ARTICULATORY AND ACOUSTIC CORRELATES OF CONTRASTIVE EMPHASIS IN FRENCH: A DEVELOPMENTAL STUDY

Lucie Ménard [1], Hélène Lœvenbruck [2], Christophe Savariaux [2]
& Henrietta Jonas-Cedergren[1]

[1] Dept. of Linguistics, Université du Québec à Montréal
[2] Institut de la Communication Parlée, Grenoble

ABSTRACT: This paper deals with the articulatory and acoustic correlates of contrastive emphasis in French, in children and adults. Nine speakers (three 4-year-old, three 8-year-old children, and three adult males and females) were recorded while producing repetitions of the [baba] sequence in two prosodic contexts: in neutral condition and under contrastive emphasis. The opening (from the consonant into the vowel) and closing (from the vowel into the following consonant) lip gestures were video-recorded and monitored using the ICP tracking system. Lip area and its first derivative were described under both prosodic conditions, in the first and second syllables. Analyses of formant frequencies were also carried out on the acoustic signal. Results show that the effect of contrastive emphasis on articulatory and acoustic measurements is smaller for children compared to adults. Such results shed light on the issue of motor control development during childhood.

INTRODUCTION

The development of speech entails the control of various components of the vocal tract. In adult's speech, the laryngeal and the supralaryngeal mechanisms can be recruited to delineate prosodic boundaries (Tabain, 2003; Fougeron and Keating, 1997) and to convey contrastive emphasis, for instance. However, the non uniform development of these mechanisms suggests that children and adults may use different articulatory strategies in order to achieve linguistic prominence.

SUPRALARYNGEAL CORRELATES OF CONTRASTIVE EMPHASIS IN ADULTS' SPEECH

Several studies have investigated the effects of contrastive emphasis (often referred to as stress or accent but dealing with the same prosodic contrast) on the kinematic patterns of supralaryngeal articulators. In a study of jaw movement in Australian English, Harrington *et al.* (1995) showed that vowels articulated in accented position had higher values of duration, amplitude and opening velocity, compared to unaccented vowels. Increased peak displacement and velocity are also found for jaw movements under contrastive emphasis (Erickson, 1998).

In order to explain the effects of prominence on the kinematic patterns of the supraglottal articulators, De Jong (1995) suggests that stress is related to a local hyperarticulation of the syllable, evidenced by reduced coarticulation. This hyperarticulation account of contrastive emphasis is also proposed by Lœvenbruck (1999) for tongue movement, in a study of the strategies used by two speakers to achieve contrastive emphasis in French.

MOTOR CONTROL DEVELOPMENT

The supralaryngeal articulators affected by contrastive emphasis in adult speech are not mastered by young children, although laryngeal control is well achieved. Indeed, it has been shown that the development of motor control is not uniform. Concerning laryngeal control, Davis *et al.* (2002) report that adult-like fundamental frequency (F0) contours and final lenghtening patterns signalling lexical stress are acquired quite early during infancy. Concerning supralaryngeal articulation, the interarticulatory dynamics of the lips, jaw, and tongue are not adult-like until the end of the first decade of life. Several studies report on the important token-to-token variability in children's speech (Green *et al.*, 2000; Smith and Goffman, 1998, for instance), and on the reduced velocity of the articulatory gestures. On the basis of these results, it can be predicted that the contrast between emphasized and unemphasized syllables will be realized by different articulatory strategies for adults and children. Since the effects of contrastive emphasis can be seen as local hyperarticulation, the stress vs unstress contrast is an interesting experimental condition to study the ability to differentiate syllables on the hyper- and hypo-articulation continuum, and hence, to explore motor control capacities.

