Coarticulatory Effects at Prosodic Boundaries: Some Acoustic Results

Marija Tabain, Guillaume Rolland & Christophe Savariaux

Institut de la Communication Parlée - Grenoble, FRANCE.

tabain@icp.inpg.fr

Abstract

Acoustic data are presented from a prosodic database containing data from 3 French speakers. The prosodic boundaries examined are the Utterance, the Intonational Phrase, the Accentual Phrase, and the Word. The aim is to study the interaction of coarticulatory effects with prosodic effects. The vowel /a/ before the prosodic boundary and the consonants /b d g f s ʃ/ after the prosodic boundary are examined. It is found that the vowel duration is greatly affected by the strength of the prosodic boundary, but consonant duration less so. The duration of the fricative consonants is more stable than the stop consonants. Formant values suggest that /a/ is lower and more back the stronger the prosodic boundary, and that the vowel is more likely to reach its low target following a bilabial consonant /b f/. Based on an examination of formant values, the velar stop /g/ appears to have much variability in the front-back dimension. Finally, there is a strong negative correlation between duration and mean velocity of the formant transition, and this effect is strongly related to the strength of the prosodic boundary.

1. Introduction

In recent years there has been an interest in describing not just the pulmonic and laryngeal correlates of stress and of prosodic boundaries, but also in the supralaryngeal effects of these prosodic categories. It has been shown (e.g. [2, 3]) that there are systematic strategies used by speakers at the supralaryngeal level to emphasize stressed syllables and to delineate prosodic boundaries. It is the purpose of this paper to describe some of the acoustic correlates of these strategies.

1.1 Supralaryngeal correlates of linguistic accent

In a landmark EPG study on English, Fougeron & Keating [4] found that there was an articulatory "strengthening" at the edges of prosodic domains, i.e. between the vowel which ends one domain, and the consonant which begins another. They found greater contact for /n/ at the beginning of a prosodic domain - and less contact for /o/ at the end of the previous domain - the greater the strength of the prosodic boundary. The four prosodic domains examined were the phonological word, the phonological (or intermediate) phrase, the intonational phrase, and the utterance. In general, three or four levels were distinguished. However, speakers differed as to what levels were distinguished in this way. Fougeron & Keating also found that the acoustic duration of /n/ was affected by prosodic position. Like linguo-palatal contact, duration increased as the prosodic boundary became stronger; however, duration and linguo-palatal contact were only weakly correlated, since more levels were distinguished by duration.

Fougeron & Keating proposed various explanations for their results. A simple explanation is that increased segment duration leads to achievement of the target. However, the weak correlation between spatial and temporal data for /n/ suggests that refinement of this theory is needed. A second explanation suggests that the increased distance between segments such as /n/ and /o/ is due to articulatory overshoot and/or greater velocity. In both cases, the result is more compression of the tongue tissues, and hence more contact. A final explanation involves increased effort or energy (more "work") on the part of the active articulators. A more forceful articulation is defined by more forceful contraction of the muscles primarily involved in the segment articulation. The three explanations outlined are of course related. Regardless, however, of the underlying strategy, the result is enhancement of the speech signal for the listener, who is able to detect some form of word or prosodic boundary. It remains to be seen to what extent these articulatory strategies reflect an acoustic reality, such as may be measured by formant frequency targets and movements (including distance and velocity – cf. [7]). This is one aim of the current paper. The other aims are discussed in section 1.3 below.

1.2 Prosodic accent in French

The prosodic structure of French is based on the accentual phrase (AP), which is dominated by the intonational phrase (IP). According to Fougeron & Jun [3], the AP has the underlying tonal representation /L Hi L H*/, with a more common phonetic realization being [L H*]. The intonational phrase is marked by a significant final lengthening. Unlike English, French does not have lexical stress; for this reason, there is no possibility of lexical effects and prosodic effects being confounded at prosodic boundaries.

