Individual differences in speech adaptation to an artificial palate

Shari R. Baum, and David H. McFarland

Individual differences in speech adaptation to an artificial palate

Shari R. Baum[a)]
School of Communication Sciences & Disorders, McGill University, 1266 Pine Avenue West,
Montréal, Québec H3G 1A8, Canada

David H. McFarland
École d'orthophonie et d'audiologie, Université de Montréal, CP 6128, Succursale Centre-ville,
Montréal, Québec H3C 3J7, Canada

This preliminary investigation examined the ability of individual speakers to adapt to a structural perturbation to the oral environment in the production of [s]. In particular, the experiment explored whether previous evidence of relatively quick adaptation subsequent to intensive practice would be replicated, whether vowel environment would influence the degree of adaptation, whether adaptive strategies would carry over to normal productions and/or similar sounds (i.e., cause negative aftereffects), and whether adaptive strategies developed during the practice phase could be recalled 1 h later. Results of acoustic and perceptual analyses generally revealed improvement after practice, few consistent effects of vowel context, few negative aftereffects, and an absence of quick recall of adaptive strategies. Moreover, extensive individual differences were found in both the degree of initial perturbation and the extent of adaptation. Implications of the results for issues in speech adaptation are briefly discussed. © *2000 Acoustical Society of America.* [S0001-4966(00)01406-5]

PACS numbers: 43.70.Aj [AL]

INTRODUCTION

In a recent series of investigations, we have been studying the adaptive responses of the speech production system to functional and structural modifications of the oral cavity (Baum and McFarland, 1997; Baum, McFarland, and Diab, 1996; McFarland and Baum, 1995; McFarland *et al.*, 1996). Of particular relevance, we examined whether adaptation to an artificial palate with a thicker than normal alveolar region in [s] production could occur in a relatively brief period, given intensive, focused practice (Baum and McFarland, 1997). The data revealed a gradual improvement in [s] production with the palate in place, as reflected in fricative spectra as well as quality ratings by phonetically trained listeners. Somewhat surprisingly, changes were also found between [s] productions in normal conditions pre- and post-practice, suggesting potential negative ''aftereffects'' (Anstis, 1995).

One of the hallmarks of previous investigations of adaptation to perturbations has been the apparent individual variability in compensatory abilities (see also Flege *et al.*, 1988; Savariaux *et al.*, 1995, 1999). For example, in the investigation just described, three of the seven subjects appeared to have adapted much more completely than the other speakers after the 1-h practice period (Baum and McFarland, 1997). Other authors have suggested that individual speakers may differ in their overall articulatory ''skill'' and their ability to modify articulatory dimensions in response to alterations in vocal tract configuration (Savariaux *et al.*, 1995, 1999).

The present investigation, with three main objectives, was designed to provide a preliminary within-subject analysis of the effects of intensive practice on the development of speech adaptation to the presence of an artificial palate. One goal was to examine whether the ability to adapt to a perturbation in fricative production is influenced by vowel environment. One might predict that certain vowel contexts may impede adaptive processes because of large or contrasting articulatory demands of the vowel and consonant. In contrast, if one adopts the hypothesis that, despite coarticulatory influences, invariant acoustic properties characterize phonetic segments, one might predict that vowel context should not affect adaptation because speakers are targeting the same main acoustic goal regardless of vowel environment. A second goal was to explore whether intensive practice will also affect other speech sounds (produced under normal conditions), particularly those with similar production characteristics, such as [š], due to the development of a distributed and adaptive mode of articulatory programming (Baum and McFarland, 1997). The third primary objective was to assess speakers' ability to accommodate without further practice to a previously adapted perturbation and the degree to which adaptive articulatory strategies are automatized, much like normal articulatory programs (see Hamlet and Stone, 1978; Hamlet *et al.*, 1978).

I. METHODS

A. Acoustic analyses

1. Subjects. Four adult female native speakers of (Quebec) French served as subjects, none of whom reported a history of speech, language, or hearing deficits. Three of the four (Speakers 2–4) had a history of orthodonture, with two (Speakers 3 and 4) wearing removable appliances for 1 year or less.

