Compensation for a labial perturbation: An acoustic and articulatory study of child and adult French speakers

Jérôme Aubin and Lucie Ménard

Laboratoire de phonétique, Université du Québec à Montréal

jerome.aubin@gmail.com, lucie.menard@uqam.ca

Abstract. *This paper deals with production-perception relationships in articulatory compensation. The high French vowel [u] was used to study compensation strategies for a lip-tube perturbation in 6-year-old children and adults. Tubes were placed between the speakers' lips to perturb vocal tract geometry. Ultrasound images and speech signals were recorded for this isolated vowel in three conditions: (1) before lip-tube insertion, (2) with the tube in place, and (3) after removal of the lip-tube. Speakers pronounced ten repetitions of the isolated vowel in each condition. To account for the quality of the compensation, perceptual experiments were carried out and revealed that 6 speakers (4 adults and 2 children) were able to completely compensate for the perturbation. The selectivity of compensatory articulation may be related to biomechanical constraints relative to each speaker, which could prevent them from achieving complete compensation. The results of this study further extend our understanding of the use of feedback in the development of speech production processes and help to specify the representation of the speech task.*

1. Introduction

A natural process in first language acquisition consists of the development of an internal model of speech production. When building this model, the child has to integrate auditory, somatosensory, and proprioceptive information to build and reinforce articulatory-acoustic links (Perkell et al., 2000). Studies of speakers with an incomplete model of speech production allow for a greater understanding of the different components of speech production processes, their nature and their development.

Perturbation experiments have been widely used to study articulatory compensation, a capacity which enables the speech production system to achieve a goal, even if this system is constrained in a certain manner (Lindblom and Sundberg, 1971). Artificial perturbations such as artificial palates or lip-tubes have the advantage of modifying important geometric characteristics of the vocal tract, and hence constraining the speakers to develop new strategies to produce the intended sound. The lip-tube is an appropriate mechanism to affect the lip area in rounded vowels, as it increases the lip area over that found in the typical configuration. Indeed, in the French vowel /u/, the tongue is elevated towards the velopalatal region and the lips are rounded so that the lip area is small (around 0.3 cm^2 for adults, Savariaux et al., 1995). It has been found that, when a 2.5-cm diameter lip-tube keeps them from rounding their lips, French adult speakers required to produce /u/ cannot immediately compensate for the perturbation.

Some speakers are better able to compensate than others, and their compensation strategies involve a backward movement of the tongue towards the palatal region (Savariaux et al., 1995, 1999). Previous studies using a lip-tube with children (Ménard et al., 2004) showed great between-speaker variability but relatively good compensatory abilities. However, unlike studies with adult subjects, lip-tube experiments in children have only investigated acoustic characteristics.

In this paper, we report an experiment in which we assess the acoustic and articulatory effects of labial perturbations on speech produced by 6-year-old children and adult French speakers. Articulatory data were collected using an ultrasound system, a non-invasive technology that allows imaging of the tongue surface (Stone, 2005). A perceptual analysis of acoustic data was carried out to determine the efficiency of the articulatory strategies applied. Two points of view—production and perception—will be examined to shed light on compensation processes from a developmental perspective.

2. Method

2.1. Production

Subjects

Two groups of speakers participated in this experiment. The first one consisted of four children, two boys (5 and 6 years old) and two girls (6 and 7 years old). The second group was made up of four adults, one male and three females, who were 21, 24, 28 and 29 years of age, respectively. A screening procedure ensured that all subjects were free of any hearing or phonological disorder.

Stimuli

In addition to the vowel /u/, which was the one used in previous lip-tube experiments, the three French vowels /a/, /i/, and /y/ were used in this corpus as they delimit the articulatory space. However, only data from the vowel /u/ will be reported in this paper.

Ten repetitions of each vowel were produced in isolation by each speaker in each of the following conditions: before the insertion of the lip-tube (N1), with the tube in place (P), and immediately after removal of the tube (N2). The order of the three conditions was the same for each vowel (N1, P, N2). Vowel order was randomized across subjects.