1.3 Hypotheses regarding the interaction of prosodic accent and intrinsic segment resistance to coarticulation

What is not clear from studies such as [4] is to what extent prosodic accent interacts with the intrinsic coarticulatory resistance of individual speech segments. It is known that certain speech sounds exhibit greater sensitivity to context effects from adjacent segments than do other speech sounds. Of particular relevance is work by Recasens [8], which shows that the greater the degree of tongue body raising in the production of a consonant or vowel, the greater the resistance of that segment to coarticulation. However, Tabain [9] has shown that active raising of the tongue body is not the only factor to determine coarticulatory resistance. The manner of articulation of the consonant is important as well. In particular, the precise tongue groove required to achieve maximum turbulence generation for sibilant fricatives constrains the tongue body in such a way that segments such as /s/ and /ʃ/ are extremely resistant to tongue body coarticulation. This leads to the result that for some speakers, the palato-alveolar /ʃ/ is more resistant to coarticulation than the alveolar /s/.

With regard to the present study, it is expected that the fricatives will behave differently than the stops, showing less variability according to prosodic context. It is further expected that the sibilant fricatives will show less variability than /f/. Finally, with regard to the stop consonant, it is expected that the alveolar /d/ will show fewer effects from the prosodic context than either /b/ or /g/, which are more sensitive to coarticulation effects.

2. Method

2.1. Stimuli and recordings

Three native speakers of metropolitan French (two male and one female) were recorded in a sound-treated room. Articulatory (EMA) and acoustic data were recorded simultaneously and time-synchronized. The acoustic data were recorded directly onto a DAT recorder at a sampling rate of

44.1 kHz, and transferred onto PC. Data were subsequently downsampled to 20 kHz.

Stimuli consisted of 5 sentences (based on [2]), each containing a prosodic boundary of interest between the 4th and 5th syllables. These sentences were (with the type of prosodic boundary listed in brackets):

1 Paul aime Tata. **B**aba les protège en secret. (utterance)
2. La pauv' Tata, **B**aba et Paul arriveront demain. (intonational phrase)
3. Tonton, Tata, **B**aba et Paul arriveront demain. (accentual phrase)
4. Paul et Tata **B**aba arriveront demain. (word)
5. Tonton et A**b**aba arriveront demain. (syllable)

The consonant in bold was varied to be one of /b d g f s ʃ/, and the vowel following this consonant was varied to be one of /a i u y œ/ (where "œ" is the schwa-like vowel found in the word "feu"). There was thus a total of 150 sentences (5 prosodic contexts * 6 consonants * 5 vowels). Two of the speakers (the female and one male) produced 4 repetitions of the corpus, giving a total of approximately 600 utterances. The second male speaker produced only 2 repetitions, giving a total of approximately 300 utterances.

The current paper focuses on the relationship between the /a/ at the end of "Tata" and the 3 stop and 3 fricative consonants listed above. Only the acoustic results are presented here. Neither the articulatory results, nor results on V-to-V coarticulation between /a/ and the 5 following vowels, will be reported here. Furthermore, results for the syllable-boundary context (sentence 5 above) will not be presented, since (in pronunciation) the vowel /a/ here is preceded by the vowel /e/, as opposed to the consonant /t/ as is the case with the other 4 sentences (this was due to a planning oversight). This rendered comparison between sentence 5 and the other sentences difficult.

2.2 Labelling and analysis environment

Acoustic data were segmented and labelled according to standard acoustic criteria, using the EMU hierarchical speech labelling tool [1]. The noise following the release of the /t/ in /ta/ was labelled separately and included as part of the /a/ duration. Formants were automatically tracked and hand-corrected. In addition, an F1 target event for the /a/ preceding the consonant of interest was labelled at the F1 peak moving *backwards* in time from the /a/-consonant boundary. Where F1 attained a steady-state after an initial rise, the first pitch period of the steady state moving backwards from the boundary was marked as the F1 target.

All analyses of the data were carried out using the EMU database speech analysis system [1, 5], interfaced with the R statistical package [6].