2. Stimuli and procedures. A specially designed artificial palate was constructed for each speaker with a 6-mm build-up of acrylic at the alveolar ridge to perturb [s] pro-

[a)]Electronic mail: c3cr@musica.mcgill.ca

duction. Details of the palate design are described elsewhere (Baum and McFarland, 1997). Speech production was sampled at three time intervals: immediately upon insertion of the palate (time 0), after 60 min of [s]-intensive practice (see below) with the palate in place (time 1), and 60 min later (time 2) after normal speech activity without the palate. At time 0, ten repetitions of the syllables [si sa su ši ša šu] were elicited prior to insertion of the artificial palate, comprising a normal baseline condition. Once the palate was inserted, an additional ten repetitions of [si sa su] were elicited for the time 0 perturbed condition. At time 1, another set of ten repetitions of [si sa su] was recorded with the palate in place to determine whether the 60 min of practice had influenced [s] production. A second unperturbed set of stimuli (both [s] and [š]) was also elicited to examine any potential aftereffects or carryover of altered articulatory patterns. Finally, at time 2, the same set of recordings was made—i.e., ten repetitions of [si sa su] with the palate in place, and ten repetitions of [si sa su ši ša šu] under normal conditions. (The normal condition was recorded first at this time interval.)

During the 60-min practice interval, subjects read [s]-laden passages aloud to facilitate adaptation to the palatal perturbation. Target stimuli were recorded on DAT tape using a Sony DTC-57ES recorder and head-mounted directional microphone (AKG-HD421U) which ensured a constant mic-to-mouth distance.

3. Analyses. All target utterances were digitized at a rate of 20 K samples/s with a 9-kHz low-pass filter and 12-bit quantization using the BLISS speech analysis system (Mertus, 1989). Centroid frequencies were computed at the midpoint of the frication noise in each stimulus as an indication of degree of compensation to the palatal perturbation (see also Baum and McFarland, 1997; McFarland and Baum, 1995; McFarland *et al.*, 1996).[1]

B. Perceptual analyses

1. Subjects. Ten native French-speaking adults with training in articulatory assessment served as raters in the perception experiment. None of the listeners had served as a speaker in the experiment.

2. Stimuli and procedures. For each individual speaker, a perceptual test was constructed consisting of all of the /sV/ tokens in both perturbed and unperturbed conditions, for a total of 180 stimuli per speaker. Ten practice training trials preceded each test: 5 ''high-quality'' exemplars and 5 ''poor'' exemplars chosen from the normal and perturbed conditions at time 0, respectively. Those stimuli with centroid frequencies furthest from normal served as examples of poor-quality productions and listeners were informed about the defined quality of the example stimuli. The perception tests were presented in four separate sessions with order of presentation counterbalanced across subjects.

Stimuli were presented to listeners over closed headphones (with a high-frequency limit of 22 kHz) with a 5-s ISI, during which listeners were required to rate the quality of the fricative on a 10-cm visual analog scale based on the examples and their articulatory assessment experience. The
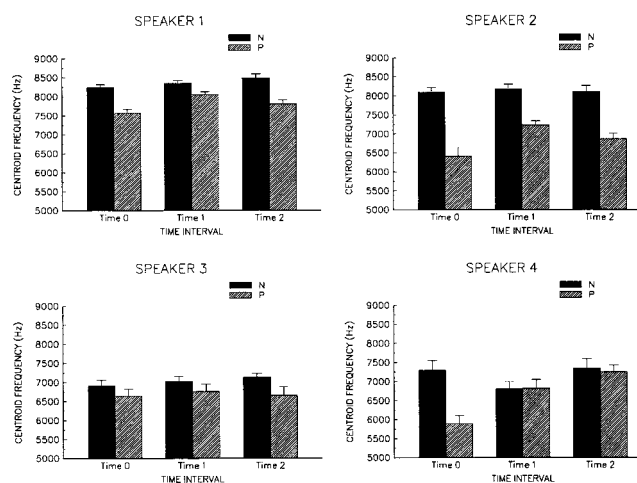

FIG. 1. Mean centroid frequencies (and standard errors) for [s] in palate and normal conditions across time intervals for each speaker.

endpoints of the scale were labeled ''unintelligible'' and ''perfect'' and listeners were asked to make a mark at an appropriate point along the scale.

II. RESULTS

Mean centroid frequencies for [s] computed at each time interval in both perturbed and unperturbed conditions are displayed in Fig. 1(a)–(d) for each speaker.[2] Separate time interval (time 0, time 1, time 2)×palatal condition (normal,perturbed)×vowel ([i a u]) analyses of variance (ANOVA) were carried out for each speaker individually. The ANOVA for Speaker 1 revealed main effects for time interval $[F(2,18)=5.268, \ p<0.05]$, palatal condition $[F(1,9)=85.166, \ p<0.001]$, and vowel $[F(2,18)=6.262, \ p<0.01]$, as well as interactions of time×condition $[F(2,18)=3.473, \ p=0.05]$ and time×vowel $[F(4,36)=2.819, p<0.05]$. *Post hoc* analyses of the time×condition interaction using the Newman–Keuls procedure ($p<0.05$) revealed significantly lower centroids in the perturbed conditions at time 0 and time 2. At time 1, the centroids in the normal and perturbed conditions did not differ significantly, suggesting some adaptation had occurred after the practice period.[3] Within the perturbed condition, only centroids at time 0 and time 1 differed significantly, confirming the compensation just noted. Because no three-way interaction emerged, the vowel-related effects were not explored further.