Material and procedure

Lip-tubes were used to create a greater lip area while pronouncing the rounded vowel /u/. The tubes were made of fine layers of plastic and had a diameter of 1.5 cm and 2 cm, respectively, for children and adults. Several simulations using an articulatory-to-acoustic model of the growing vocal tract (Variable Linear Articulatory Model, developed by S. Maeda) revealed that those diameters were optimal for each speaker group. The effect of the lip-tube was to increase F1 (mean value of 28%), F2 (mean value of 59%), and, to a much lesser extent, F3 (mean value of 1%). Lip-tube lengths were adjusted for each speaker to prevent lengthening of the oral cavity.

The subjects were seated comfortably in a quiet room. Adult subjects had their heads kept still by a helmet, while child subjects put their heads on a headrest. Articulatory data were obtained using a Sonosite 180 Plus ultrasound system. The system's transducer (84-degree curved array) was attached to a microphone stand or was held stable by an experimenter. Both methods involved negligible movement of the transducer relative to the head. The acoustic signal was captured by an unidirectional microphone. Both ultrasound and microphone signals were recorded by a miniDV Panasonic AG-DVC 30 camcorder, in NTSC format.

2.2. Perception

A subset of the corpus produced (50%) served as stimuli for a perceptual experiment. In each of the experimental conditions (N1, P, N2), only repetitions 1, 3, 5, 7 and 9 were selected for the listening experiment. Stimuli were presented once via high-quality headphones in a blocked condition by speaker. Within a block, each speaker's stimuli were randomized and blocks order was randomized. For technical reasons, vowels produced by one adult speaker (AL_a) could not be used in this test. Twenty French listeners (10 males, 10 females) aged from 19 to 25 years old participated in this experiment. Listeners had to identify the perceived vowel and to evaluate its quality on a scale ranging from 1 (bad) to 5 (excellent).

3. Data analysis

3.1. Production

Acoustic data

The audio signals were labeled and vowels were extracted using Praat. Fundamental frequency and formant values were extracted at the center of each vowel. The acoustic signal was downsampled to 22 KHz after low-pass filtering (cut-off frequency of 10 KHz) and an LPC analysis detected up to four formants (the number of poles ranged from 12 to 18). Fundamental frequency measurements were made using the autocorrelation method.

Articulatory data

Ultrasound images corresponding to the center of the vowels were extracted using Adobe Premiere Pro. Tongue surface contours were measured using EdgeTrak, a semi-automatic system for the extraction and tracking of tongue contours (Li et al., 2003). The 100-point contours were exported to a homemade Matlab application, Lingua, which extracts several parameters quantifying tongue contours. In this paper, we will focus on four of them. Figure 1 shows examples of the parameters. For the sake of clarity, the tongue contour is depicted by the solid line, and the palate trace corresponds to the dotted line. Each contour is first reshaped as a triangle using the extremities of the contour as the triangle base (dashed lines). From this triangle, the x and y coordinates of the highest point (on the vertical axis) of the contour are first extracted (point E in Figure 1). Note that this point can be interpreted as an absolute measure of tongue height but does not correspond to a measure of tongue height relative to the palate. Then, measures of tongue curvature and tongue curvature position are determined from points A to D represented in Figure 1. Tongue curvature is defined as the ratio of the distance CD over the distance AB. It can be seen that contour (b) in Figure 1 has a larger value for tongue curvature than contour (a). Tongue curvature position is defined as the ratio of the distance AD over the distance DB. Contour (b) in Figure 1 has a smaller value for tongue curvature position than contour (a). Contrary to the x value of the highest point of the tongue (point E), curvature position is a measure of the position of the mass of the tongue relative to the whole tongue.