3. Results

Table I shows mean and standard deviations for vowel duration, consonant duration, and for the percentage of the vowel taken up by the syllable. For each speaker, results are listed separately for each of the prosodic contexts (Ub=Utterance boundary, Ib=Intonational Phrase boundary, Ab=Accentual Phrase boundary, Wb=Word boundary). It can be seen that for speakers AV (female) and GR (male), there is a significant decrease in vowel duration the weaker the prosodic boundary, whereas speaker CV (male) groups the three higher boundaries together, with the Word boundary having a significantly shorter vowel duration. For the consonant durations (stop and fricative contexts are combined here), there is less of an effect according to the prosodic boundary, although speakers AV and GR distinguish the Word boundary from the higher level boundaries. In terms of the percentage of the syllable duration taken up by the vowel, it is clear that the stronger the prosodic boundary, the greater the percentage duration of the vowel. Speaker AV distinguishes all four boundaries, whereas speaker CV distinguishes the Word boundary from the higher boundaries, and speaker GR distinguishes Wb and Ab from Ub and Ib. It is therefore clear that in acoustic terms, the vowel is temporally reduced the weaker the prosodic boundary, whereas the duration of the consonant is relatively invariant.

		Vowel Duration (ms)		Consonant Duration (ms)		Percentage Duration: Vowel/Syllable		
		Mean	S.D.	Mean	S.D.	Mean	S.D.	N
AV	Ub	216.9	21.50	140.2	62.95	62.68	11.78	126
	Ib	163.6	20.74	136.2	41.42	55.49	6.88	126
	Ab	139.2	17.67	133.4	42.59	52.11	8.16	126
	Wb	105.0	11.11	116.1	41.73	49.12	9.20	127
CV	Ub	167.5	15.37	122.4	42.99	59.11	11.04	62
	Ib	169.7	12.94	109.9	40.89	61.94	9.07	63
	Ab	165.2	14.00	110.8	42.70	61.20	9.76	63
	Wb	98.6	7.92	99.0	40.05	51.93	11.11	63
GR	Ub	196.2	17.51	126.9	27.93	61.10	5.71	126
	Ib	179.0	24.47	133.7	28.04	57.53	5.19	127
	Ab	169.8	24.18	127.1	28.76	57.55	5.62	128
	Wb	109.8	10.27	106.0	34.80	52.24	8.86	127

Table I: Acoustic durational data for 3 speakers of metropolitan French (AV=female; CV & GR=male). Results are listed separately for each prosodic boundary (see text for details of boundaries).

		Vowel Duration (ms)		Consonant Duration (ms)		Percentage Duration: Vowel/Syllable		
		Mean	S.D.	Mean	S.D.	Mean	S.D.	N
AV	/b/	148.5	46.06	85.17	18.55	62.45	9.15	84
	/d/	154.5	44.53	91.31	21.78	62.26	7.31	81
	/g/	150.9	49.22	92.26	31.11	61.64	10.31	87
	/f/	160.5	42.53	161.4	27.22	49.35	6.33	80
	/s/	162.0	41.73	175.1	25.90	47.51	5.34	84
	/ʃ/	159.7	42.58	180.2	22.62	46.35	5.28	92
CV	/b/	147.3	30.35	65.75	18.01	68.91	7.38	40
	/d/	150.0	35.35	74.32	27.50	66.85	8.63	39
	/g/	152.1	29.18	73.40	21.03	67.59	5.40	40
	/f/	146.0	33.60	140.6	17.13	50.37	6.38	44
	/s/	154.2	33.02	150.2	13.93	50.12	5.47	48
	/ʃ/	151.0	34.77	146.8	17.20	50.14	5.88	40
GR	/b/	155.6	33.79	105.6	22.97	59.41	4.57	84
	/d/	159.1	35.55	107.8	26.56	59.67	4.17	88
	/g/	155.8	38.27	92.5	29.08	62.94	8.73	89
	/f/	165.3	40.08	143.0	16.49	52.96	6.53	84
	/s/	175.6	39.81	146.3	16.52	53.98	5.30	80
	/ʃ/	171.6	38.04	150.0	15.81	52.79	6.01	80

Table II: Acoustic durational data for 3 speakers. Results are listed separately for each consonant context.