METHODOLOGY

Speakers and corpus

Six children (three 4-year-olds and three 8-year-olds) and three adult speakers were audio-visually recorded. All subjects were native speakers of French and had no known speech production or hearing disability. The six children were evaluated as normal for language development by a vocabulary test (EVIP-P Body). The target word was [baba], produced in two prosodic conditions: in a neutral condition and under contrastive emphasis. The target word was embedded in two carrier sentences: “J’ai vu Baba qui mange un gâteau” (*I saw Baba eating a cookie*), in which [baba] is in the neutral condition, and “Non, c’est <u>Baba</u> qui mange un gâteau” (*No, it’s <u>Baba</u> who is eating a cookie*), in which [baba] is emphasized. Note that the second syllable of the word [baba] is located at the edge of a prosodic group, in both conditions. A primary accent is expected in this condition on the last syllable [ba]. These two utterances were part of a story told by the experimenter. In the neutral condition, the speaker was asked to repeat what one of the characters said. In the emphasis condition, the experimenter asked a question to the speaker, while replacing the target word [baba] by another word. The speaker was asked to correct the wrong utterance by using a syntactic extraction (“No, it’s ... who...”) and producing the word [baba] with contrastive emphasis. Between five and ten repetitions of each sentence were obtained, for each subject.

Articulatory and acoustic recordings

Audio and visual data were recorded using the Lip track system developed at ICP. The images were recorded via one fixed camera and stored on a betacam-ST tape. The speakers wore glasses with two calibration pellets, and sat on a chair. Their head was maintained by a helmet, fixed to the chair. Their lips were painted in blue. The images were digitized at 50 Hz (one video image consists of two interleaved frames). The acoustic signal was also digitized at a frequency of 22050 Hz. Using a program designed at ICP, the internal lip contour was automatically detected for the total duration of the target word [baba] and lip area (S) was calculated. Each image was examined by the experimenter, in order to detect computation errors. For some children, the lip contours were hand-corrected, when part of the blue make up had been removed from the lips by the tongue.

Analysis of lip area, peak velocity, and duration of the opening gesture

For each target sequence [baba], three synchronous signals were available: the evolution of S, the first derivative of S (referred to as velocity), and the acoustic signal. A Matlab program, developed at ICP, allowed the visualization of the signals and the labelling of various articulatory and acoustic events. Figure 1 shows the labelling points, based on minimal values, maximal values and zero-crossing values of S and velocity. For the present study, we analyzed the following parameters: (i) the maximal S value reached for [a] in both positions (value of S at t=v3, for the first [a], and at t=v8, for the second [a], (ii) the peak velocity of the opening gesture from [b] to [a] (v2, for the first [a], and v7, for the second [a], (iii) the duration of the opening gesture (the duration of the interval between v1 and v3, for the first [a], and between v6 and v8, for the second [a]).

In addition, the acoustic boundaries for the four segments in [baba] were determined. Formant frequencies at the time of maximal lip area of the vowel were then extracted, using the Burg formant extraction algorithm integrated in the Praat software (http://www.fon.hum.uva.nl/praat/). For all vowels, the results were compared to formant values based on a visual inspection of the spectrogram. In case of important discrepancies between the two methods, parameters for the detection algorithm were readjusted and the analysis was performed again. F0 values were also examined in order to identify the accented syllables. Since for some speakers, the [baba] sequences in the emphasis condition were uttered with emphasis on either the first or the second syllable, we decided to group the utterances according to their accentual patterns. In the present study, only the sequences realized with a perceived accent on the second syllable, for the neutral condition, and a perceived accent on the first syllable, for the emphasis condition, are analyzed. These patterns were by far the most frequent ones. For all the sentences in the emphasis condition, we ensured that the target word [baba] was associated with a high F0 value, with a low F0 plateau on the preceding and subsequent syllables (Touati, 1987).