The ANOVA for Speaker 2 yielded significant main effects and interactions of all variables. Of particular interest was the time×condition×vowel interaction $[F(4,36)=2.639, p=0.05]$, which was subjected to *post hoc* analysis as described above. These tests revealed significant differences in centroid values computed in the normal and perturbed conditions for all vowels at all time intervals, with the exception of [su] stimuli at time 1 (for which a trend emerged). For this speaker, despite apparent improvements in adaptation over the practice interval, significant perturbation remained.

The ANOVA computed on Speaker 3's data revealed main effects of palatal condition $[F(1,9)=21.955, \ p<0.001]$ and vowel $[F(2,18)=352.382, \ p<0.001]$ and interactions of condition×vowel $[F(2,18)=24.747, p<0.001]$
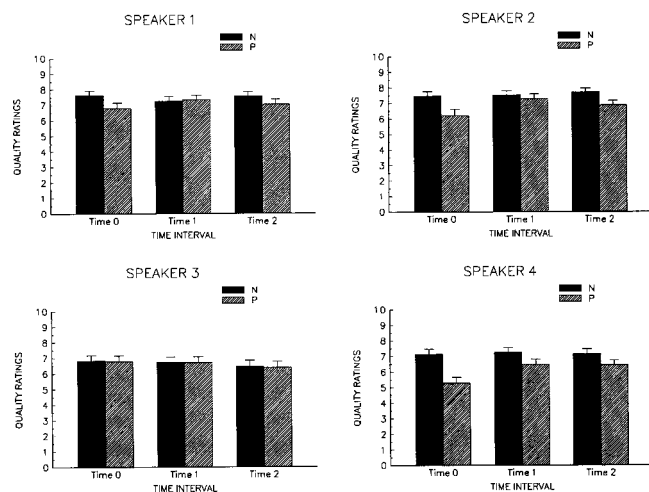
FIG. 2. Mean quality ratings (and standard errors) for [s] in palate and normal conditions across time intervals for each speaker.

and time×condition×vowel $[F(4,26)=3.23, p<0.05]$. *Post hoc* analysis of the three-way interaction yielded only a few significant comparisons: at each time interval, centroids for [su] stimuli only were significantly lower in the perturbed relative to the normal condition. Thus, even immediately after insertion of the artificial palate, this speaker's [s] articulation was not greatly affected.

Finally, like Speaker 2, the ANOVA for Speaker 4 yielded three significant main effects and numerous interactions. *Post hoc* testing of the time×condition×vowel interaction $[F(4,36)=4.540, p=0.005]$ revealed significant effects of the perturbation at time 0 for all vowels, no significant differences at time 1 (except for [si] stimuli, which displayed a difference in centroid frequencies in the unexpected direction), and differences between perturbed and normal conditions at time 2 for all but [sa] stimuli. For this speaker, within the normal condition, centroids for [si] and [sa] stimuli were lower at time 1 as compared to time 0 and time 2, suggesting an influence of the palate on normal [s] production. Within the perturbed condition, centroids were higher at time 1 relative to time 0 across vowels and remained high at time 2.

Mean centroid frequencies were also calculated at each time interval for normal (unperturbed) [š] stimuli to determine whether the altered articulatory patterns developed in compensation for the palate would affect nonalveolar sounds. Individual time interval×vowel ANOVAs conducted on [š] centroid frequencies revealed few significant differences of interest for any of the speakers, despite apparent variability in mean centroid frequencies of up to several hundred Hz across the time intervals. (Details are available upon request.)