Table II gives the same data as listed in Table I, but with the durational data listed according to the consonant context, rather than the prosodic context. Of note here is the significantly greater duration of the fricative consonants as opposed to the stop consonants, but accompanied by *less* variability, at least for speakers CV and GR. For speaker AV, the longer fricatives are not accompanied by less variability, although her percentage durational data show significantly less variability for the fricatives. These results suggest much more stable timing for the fricatives than for the stops.

Table III shows F1 and F2 values for the vowel /a/ according to prosodic context. The formant values were sampled at the temporal midpoint of the vowel. There is a very clear pattern of increasing F1 values the stronger the prosodic boundary (the only exception is speaker GR's Ub value). There is also a very clear pattern of decreasing F2 values the stronger the prosodic boundary. These results suggest that, in articulatory terms, the /a/ vowel becomes lower and more back the stronger the prosodic boundary.

		F1 (Hz)		F2 (Hz)	
VOWEL		Mean	S.D.	Mean	S.D.
AV	Ub	804.9	47.10	1465	95.51
	Ib	702.8	44.80	1516	96.66
	Ab	693.9	42.84	1590	109.86
	Wb	641.6	38.75	1617	116.28
CV	Ub	639.4	47.77	1259	62.62
	Ib	612.0	24.53	1302	53.00
	Ab	600.4	32.18	1311	71.96
	Wb	504.5	32.27	1435	61.59
GR	Ub	619.6	25.94	1150	42.89
	Ib	658.4	36.27	1161	63.77
	Ab	636.9	38.29	1148	88.38
	Wb	596.7	24.05	1223	111.70

Table III: F1 and F2 vowel data for 3 speakers. Results are listed separately for each prosodic boundary.

		F1 Vowel (HZ)		F2 Vowel (Hz)		F2 Consonant (Hz)	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
AV	/b/	732.8	58.72	1533	117.73	1471	201.94
	/d/	707.0	76.00	1547	110.61	1629	188.92
	/g/	700.3	74.68	1605	143.57	1820	306.24
	/f/	716.3	71.67	1492	108.41	1476	240.25
	/s/	701.5	77.94	1543	97.79	1672	199.79
	/ʃ/	706.5	75.67	1557	114.94	1732	241.74
CV	/b/	598.1	49.47	1344	86.78	1307	151.35
	/d/	580.8	61.94	1321	87.17	1433	158.61
	/g/	579.8	59.19	1349	104.00	1672	292.50
	/f/	598.6	59.47	1308	85.56	1309	170.17
	/s/	587.2	67.00	1328	90.08	1454	167.60
	/ʃ/	587.9	71.50	1313	86.39	1447	192.12
GR	/b/	635.9	37.34	1176	74.15	1139	171.65
	/d/	625.4	35.64	1183	97.73	1370	178.41
	/g/	628.0	38.28	1219	98.41	1421	187.74
	/f/	630.3	40.22	1155	67.30	1178	152.80
	/s/	627.7	40.60	1137	58.46	1310	156.48
	/ʃ/	620.2	41.10	1149	87.55	1480	179.33

Table IV: F1 and F2 vowel data, and F2 consonant data, for 3 speakers. Results are listed separately for each consonant context.

Table IV shows the same data as Table III, but listed according to consonant context rather than prosodic context. In addition, data are provided for the F2 consonant values, sampled at the vowel-consonant boundary (these values would have little meaning if listed according to prosodic context alone, hence their exclusion from the previous table). Several tendencies can be observed here. Regarding F1, there is a slight tendency for this formant value to be higher in the labial consonant contexts /b f/, suggesting that the low vowel target is attained in this context. For two speakers (AV and CV), the vowel F1 in the /b/ context is accompanied by less variability than in the other contexts. It might also be noted that for all three speakers, it is a sibilant fricative context that induces the greatest variability in F1 for the vowel. One could hypothesize that there is a trade-off between the temporal stability for the fricatives (reported above), and a high-low dimension articulatory stability for the low vowel.