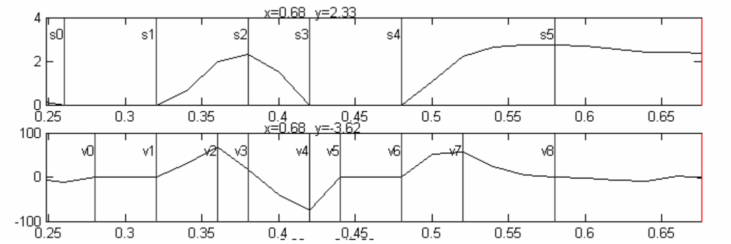

Figure 1. Labelling points on the [baba] sequence. Upper panel: Lip area (S) (in cm^2); lower panel: velocity (first derivative of the S values). S2 and S5 represent maximal lip area reached for the two successive [a]s.

RESULTS

Articulatory data

Mean values and standard deviations of the maximal lip area value, peak velocity and duration of the opening gesture, are plotted in Figure 2, for each speaker. Data were grouped according to position (first and second [ba] syllables) and prosodic condition.

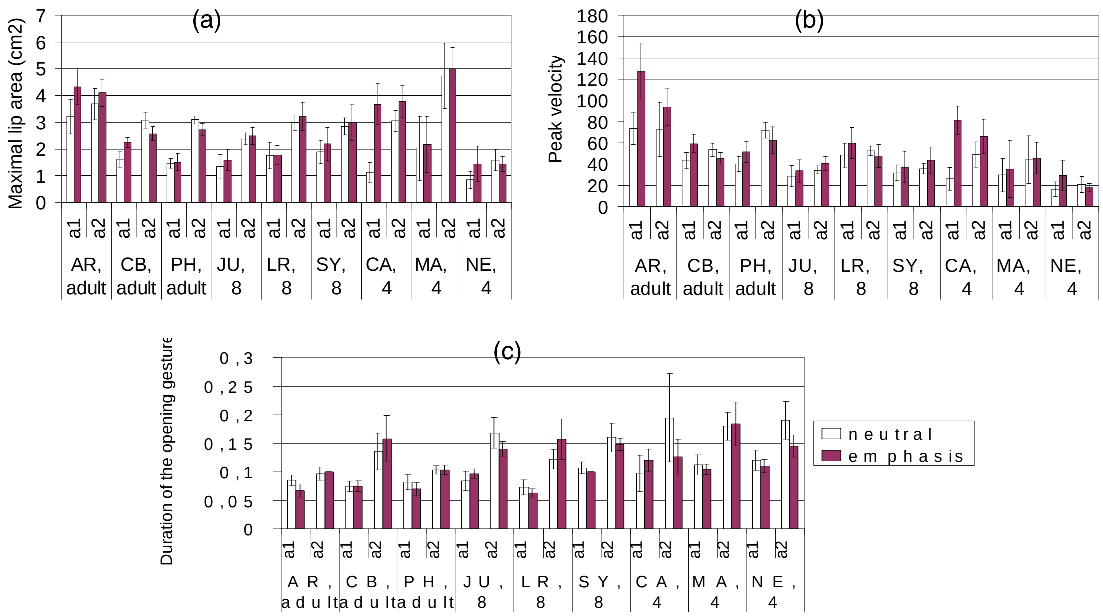

Figure 2. Mean values of articulatory measurements (maximal S, peak velocity and duration of the opening gesture, in sec) for the nine speakers, in neutral condition (light bar) and under contrastive emphasis (shadow bar), for the first vowel (labelled "a1") and the second vowel (labelled "a2") in [baba].