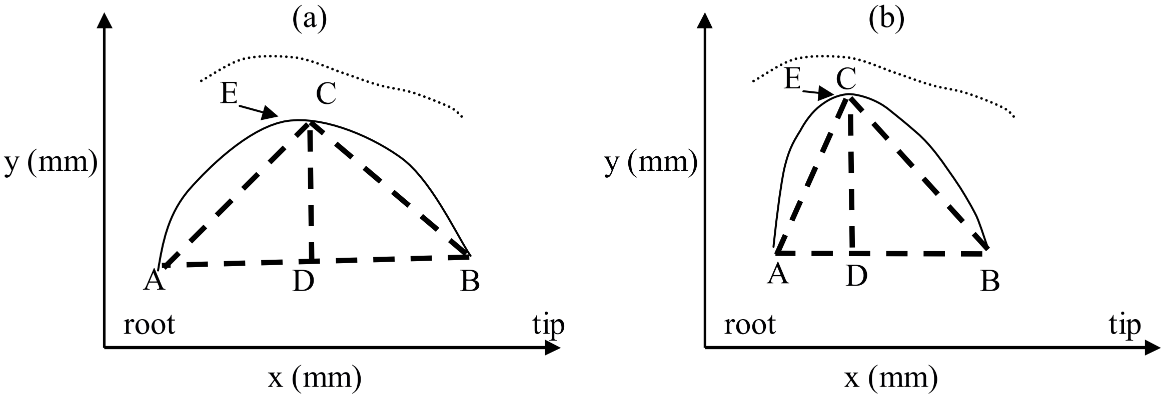

Figure 1. Schematic representations of two sagittal contours in xy coordinates and parameters extracted for the analysis.

These four parameters allowed the characterization of the articulatory space along two separate tongue characteristics: tongue position (x and y coordinates of the highest point of the tongue) and tongue shape (tongue curvature and tongue curvature position). Tongue contours across experimental conditions will be compared along those dimensions.

3.2. Perception

To assess the efficiency of the articulatory strategies, the quality of the best perceived vowel in the perturbed condition (P) was compared to the mean quality of the vowels produced in the normal condition (N1). Quality values of vowels pronounced in the P condition that were equal to or higher than the mean value for the N1 condition were the criteria used to determine that complete compensation had taken place.

4. Results

4.1. Production

The effect of the experimental condition was calculated for each acoustic parameter for each vowel and for each speaker using T-tests. The results are shown in Table 1, where mean formant values in the P condition are compared to those in the N1 condition.

The data presented in Table 1 show considerable between-speaker variability, a pattern in line with previous studies (Savariaux et al., 1995; Ménard et al., 2004). In the F1 dimension, only two speakers significantly increased this formant in the P condition compared to the N1 condition (AL_a and LB_c). However, the percentage increase for LB_c is below the 10% threshold considered by Savariaux et al. (1995) as the acoustic criterion for successful compensation. With regard to F2, one adult (MB_a) and one child (LB_c) produced /u/ in the P condition with values significantly more than 10% greater than in the N1 condition. Thus, those three speakers could not achieve complete compensation. JC_c was the only speaker who did not significantly change the F1, F2 and F3 values, thus showing perfect compensatory abilities at the acoustic level. Note that differences that could appear considerable, like the F1 values for the subject MB_a, are not significant due to high standard deviations.

Subject	**f0**		**F1**		**F2**		**F3**	
	mean	**var. %**	**mean**	**var. %**	**mean**	**var. %**	**mean**	**var. %**
AC_a	186	+8,6*	351	-2.4	912	-10.7*	2073	+1.1
AL_a	310	+1.8*	380	+21*	938	+1.3	2743	+2.0
JR_a	297	+6.2*	399	+0.9	889	+7.5*	2749	-1.5
MB_a	220	-0.7	396	-12.3	724	+20.1*	2813	-0.4
BP_c	267	-3.6	537	+7.7	1205	-15.3	3332	+10.7*
JC_c	353	+4.3*	402	-3.2	1105	+3.7	3932	-2.8
LB_c	288	-2.3*	513	+4.9*	956	+31.0*	3443	-0.6
MC_c	297	+6.2*	399	+2.0	889	+7.5*	2749	-1.5

Table 1. Variation observed for /u/ in the perturbed condition (P) relative to the normal condition (N1) for each speaker (_a = adult, _c = children). All values are averaged over ten repetitions and are in Hertz. Significant differences are indicated by *.

