10-28-2002

Learning to Produce Speech with an Altered Vocal Tract: The Role of Auditory Feedback

Jeffery A. Jones
Wilfrid Laurier University, jjones@wlu.ca

Kevin Munhall
Queen's University - Kingston, Ontario

Recommended Citation

Jones, J.A. & Munhall, K.G. (2002). Learning to produce speech with an altered vocal tract: The role of auditory feedback. The Journal of the Acoustical Society of America 113, 532. doi: 10.1121/1.1529670

Learning to produce speech with an altered vocal tract: The role of auditory feedback

Jeffery A. Jones[a)]
ATR International—Human Information Science Laboratories, Communication Dynamics Project, 2-2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-0288, Japan

K. G. Munhall
Department of Psychology, Department of Otolaryngology, Queen's University, Kingston, Ontario K7L 3N6, Canada

(Received 21 February 2002; revised 28 October 2002; accepted 28 October 2002)

Modifying the vocal tract alters a speaker's previously learned acoustic–articulatory relationship. This study investigated the contribution of auditory feedback to the process of adapting to vocal-tract modifications. Subjects said the word /tɑs/ while wearing a dental prosthesis that extended the length of their maxillary incisor teeth. The prosthesis affected /s/ productions and the subjects were asked to learn to produce ''normal'' /s/'s. They alternately received normal auditory feedback and noise that masked their natural feedback during productions. Acoustic analysis of the speakers' /s/ productions showed that the distribution of energy across the spectra moved toward that of normal, unperturbed production with increased experience with the prosthesis. However, the acoustic analysis did not show any significant differences in learning dependent on auditory feedback. By contrast, when naive listeners were asked to rate the quality of the speakers' utterances, productions made when auditory feedback was available were evaluated to be closer to the subjects' normal productions than when feedback was masked. The perceptual analysis showed that speakers were able to use auditory information to partially compensate for the vocal-tract modification. Furthermore, utterances produced during the masked conditions also improved over a session, demonstrating that the compensatory articulations were learned and available after auditory feedback was removed. [DOI: 10.1121/1.1529670]

PACS numbers: 43.70.Fq, 43.70.Aj, 43.70.Dn [AL]

I. INTRODUCTION

In order to learn to produce speech, children must learn the unique configuration of their vocal tracts. Even after speech acquisition, children's speech motor control systems must adapt to gradual changes in the shape and size of their vocal tract due to growth. By comparison, changes to the vocal tract are relatively minor after puberty (Benjamin, 1997). However, adults may still be confronted with severe vocal-tract modifications if they lose teeth, acquire dentures, or wear other types of dental appliances. These vocal-tract alterations often mean that previously learned articulations do not produce speech sounds of the same quality. To adjust to the new articulatory–acoustic relationship resulting from vocal-tract modifications, speakers must modify their previously learned articulations in order to produce perceptually adequate speech sounds.

A number of studies have demonstrated that adult speakers can compensate to some degree for structural changes to the oral cavity. Laboratory manipulations have involved subjects wearing dental prostheses. For instance, Hamlet and her colleagues conducted a series of studies in which subjects had to learn to speak while wearing an ''artificial palate'' that covered the alveolar ridge region of the mouth (Hamlet, 1973; Hamlet, Cullison, and Stone, 1979; Hamlet and Stone, 1976, 1978; Hamlet, Stone, and McCarty, 1978). Observations of sibilant articulations using techniques such as electropalatography, ultrasound, and magnetic resonance imaging have shown that accurate sibilant production relies on the exact placement of the tongue relative to the palate in order to form a medial groove (Fletcher and Newman, 1991; Narayanan, Alwan, and Haker, 1995; Stone *et al.*, 1992). The presence of the artificial palate caused the tongue to contact the alveolar ridge sooner than it would normally and to release contact later than it should, lengthening frication and altering the width of the groove of the tongue (Hamlet *et al.*, 1979). These deleterious effects tend to be greater when the thickness of the palate is increased.

Subjects eventually do improve the quality of their speech in the presence of the artificial palate (Hamlet, 1973; Hamlet and Stone, 1976, 1978; Hamlet *et al.*, 1978, 1979). Small improvements are apparent after a relatively small number of practice trials that occur within an hour-long experimental session (McFarland, Baum, and Chabot, 1996). However, several days to weeks are often needed to achieve normal sounding productions (Hamlet and Stone, 1976; Hamlet *et al.*, 1978). Once adaptation has occurred, it takes only a few minutes of practice to readapt to the artificial palate even if months have elapsed since a subject's previous exposure to the altered oral environment (Hamlet *et al.*, 1978).