Turning to F2, of note for the vowel is the greater F2 value in the velar stop context. This suggests a more forward vowel articulation, consistent with previous data reported for French [10]. With regard to the F2 consonant data (for which the mean values are consistent with values reported previously in the literature) of note is the significantly greater variability in the velar stop. This is true for all three speakers. One could once again hypothesize that this greater variability in F2 at the VC boundary for /g/ suggests greater variability on the front-back contact dimension for this consonant.

Table V presents data for the transition (moving backwards in time) from the VC boundary to the manually-labelled F1 target (described above in the Method section). Data are listed according to the prosodic context, and only stop consonant-context data are included. The distance of the transition in an acoustic space was calculated using a Euclidean distance measure in the F1-F2 plane between data sampled at the VC boundary and at the F1 target (the F2 target was not marked separately). The transition duration was also calculated. Finally, the mean velocity of the transition was calculated by dividing the distance values by the duration values. It can be seen in Table V that there is no clear pattern in the distance measure across the prosodic boundaries, perhaps because the consonant classes are collapsed in this table. However, the transition duration increases significantly the stronger the prosodic boundary (although speakers CV and GR seem to group Ib and Ab together in this case). For the mean velocity data, it appears that the Word boundary context has significantly greater mean velocities than the other prosodic contexts. Within the three other prosodic contexts, there does seem to be a tendency for the velocity to increase the weaker the prosodic boundary. To what extent this greater velocity is correlated with decreasing duration will be examined below. Table VI presents the same results as Table V, but according to consonant context. Of note is the greater variability in transition distance for the velar stop, once again suggesting greater variability in place of articulation.

Finally, Table VII shows correlation data (with significance values) for distance~velocity, distance~duration, and duration~velocity. It can be seen that for all three speakers, there is a significant positive correlation for distance~velocity and a strong negative correlation for duration~velocity. Only one speaker, GR, has a significant positive correlation for distance~duration. Of course, velocity is calculated based on distance and duration values. However, the tendencies observed in Table V for duration to increase and velocity to decrease the stronger the prosodic boundary suggest that these results are to be taken seriously. They do, indeed, echo much work on articulatory strategies used by speakers producing the stressed vs. unstressed contrast in languages such as English.

Some other acoustic cues used to describe our acoustic data are worth mentioning. RMS energy in the fricative and in the vowel was examined and found to have no relationship with the prosodic boundary. We are currently testing to see if the spectral slope of the fricative changes more rapidly in the weaker prosodic contexts (such as the Word); we predict this to be the case based on the formant transition velocity results found for the stop consonants.

		Distance in F1-F2 (Hz)		Transition Duration (ms)		Mean Velocity: Hz/ms	
		Mean	S.D.	Mean	S.D.	Mean	S.D.
AV	Ub	375.2	122.1	123.9	27.58	3.266	1.424
	Ib	214.2	141.9	88.14	18.70	2.571	1.853
	Ab	255.8	178.9	63.67	13.59	4.374	3.382
	Wb	294.9	193.8	42.27	7.91	7.570	5.559
CV	Ub	331.2	98.00	101.7	13.77	3.366	1.437
	Ib	257.8	178.8	63.49	15.72	4.547	3.644
	Ab	292.1	158.6	61.23	17.21	5.540	4.074
	Wb	245.1	147.2	19.25	7.04	15.480	12.708
GR	Ub	281.0	106.0	113.10	22.80	2.623	1.238
	Ib	209.3	150.6	78.39	19.98	2.808	2.263
	Ab	204.0	123.6	79.37	24.67	2.782	1.741
	Wb	219.0	123.9	40.98	8.60	5.542	3.083

Table V: Transition distance, duration and mean velocity data for 3 speakers. Results are listed separately for each prosodic boundary. Only stop consonant contexts are included.