Individual ratings along the visual analog scale on the perception tests for each speaker were computed in cm and averaged across the ten repetitions of each target stimulus. These values (displayed in Fig. 2) were submitted to separate time interval×palatal condition×vowel ANOVAs for each speaker. The ANOVA for Speaker 1 revealed no significant main effects or interactions. For Speaker 2, all main effects and interactions were significant. Of particular interest was the three-way interaction of time×condition×vowel

$[F(4,40)=9.043, p<0.001]$, which was further analyzed using the Newman–Keuls procedure. Initial pairwise comparisons focused on the quality ratings in the normal versus the palatal conditions at each time interval. At time 0, ratings were significantly lower in the palate condition relative to the normal condition for all but [sa] stimuli. At time 1, only the [si] stimuli in the palate condition were rated significantly more poorly than in the normal condition. Finally, at time 2, ratings for the stimuli produced under normal and palate conditions differed for all but the [su] stimuli. Comparisons were also made across the time intervals within conditions. In the normal condition, the ratings for [si] stimuli at time 0 were significantly lower than at times 1 and 2. No other significant differences emerged. In the palate condition, ratings for stimuli produced at time 0 were significantly lower than at both times 1 and 2 for almost all comparisons, with the exception of the [sa] stimuli at time 0 (where no effect of the palatal perturbation was seen). Surprisingly, for [sa] stimuli at time 2, quality ratings were worse than at time 0.

The ANOVA for speaker 3 revealed a main effect of time $[F(2,16)=8.324, p<0.003]$ and interactions of time×vowel $[F(4,32)=2.826, p<0.05]$ and condition×vowel $[F(2,16)=20.567, p<0.001]$. *Post hoc* analyses of the latter interaction revealed significantly lower quality ratings for [su] stimuli in the palate compared to the normal condition, collapsed over time intervals. In contrast, for [si] stimuli, ratings were unexpectedly higher in the palate condition relative to the normal condition, while for [sa] stimuli, no significant differences across condition emerged. For speaker 4, all main effects and interactions were significant with the exception of the condition×vowel interaction $(F<1)$. *Post hoc* analysis of the three-way interaction $[F(4,40)=4.262, p<0.01]$ revealed a significant effect of the palatal manipulation on quality ratings at time 0 for all vowels; at time 1, only [si] stimuli were still rated lower in the palate compared to the normal condition. At time 2, ratings for both [si] and [sa] productions were again lower in the perturbed condition. Within the normal condition, no differences emerged across time intervals; however, within the palate condition, for both [sa] and [su], ratings at time 0 were significantly lower than at times 1 and 2.

III. DISCUSSION

This investigation was designed to provide preliminary data representing detailed within-subject comparisons of the development of speech adaptation to a structural modification of the oral cavity. Emerging from these analyses were sometimes striking differences in individual compensatory responses. For example, evidence of speech adaptation for [s] production after 1 h of intensive practice was observed in the acoustic and perceptual data for Subject 4, in the acoustic data only for Subject 1, and in the perceptual data only for Subject 2 (with a trend towards adaptation observed in the acoustic results). Further, Subject 3 showed essentially no effect of the artificial palate on speech articulation at the first and subsequent measurement intervals. At present, we can only speculate as to some of the many factors that may contribute to these individual differences. We and others (Baum *et al.*, 1996; McFarland *et al.*, 1996; Savariaux *et al.*, 1995,

3574   J. Acoust. Soc. Am., Vol. 107, No. 6, June 2000

S. Baum and D. H. McFarland: Letters to the Editor   3574

1999) have suggested that individual subjects may differ in their ability to integrate altered sensory feedback in the modification of articulatory gestures appropriate to the change in oral form. An additional, and related, possibility is that individual differences in oral form or articulatory placement may have rendered the artificial palate more or less perturbing to [s] production. However, consistent with our previous findings (McFarland *et al.*, 1996), no consistent relationship was observed between measures of palatal dimensions and individual responses to articulatory perturbation.[4] This does not rule out the possibility that articulatory positioning for [s] production may have influenced the speech perturbation of the artificial palate. For example, certain subjects may normally adopt a more posterior positioning of the tongue for [s] outside of the perturbed ''zone'' (Savariaux *et al.*, 1999).[5]

Individual differences were also observed in the extent to which speech adaptation to perturbation subsequent to an hour of intensive practice influenced nonperturbed articulatory planning. In fact, changes in [s] production in the normal conditions pre- and post-practice, suggesting negative ''aftereffects,'' were observed consistently in the acoustic data of only one experimental subject. In our previous investigation we found that five of the seven subjects exhibited negative aftereffects using a criterion of a 1000-Hz decrease in normal centroid frequencies after the practice period (see Table II, Baum and McFarland, 1997). Based on the perceptual data, three of the seven subjects exhibited such aftereffects. These data suggest that there are individual differences in the distributed influence of the development of adapted articulation to global articulatory programming. Similarly, intensive practice intended to facilitate [s] production under conditions of perturbation had little effect on other speech sounds with similar production characteristics. No consistent effects of the practice interval were observed on [š] production, suggesting that the altered articulatory processes associated with speech adaptation were specific to the highly practiced sounds. Our data also suggest that speakers targeted a specific consonantal acoustic goal under conditions of perturbation regardless of vowel context, as no systematic effect of vowel environment on the adaptive response to the oral articulatory perturbation was found.