It is striking to observe, first, that the global values of maximal S do not show a tendency to decrease with speaker age. If we consider the size of the articulators for children relative to adults, one could expect that the maximal lip area should be reduced for children. An ANOVA was carried out on the values reported in Figure 2a, and the differences related to age do not reach significance. This result is in line with data presented by Smith and Goffman (1998) who examined lip and jaw displacements for CV sequences in children. Concerning the variation of S according to position, for all speakers but one (AR, adult, in emphasis), [a] is realized with an increased lip area in the second position (labelled "a2" in Figure 2), compared to the first position (labelled "a1" in Figure 2). This pattern shows the effect of the primary accent on the last syllable of the prosodic group (the last [a] is often longer and more open, see Tabain, 2003). As to the effect of condition, in contrastive emphasis compared to the neutral condition, an increased S value for the first [a] is predicted by the hyperarticulation hypothesis. This is the case for all subjects. However, as can be seen in Table 1, the differences depicted in Figure 2a are not always statistically significant. Results of two-way ANOVAs on the S values, with two factors position and condition, reveal that position has a significant effect on the maximal S reached in [a] for all speakers but two (AR, adult and NE, 4). Regarding condition, a significant effect is found for three speakers. For the two adults CB and PH, the interaction is significant. For those speakers, the decrease of S for the second [a] under emphasis (observed in Figure 2a) is significantly different from the variation of S for the first [a]. This may contribute to enhance the prominence of the first syllable, under contrastive emphasis. Most importantly, for all children but one (CA, 4), contrastive emphasis (condition factor) does not have a significant effect on the maximal S value, neither as a principal effect or in interaction with position.

The mean values of peak velocity of the opening gesture from [b] to [a] are plotted in Figure 2b. According to this figure, peak velocity values show a clear tendency to decrease with age, except for the 4-year-old speaker CA. We shall return to this speaker later in the discussion. This global tendency is significant, based on an ANOVA with age as the factor ($F(2, 35)=7.727$; $p<0.01$). Concerning the variation of peak velocity according to condition, for each speaker, results of the ANOVA are presented in Table 1. It can be seen that while for two of the three adult speakers, the prosodic condition has a significant effect on the peak velocity of the opening gesture, values of all children but one (CA, 4) do not show such an effect.

Finally, mean duration values (in sec) of the opening gesture are presented in Figure 2c. The global increasing mean values of duration with age are significant, based on the results of an ANOVA with age as the factor ($F(2,35)=5.223$; $p=0.01$). However, as can be seen from the results of the within-speaker ANOVA (Table 1), all speakers significantly increase the duration of this gesture for the second [a] (row "pos."). However, unlike the pattern revealed by S, duration is not always increased for [a] in the first syllable, under contrastive emphasis, compared to the neutral condition (see Figure 2c). No development trends are noticeable.

Table 1. F-values and p-level of within-speaker two-way ANOVAs performed on measured parameters (S, peak velocity and duration of the opening gesture), with prosodic condition (neutral and emphasis) and position (first and second syllables) as factors. ***: p<0.000; **: p<0,001; *p<0,05.

	AR, ad. F(1,26)	**CB, ad.** F(1,54)	**PH, ad.** F(1,32)	**JU, 8** F(1,18)	**LR, 8** F(1,32)	**SY, 8** F(1,18)	**CA, 4** F(1,16)	**MA, 4** F(1,16)	**NE, 4** F(1,16)
Maximal lip area									
pos	0.24	150.82***	376.14***	39.91***	83.47***	15.78**	18.65**	32.72***	3.72
cond	7.79**	0.66	4.56*	1.27	0.72	0.97	49.12***	0.17	1.50
pos*cond	1.6	60.36***	7.78**	0.17	0.56	0.11	15.47**	0.02	3.73
Peak velocity of the opening gesture									
pos	3.20	0.87	47.39***	3.12	83.47***	1.39	0.40	1.82	0.84
cond	15.40**	4.30*	0.10	2.32	0.77	2.08	35.55***	0.16	1.81
pos*cond	2.90	37.91***	11.20**	0.03	0.77	0.09	10.00**	0.05	4.10
Duration of the opening gesture									
pos	23.99***	105.17***	54.28***	77.53***	112.85***	59.82***	4.08*	44.16***	24.90***
cond	2.67	2.59	2.44	1.12	3.56	2.03	0.76	0.03	6.83*
pos*cond	5.56*	2.55	2.44	7.91*	11.21*	0.17	3.10	0.29	2.77
F1 (bark)									
pos	0.37	0.04	120.61***	5.79*	0.20	0.18	4.05	0.14	0.01
cond	19.72***	117.55***	18.54***	0.35	3.27	0.72	19.04***	0.31	1.09
pos*cond	6.47	204.44***	10.44**	4.38	0.25	0.53	0.39	0.13	4.74
F2 (bark)									

pos	13.91**	3.78	26.31***	0.12	1.72	0.25	0.62	18.30**	7.56*
cond	4.57*	5.68*	49.11***	0.92	6.14*	0.01	4.80*	0.01	2.87
pos*cond	0.15	40.25***	16.74***	4.17	0.01	3.37	2.27	0.32	0.01