The contribution of auditory feedback to learning to produce normal speech in the presence of these novel vocal-tract manipulations is not known. It is widely believed that the

a)Electronic mail: jones@atr.co.jp

availability of auditory feedback regarding speech performance is particularly important for the development of normal speech in children (Borden, 1979; Oller and Eilers, 1988; Osberger and McGarr, 1982; Smith, 1975). However, longitudinal studies of postlingually deafened individuals suggest that auditory feedback is also a factor in the long-term maintenance of accurate speech in adults. Abnormalities in the control of pitch, loudness, and the rate of speech appear quite soon after hearing is lost. Longer periods of deafness lead to increased variability in consonant and vowel production (Binnie, Daniloff, and Buckingham, 1982; Cowie and Douglas-Cowie, 1992; Lane and Webster, 1991; Waldstein, 1990).

Evidence from the experimental manipulation of the auditory feedback received by normal-hearing individuals confirms these clinical data. For example, masking the auditory feedback of hearing individuals affects aspects of speech such as pitch (e.g., Rivers and Rastatter, 1985; Ternström, Sundberg, and Collden, 1988). Modifications of the spectra of feedback often lead to immediate changes in speech that are dependent on the frequencies filtered (Garber and Moller, 1979). If subjects' feedback regarding their F_0 is artificially raised or lowered, they tend to compensate by shifting their vocal pitch in the opposite direction of the perturbation (Burnett *et al.*, 1998; Donath, Natke, and Kalveram, 2002; Elman, 1981; Jones and Munhall, 2000, 2002; Kawahara, 1995a, 1995b).

Longer-term effects have also been induced. For example, Houde and Jordan (1998) asked subjects to whisper one-syllable words while they received altered auditory feedback regarding their vowel productions. Subjects heard feedback in which the formants of the vowels they were producing were gradually shifted enough over time to change the vowels' phonetic identity. Subjects compensated for the formant transformations. These compensations persisted even during trials in which feedback was masked by noise. Subjects either modified their existing mapping between their vocal-tract productions and their acoustic feedback or developed a new mapping. Analogous results were obtained when Jones and Munhall (2000, 2002) gave subjects altered auditory feedback regarding their fundamental frequency productions.

These data suggest that auditory feedback is used both to make online corrections and for the longer-term calibration of the mapping between speech gestures and the resulting acoustic feedback. Feedback may become even more crucial under circumstances where the characteristics of the vocal tract or motor system are altered.

There is little previous work on the specific importance of auditory feedback in adapting to the novel acoustic-motor mapping brought about by altering the vocal tract. However, a number of clinical observations indicate that recovery without auditory feedback is very difficult. For example, Perkell *et al.* (1995, 2000a) described a subject who became deaf as a result of surgery to remove bilateral acoustic neuromas. During the surgery, the subject received an auditory brainstem implant that provided her with information regarding the auditory envelope but did not provide information regarding spectral cues. Despite her situation, the subject still maintained a good /s/ versus /ʃ/ contrast. However, when a subsequent operation caused her to suffer a slight weakness on the left side of her tongue due to denervation of the tongue muscles, she lost the /s/ and /ʃ/ contrast and could not regain accuracy. Perkell and his colleagues concluded that the loss of the important auditory information did not allow her to correct for the altered acoustic-motor relationship.

Experimental data have so far not supported these clinical findings. For example, Garber *et al.* (1980b) conducted one of the few investigations on the effect that noise has on adapting to an artificial palate. They found that masking noise did not differentially affect productions with the appliance in the mouth compared to productions made without it. More recently, Honda and Kaburagi (2000) examined the effect masking noise had on compensations to rapid changes in palatal thickness. Immediate but incomplete compensations of tongue position were found when the thickness of an artificial palate was suddenly changed during production of fricatives. Although only a small number of subjects participated in a perceptual experiment assessing the quality of the speakers' productions, the authors concluded that auditory information did not play a significant role in compensations and that tactile information regarding tongue–palate contact or intraoral pressure is likely essential for the process.