At the articulatory level, between-speaker variability was observed as well, as different speakers used different strategies to compensate for the lip-tube. Changes in articulatory parameters describing tongue configurations are presented in Table 2. Only significant differences based on T-tests are indicated.

Table 2 shows that the speakers used various strategies to compensate for the lip-tube. For instance, for adult speakers, significant differences in terms of tongue position are produced by lowering (AL_a and MB_a) and fronting the tongue (MB_a). Those strategies were not successful, however, since, acoustically, both speakers produced formant values in the P condition that were significantly higher (by more than 10%) than in the N1 condition (see Table 1). For children, modifications in tongue position involved higher (BP_c, JC_c, MC_c) and more front (BP_c, MC_c) positions. Concerning tongue shape parameters, all speakers but one (AL_a) significantly modified their tongue curvature value and position. One adult (AC_a) and one child (MC_c) placed their tongue in a less bunched position in the P condition compared to the N1 condition, a pattern that is likely related to an increased constriction length. The remaining speakers adopted the reverse strategy, namely a more bunched tongue position in the P condition compared to the N1 condition. The latter tongue shape likely reduces the length of the constriction and also increases the volume of the front and/or back cavities.

Speak	Tongue position		Tongue shape	
	y max. (↓ = lower; ↑ = higher)	x of y max. (→ = more front)	Curvature (↑ = flatter; ↓ = more bunched)	Curvature position (→ = more front; ← = more back)
AC_a			↑*	←***
AL_a	↓***			
JR_a			↓***	→**
MB_a	↓**	→***	↓***	←*
BP_c	↑***	→**	↓***	→***
JC_c	↑***		↓***	→*
LB_c			↓***	→***
MC_c	↑***	→***	↑***	←**

Table 2. Significant differences observed in the perturbed condition (P) relative to the normal condition (N1) for /u/. (* = $p < .05$; ** = $p < .01$; ***=< .001)

For tongue curvature position, Table 2 shows that for two adults (AC_a and MB_a), tongue shape was modified in such a way that the mass of the tongue was more back. Note that speaker MB_a produces this pattern while fronting the highest point of the tongue, in x coordinates. These results depict a strategy whereby the highest point of the tongue is more front, relative to the palate, but relative to the whole mass of the tongue, the tongue root is advanced such that proportionally less volume is located behind this highest point. Indeed, a forward position for the highest point of the tongue (column labeled "x of max. point") may be a consequence of modifications in tongue shape. One adult (JR_a) and three children (BP_c, JC_c and LB_c) modified curvature position in the P condition, such that the mass of the tongue was more forward compared to the N1 condition. Note however that for some speakers (AC_a and BP_c), length differences in tongue contours can be observed between experimental conditions. This may have to do with an increased tension of the tongue muscles as they make an effort to compensate for the perturbation, thus affecting the propagation of the ultrasound sound waves.

4.2. Perception

The results for perception are shown in Table 3. Recall that the numbers represent the quality of the vowel, with 5 being the highest quality. The data shown in Table 3 reveal that the three adults (vowels produced by the speaker AL_a could not be part of the perception test for technical reasons) and two children (JC_c and MC_c) were able to achieve complete compensation for the vowel /u/. Indeed, for those speakers, the quality rating score in the perturbed condition was equal to or higher than the mean for the normal condition. Note that for all those subjects, vowels representing complete compensation did not occur before the fifth repetition, showing the gradual process of developing an efficient compensatory strategy.

Subject	Vowel quality		Repetition number
	Normal condition	Perturbed condition	
AC_a	3.98	4.25	9
JR_a	4.17	4.45	5
MB_a	3.25	3.29	7
BP_c	3.36	-	-
JC_c	4.16	4.20	5
LB_c	3.3	-	-
MC_c	3.46	3.29	9

Table 3. Vowel quality ratings for vowels produced in normal (N1) and perturbed (P) conditions for all speakers (1 = bad; 5 = excellent). Only the first trial representing complete compensation is shown.