Acoustic data

Dispersion ellipses of the formant values (in Bark), in the F1 vs. F2 space, are plotted in Figure 3, for both conditions and positions. While for the adult speakers, [a] in the first syllable is associated with increased F1 values in the emphasis condition compared to the neutral condition, no clear distinction is observed for the children. The effect of emphasis on the second [a] is not consistent across speakers. As can be observed in Table 1, the prosodic condition has a significant effect on F1 for the three adults, whereas five out of six children do not show such an effect.

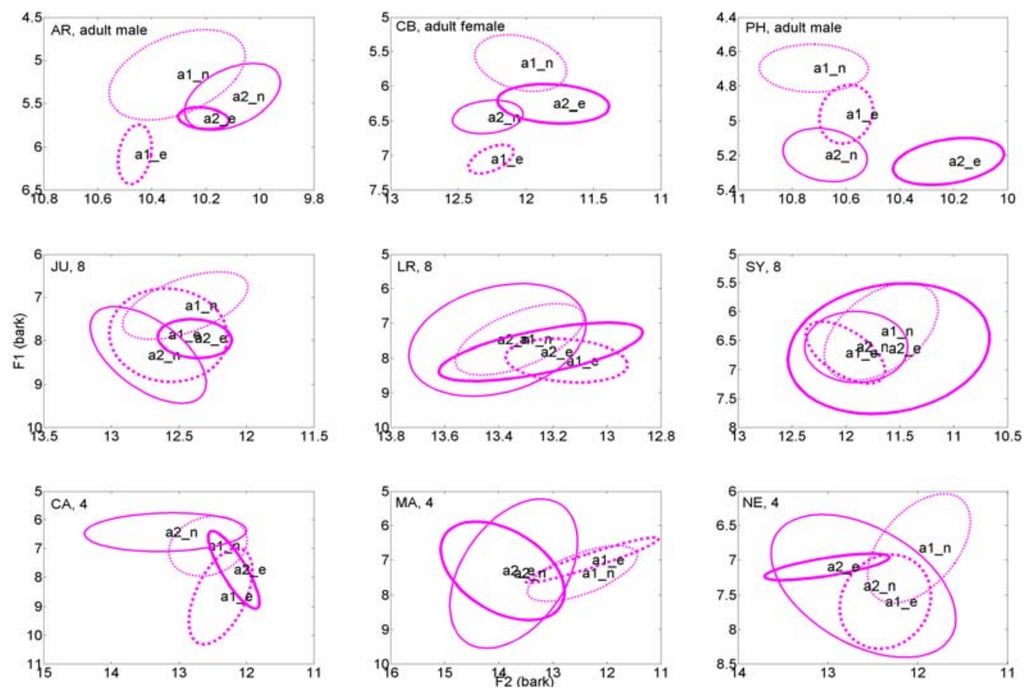

Figure 3. Dispersion ellipses (±1.5 st. dev.) of F1 (y-axis) and F2 (x-axis) for the vowel [a], for the nine speakers, in the first (labelled ‘a1’) and second (labelled ‘a2’) positions, and in neutral (labelled ‘_n’) and emphasis (labelled ‘_e’) conditions. Dotted thin line: first position, neutral condition; Solid thin line: second position, neutral condition; Dotted thick line: first position, emphasis condition; Solid thick line: second position, emphasis condition.