Finally, the present results suggest that adapted speech articulation resulting from intensive target-specific practice results in compensatory responses that are fragile and not easily recalled 1 h after the practice interval. This is in contrast to an earlier investigation of speech adaptation using an artificial palate very similar to that used in the current study (Hamlet *et al.*, 1978) with the exception that there was a 4-mm build-up of acrylic as contrasted to 6 mm in the present work. In that investigation, however, subjects were provided with a much longer period of adaptation during which they wore the artificial palate for 2 weeks except while sleeping. After the 2-week adaptation period, subjects were tested again and phonetic/perceptual judgments revealed more normal consonant production after only the third repetition of the test utterance, suggesting a rapid retrieval of the modified articulatory strategies. Although the period of intensive practice used in our previous and current investigation appears to accelerate the adaptation process, longer exposure to the presence of the change in oral form may be necessary to stabilize the modified articulatory programs.

[1]Although there are obviously good reasons to compute both temporal measures and additional spectral moments (e.g., skewness and kurtosis) in providing an accurate characterization of fricatives, we limited the reported measures to centroid frequencies for several reasons. First, our previous data did not reveal differences across conditions in either temporal or other spectral measures. Second, our primary goal was a comparison of perturbed and unperturbed conditions, not specifically the most comprehensive spectral characterization of fricatives.

[2]It should be noted that some of the speakers' productions yielded high centroid values, approaching the cutoff frequency of the filter. It is, therefore, possible that the spectra are not fully representative of the [s] productions. However, because all comparisons are within-subject, across perturbation conditions, this was unlikely to have affected the specific questions under investigation.

[3]One must, of course, always be cautious about interpreting a null result. However, given the significant differences which were found at other time intervals, the question of sufficient statistical power does not appear to be at issue.

[4]Interestingly, speakers' history of orthodonture also showed no consistent influence on adaptation.

[5]It is interesting to note that the acoustic and perceptual findings, although largely consistent with one another, were not exactly parallel. Such a pattern is in keeping with previous investigations (e.g., Savariaux *et al.*, 1999) and emphasizes the need for multiple, complementary measures in assessing articulatory compensation.

Anstis, S. (**1995**). ''Aftereffects from jogging,'' Exp. Brain Res. **103**, 476–478.

Baum, S., and McFarland, D. (**1997**). ''The development of speech adaptation to an artificial palate,'' J. Acoust. Soc. Am. **102**, 2353–2359.

Baum, S., McFarland, D., and Diab, M. (**1996**). ''Compensation to articulatory perturbation: Perceptual data,'' J. Acoust. Soc. Am. **99**, 3791–3794.

Flege, J., Fletcher, S., and Homiedan, A. (**1988**). ''Compensating for a bite block in /s/ and /t/ production: Palatographic, acoustic, and perceptual data,'' J. Acoust. Soc. Am. **83**, 212–228.

Hamlet, S., and Stone, M. (**1978**). ''Compensatory alveolar production induced by wearing a dental prosthesis,'' J. Phonetics **6**, 227–248.

Hamlet, S., Stone, M., and McCarty, T. (**1978**). ''Conditioning prostheses viewed from the standpoint of speech adaptation,'' J. Prosthet. Dent. **40**, 60–66.

McFarland, D., and Baum, S. (**1995**). ''Incomplete compensation to articulatory perturbation,'' J. Acoust. Soc. Am. **97**, 1865–1873.

McFarland, D., Baum, S., and Chabot, C. (**1996**). ''Speech compensation to structural modifications of the oral cavity,'' J. Acoust. Soc. Am. **100**, 1093–1104.

Mertus, J. (**1989**). BLISS User's Manual (Brown University, Providence).

Savariaux, C., Perrier, P., and Orliaguet, J. (**1995**). ''Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production,'' J. Acoust. Soc. Am. **98**, 2428–2442.

Savariaux, C., Perrier, P., Orliaguet, J-P., and Schwartz, J-L. (**1999**). ''Compensation strategies for the perturbation of French [u] using a lip tube. II. Perceptual analysis,'' J. Acoust. Soc. Am. **106**, 381–393.