5. General discussion

The results indicate that children show compensatory behavior at the acoustic and articulatory levels. Indeed, according to Table 1, in F1 and/or F2, all children produced a percentage variation in the P condition compared to the N1 condition that was lower than 10%. The same was true for three adult speakers but one (MB_a). Given the similar results obtained by Ménard et al. (2004) and Baum and Katz (1988), it seems likely that compensatory articulation abilities emerge at the same time as the child learns to produce perceptually correct phonological segments. At the articulatory level, several strategies were observed. Those strategies may be understood when considering the fact that F1 and F2 in /u/ are affiliated to Helmholtz resonators, whose resonant frequencies can be calculated as ($F = (c/2\pi k)*(A_{co}/L_{co}*V_{ca})^{1/2}$, where A_{co} = constriction area, L_{co} = constriction length, V_{ca} = cavity volume). Modifications in tongue position and tongue shape likely resulted in increased volumes in the front and/or back cavities, and thus decreased F1 and F2 values. The strategies observed (Table 2) may also be related to increased constriction length, which results in lower formant values. However, a higher tongue position in the P condition compared to the N1 condition, as observed for some speakers reduces constriction area and the affiliated formant values. Thus, both children and adults showed compensatory abilities since both groups were able to modify tongue position and/or shape in the perturbed condition, compared to the normal condition N1. However, when we consider compensation at the perceptual level, adults performed slightly better. Indeed, all three adults produced at least one perturbed vowel for which the quality was judged equivalent to or better than in the normal condition N1, whereas only two out of four children produced such vowels. The variability observed in both acoustic and articulatory results tends to support previous hypotheses on the selective nature of articulatory compensation, limited in part by the speaker's choices and by biomechanical constraints relative to each speaker. This variability, added to the fact that all of the speakers attempted to compensate for the perturbations, supports the hypothesis that the goal of the speech production process is neither purely acoustic nor purely articulatory but multimodal, as proposed by Perrier (2005).

6. Conclusion

This study evaluated the compensatory articulation capacity of children and adults using an artificial perturbation. Although articulatory simulations showed that total compensation was possible, two of the children were not able to completely compensate. However, children showed compensatory skills as young as 5 years of age. The evolution of this capacity seems to parallel speech motor control development. Other perturbations could be used to validate this hypothesis and to further investigate the selective nature of articulatory compensation.

References

Baum, S. R., and Katz, W. F. Acoustic analysis of compensatory articulation in children. *Journal of the Acoustical Society of America*, 84(5):1662-1668, 1988.

Li, M., Kambhamettu, C., and Stone, M. EdgeTrak, a program for band-edge extraction and its applications. Sixth IASTED International Conference on Computers, Graphics and Imaging, Honolulu, HI, August 13–15, 2003.

Lindblom, B., and Sundberg, J. Acoustical consequences of lip, tongue, jaw, and larynx movement. *Journal of the Acoustical Society of America*, 50(4):1166-1179, 1971.

Ménard, L., Perrier, P., and Savariaux, C. Exploring production-perception relationships during speech development: A study of compensation strategies to a lip-tube perturbation. *147th Congress of the Acoustical Society of America,* 115(5):2629, 2004.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M., Perrier, P., Vick, J., Wilhelms-Tricarico, R., and Zandipour, M. A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss. *Journal of Phonetics*, 28:233-272, 2000.

Perrier, P. Control and representations in speech production. *ZAS Papers in Linguistics*, 40:109-132, 2005.

Savariaux, C., Perrier, P., and Orliaguet, J.-P. Compensatory strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production. *Journal of the Acoustical Society of America*, 98(5):2428-2442, 1995.

Savariaux, C., Perrier, P., Orliaguet, J.-P., and Schwartz, J.-L. Compensation strategies for the perturbation of French [u] using a lip tube. II. Perceptual analysis. *Journal of the Acoustical Society of America*, 106(1):381-393, 1999.

Stone, M. 2005. A guide to analyzing tongue motion from ultrasound images. *Clinical Linguistics and Phonetics*, 19:455-502.