		Distance in F1-F2 (Hz)		Transition Duration (ms)		Mean Velocity: Hz/ms		
		Mean	S.D.	Mean	S.D.	Mean	S.D.	N
AV	/b/	253.9	162.54	76.69	33.61	3.760	3.192	84
	/d/	218.1	142.92	83.89	35.43	2.853	1.870	81
	/g/	376.1	166.83	74.97	36.81	6.688	4.915	87
CV	/b/	149.4	104.03	66.24	32.14	2.130	0.870	40
	/d/	146.2	104.92	62.25	32.42	2.168	1.167	39
	/g/	142.3	115.79	54.81	32.10	2.203	1.195	40
GR	/b/	256.0	76.87	81.61	31.43	3.575	1.626	84
	/d/	227.9	90.59	77.17	33.04	3.345	1.672	88
	/g/	216.2	96.46	74.84	32.68	3.186	1.491	89

Table VI: Transition distance, duration and mean velocity data for 3 speakers. Results are listed separately for each consonant context.

	Speaker	/b d g/
Distance, Velocity	AV	0.7521 **** (t=18.04, df=250)
	CV	0.4637 **** (t=5.66, df=117)
	GR	0.6796 **** (t=14.91, df=259)
Distance, Duration	AV	0.0274 n.s.
	CV	0.0588 n.s.
	GR	0.1676 ** (t=2.73, df=259)
Velocity, Duration	AV	-0.4907 **** (t=-8.91, df=250)
	CV	-0.6026 **** (t=-8.16, df=117)
	GR	-0.4940 **** (t=-9.14, df=259)

Table VII: Correlation values for pairwise comparisons of transition distance, duration and mean velocity for 3 speakers. Only stop consonants were used to calculate results. **** = ($p<0.0001$), *** = ($p<0.001$), ** = ($p<0.01$), * = ($p<0.0167$), n.s. = not significant.

4. Conclusion

Our results have shown that there is an interaction between acoustic segment durations and the prosodic hierarchy, with consonants relatively unaffected by the prosodic boundary and vowels highly affected by the prosodic boundary. In addition, the stability of the fricative consonants is much greater than the stop consonants in terms of duration. In addition, vowel formant values suggest a lower and more back articulation for the /a/ the stronger the prosodic boundary. This low /a/ seemed to be lowest in the context of the bilabials /b f/, where the tongue is most free to anticipate the vowel. With regard to the front-back dimension, the velar stop /g/ seemed to show the most variability on the front-back dimension as suggested by various F1 and F2 measures. Most importantly, there seemed to be a strong inverse correlation between transition duration and velocity, which recalls articulatory results in the literature.

Our current results therefore provide several hypotheses regarding articulatory strategies which we are currently testing using the EMA data recorded at the same time as the acoustic data presented here. Already, however, we have provided a picture of the various acoustic cues which listeners may use to determine the various prosodic boundaries.

5. References

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

[2] Fougeron, C. (1998) Variations articulatoires en début de constituants prosodiques de différents niveaux en français Doctorat de Phonétique, Paris III - Sorbonne Nouvelle.

[3] Fougeron, C. & Jun. S-A. (1998) "Rate effects on French intonation: prosodic organization and phonetic realization" Journal of Phonetics **26** 45-69.

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

[5] Harrington, J., Cassidy, S., Fletcher, J. & McVeigh, A. (1993) "The mu+ system for corpus based speech research," Computer Speech and Language **7** 305-331.

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

[7] Moon, S-J. & Lindblom, B. (1994) "Interaction between duration, context and speaking style in English stressed vowels" Journal of the Acoustical Society of America **96** 40-55.

[8] Recasens, D. (1985) "Coarticulatory patterns and degrees of coarticulatory resistance in Catalan CV sequences" Language and Speech **28** 97-114.

[9] Tabain, M. (2001) "Variability in fricative production and spectra: implications for the Hyper- & Hypo- and Quantal theories of speech production" Language and Speech **44** 58-93.

[10] Vaissière J. (1987) "Effect of phonetic context and timing on the F-pattern of the vowels in continuous speech" Proceedings of the 11th International Congress of Phonetic Sciences 43-46.

Acknowledgements

This work was supported by the French Ministère de l'Education Nationale, de la Recherche et de la Technologie (MENRT) and by the Fondation Fyssen. We would like to thank Pascal Perrier and Hélène Loevenbruck for support, and our speakers for their time.