Indeed, the importance of tactile information for compensations to vocal-tract perturbations has been shown for other types of manipulations. For example, a number of influential studies have shown that when subjects are asked to produce vowels with a ''bite block'' inserted between their teeth, they compensate for the bite block's presence with very little or no practice (Fowler and Turvey, 1980; Gay, Lindblom, and Lubker, 1981; Kelso and Tuller, 1983; Lindblom and Sundberg, 1971), even from the first glottal pulse (Lindblom, Lubker, and Gay, 1979; cf. Flege *et al.*, 1988; McFarland and Baum, 1995). In order for the perceptual identity of a phoneme to be maintained with a bite block in a speaker's mouth, an unnatural articulator configuration must be used. Somatosensory and proprioceptive information is available regarding the position of the articulators and the nature of the bite block restricting movement before speakers speak. This information helps the speech motor control system reorganize speech even prior to movement initiation.

The results from bite block studies highlight the importance of somatosensory information in adjusting to novel speech conditions. They also illustrate a potential confound that exists in many of the studies that have experimentally altered vocal tracts in ways that reduce or affect tactile feedback. In particular, studies that have involved artificial palates as a tool to explore adaptation have all been confounded by a reduction of tactile information. Covering the palate with an acrylic shield results in a loss of sensory information and may affect the strategies that subjects use during adaptation. Therefore, discerning the precise contributions of auditory and tactile feedback to the adaptation process is very difficult using these kinds of manipulations.

The goal of the present investigation was to examine the contribution of auditory feedback to learning a novel acoustic-motor relationship by modifying the vocal tract in a way that did not hinder movement or reduce somatosensory

information. To that end, speakers wore a dental prosthesis that extended the length of their teeth by a few millimeters. The prosthesis did not affect the speakers' bite. In addition, it was only in contact with the teeth and did not cover any oral tissues so that tactile information normally available was still present with the prosthesis inserted.

Because the prosthesis extended the teeth, production of sibilants was altered. To produce an /s/, speakers must position their tongue against the dento-alveolar ridge and force air through a short midsagittal groove along the anterior of their tongue blade. Sound is generated when the airstream hits the lower or upper incisors (Shadle, 1985). The presence of the teeth causes an increase in the amplitude of the noise and generates an antiresonance in lower frequencies (Shadle, 1991). The lengthened teeth provided an abnormal obstruction to the airflow normally required for sibilant production and modified the turbulence. The small cavity in front of the constriction would therefore be increased, causing the resonance frequencies to be lower. Speakers would have to find a way to increase the power of higher frequencies and would likely do this by moving their tongue blade to a more anterior position.

In the first experiment, subjects were asked to produce normal sounding /s/'s while wearing the dental prosthesis. This task required subjects to modify their normal /s/ tongue position in order to produce a good-sounding sibilant. The quality of the /s/'s was measured by having subjects say the monosyllable /tɑs/. Incorporating the /s/ into the word /tɑs/ prevented subjects from simply maintaining a static tongue position for the entire experiment; in order to say the word /tɑs/, the tongue must move from the position necessary to produce the open vowel /ɑ/, up to the dento-alveolar ridge to produce the /s/. Thus, the /s/ production had to be coordinated for each trial.

During the experiment, subjects were allowed to practice with the dental appliance while hearing their speech and then were tested in the presence of masking noise in order to track their adaptation to the device. Acoustic analyses were used to parametrize the changes in the power spectrum of the /s/ over time (see Stoica and Moses, 1997, for discussion of the computation of power spectral density). In a second experiment, the perceptual judgments of naive listeners were used to evaluate the quality of the /s/'s speakers produced over the course of the experimental session.

Our design allowed us to tease apart the contribution of auditory feedback from that of other sources of feedback. If auditory feedback is the primary vehicle for learning, then we should observe that greater improvement occurs during blocks when utterances are produced with feedback available in comparison to blocks in which utterances are produced with feedback unavailable. Thus, the learning we observe over a session should occur in a stepwise fashion, with incremental improvements only occurring during blocks when feedback is available. On the other hand, if auditory feedback is not crucial for learning the compensations necessary in the presence of the prosthesis, then any improvements observed should be equivalent for the feedback and masked conditions.