DISCUSSION

The results presented so far suggest that the contrast between emphasized and unemphasized syllables is realized by different articulatory strategies during speech development. For adult speakers, the increased maximal lip area for the first syllable observed in contrastive emphasis with respect to the neutral condition is related to higher peak velocity (see the effect of condition and the interaction of condition and position for peak velocity in Table 1). These articulatory parameters correspond to significant effects on acoustic patterns such as F1. When articulated under constrastive emphasis, the first vowel [a] is located in a more peripheral position (lower) in the acoustic space, compared to the neutral condition. These effects for adults can be interpreted as local hyperarticulation (De Jong, 1995). Contrastively, five children out of six do not produce the vowel under emphasis condition with an increased lip area or opening peak velocity, compared to the neutral condition. For most children, concerning position, even though [a] in the second syllable is produced with significantly greater S and duration of the opening gesture compared to the first position, no resulting significant effect on the mean F1 value is observed.

The case of the 4-year-old speaker CA, is particularly striking. Unlike the other children, this speaker shows important variations in lip area, opening peak velocity, and duration under contrastive emphasis (cf Figure 2). Note that the mean values of opening peak velocity are even higher for this child compared to adults. We believe that this pattern reflects one of the stages, in motor control development, where simple movements (such as the opening/closing gesture) are mastered, while other movements still have to be integrated (Smith and Goffman, 1998).

Nevertheless, on the basis of the global patterns reported here, one conclusion could be that the lack of articulatory and acoustic effects of emphasis is related to the inability of the children to hyperarticulate in this condition. However, the greater values of the lip area depicted in Figure 2 suggest that, in line with Smith and Goffman's observations (1998), children produce relatively large articulatory movements considering the size of their oral structures, compared to adults. Thus, the observed undifferentiated kinematic and acoustic patterns according to condition could rather be due to the inability to reduce or hypoarticulate unaccented syllables. Further articulatory and acoustic analyses of the full utterances are currently in progress in order to further investigate this hypothesis. Finally, another important question particularly relevant to the present work is the fact that lip area, achieved by jaw and lip movements, is not the sole articulatory parameter related to F1. The tongue position also plays an important part in the variation of this acoustic parameter. Studies of tongue positions coupled with jaw and lip positions are needed in order to provide a complete picture of motor control abilities in the realization of contrastive emphasis in children.

REFERENCES

Davis, B. L., MacNeilage, P. F., Matyear, C. L., & Powell, J, K. (2002) "Prosodic Correlates of Stress in Babbling: An Acoustical Study" *Child Developmment* **71** 1258-1270.

De Jong, K. (1995) "The supraglottal articulation of prominence in English : linguistic stress as localized hyperarticulation" *Journal of the Acoustical Society of America* **97** 491-504.

Erickson, D. (1998) "Effects of Contrastive Emphasis on Jaw Opening" *Phonetica* **55** 147-169.

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

Green, J. R., Moore, C. A., Higashikawa, M. & Steeve, R. W. (2000) "The Physiologic Development of Speech Motor Control: Lip and Jaw Coordination" *Journal of Speech, Language and Hearing Research* **43** 239-255.

Harrington, J., Fletcher, J. & Roberts, C. (1995). "Coarticulation and the accented/unaccented distinction" *Journal of Phonetics* **23** 305-322.

Lœvenbruck, H. (1999) "An investigation of articulatory correlates of the accentual phrase in French" *International Congress of Phonetic Sciences* **1** 667-670.

Smith, A. & Goffman, L. (1998) "Stability and patterning of speech movement sequences in children and adults" *Journal of Speech, Language, and Hearing Research* **41** 18-30.

Tabain, M. (2003) "Effects of prosodic boundary on /aC/ sequences: articulatory results" *Journal of the Acoustical Society of America* **113** 2834-2849.

Touati, P. (1987) *Structures prosodiques du suédois et du français: Profils temporels et configurations tonales,* Ph. D. thesis, Travaux de l'Institut de linguistique de Lund, Lund University Press.