeron, C. & Keating, P. (1997) "Articulatory strengthening at edges of prosodic domains," Journal of the Acoustical Society of America **101** 3728-3740.

Harrington, J. Fletcher, J. & Beckman, M. (2000) "Manner and place conflicts in the articulation of accent in Australian English," in M. Broe & J. Pierrehumbert (eds) Papers in Laboratory Phonology V: acquisition and the lexicon Cambridge: CUP, 40-51.

Ihaka, R. & Gentleman, R. (1996) "R: A Language for Data Analysis and Graphics," Journal of Computational and Graphical Statistics **5** 299-314.

Perkell, J. (1996) "Properties of the tongue help to define vowel categories: hypotheses based on physiologically-oriented modelling," Journal of Phonetics **24** 3-22.

Recasens, D. (1999) "Lingual coarticulation," in W. Hardcastle and N. Hewlett (eds) Coarticulation: theory, data and techniques Cambridge: CUP, 80-104.

Stevens, K. (1989) "On the quantal nature of speech," Journal of Phonetics 17, 3-46.

Tabain (in press) "Effects of prosodic boundary on /aC/ sequences: articulatory results," Journal of the Acoustical Society of America

Tabain, M., Perrier, P. & Savariaux, C. (2002) "Prosodic boundary effects on /iC/ sequences in French: acoustic results," *9th Australian International Conference on Speech Science and Technology* Melbourne: Australia, 497-502.

Vayra & Fowler (1992) "Declension of supralaryngeal gestures in spoken Italian," Phonetica **49** 48-60.