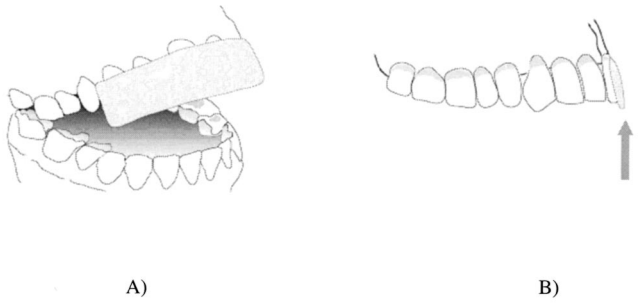

FIG. 1. (A) Depiction of the dental prosthesis in the subject's mouth. (B) A sagittal view of the maxillary teeth with the prosthesis in position (prosthesis indicated by the gray arrow). Note that the prosthesis did not affect the subject's bite when the mouth was closed.

II. EXPERIMENT 1

In the first experiment, speakers were asked to learn to produce adequate /s/ sounds in the context of the word /tɑs/ over the course of training trials.

A. Method

1. Vocal tract modification

a. Subjects. Six females between 22 and 36 years of age (mean age=27 years) participated. All subjects were graduate students at Queen's University in Canada and were native speakers of Canadian English. They reported having no history of hearing, speech, or language difficulties or disorders. Five of the subjects had received orthodontic treatment for an average of 2.4 years during their teenage years. All the subjects had a Class I occlusion (''normal bite'') and thus their maxillary incisors were situated anterior to the mandibular incisors when their mouth was closed.

b. Dental prosthesis. Dental impressions were made of each subject's maxillary and mandibular teeth. Using the impressions, an acrylic prosthesis was constructed. The prosthesis lengthened the teeth between 5 and 6 millimeters but did not affect the subjects' bite. The prosthesis fit onto the buccal and occlusal surfaces of the maxillary incisor teeth and did not require an adhesive or wire clasps to remain fixed in place. Figure 1 is a depiction of the prosthesis on a subject's teeth.

c. Recording equipment. Sessions took place in a double-walled soundproof booth (Industrial Acoustics Corporation, model 1204). The sessions were recorded on digital audiotape so that analysis of the signals could later be carried out using algorithms incorporated into the PRAAT software program (Boersma, 1993). Subjects' speech sounds were transduced with a headset microphone (Shure WH20) positioned a fixed distance from their mouth (approximately 5 cm). The speech signals were amplified (Tucker-Davis MA2 microphone amplifier) and filtered (Tucker-Davis FT6-2) with a 9 kHz cutoff. The signals were then routed to a mixer (Rolls RA62) where they were mixed with white noise (Grason-Stadler 901B). The combined noise and signal were together sent to a Yorkville reference amplifier (model SR 300) that transmitted the sound through Etymotic (ER-2) earphones foam inserts placed in the subjects' ear canals. The masking noise was approximately 75 dB SPL. Our pilot work showed that this level effectively masked voiceless

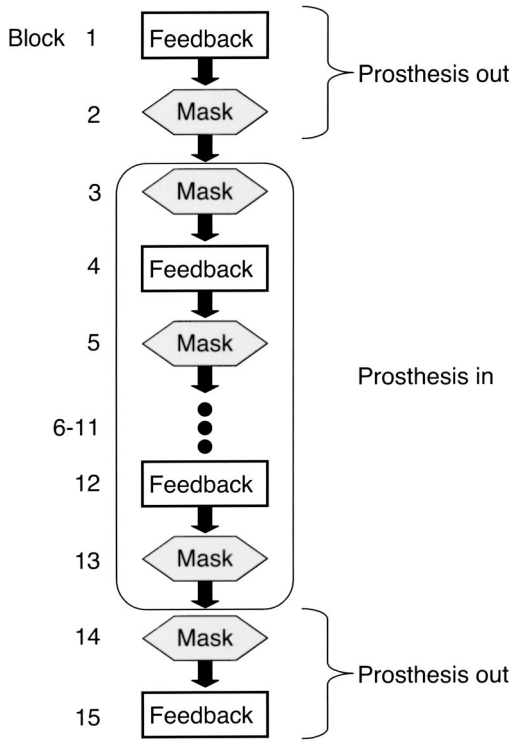

FIG. 2. Flow diagram depicting the order of conditions subjects encountered in each of the two sessions in the experiment.

sounds. The white noise was absent during trials in which the subjects were to receive auditory feedback regarding their utterances. Subjects monitored a vertical array of light-emitting diodes located in front of them. The array indicated the sound level of their productions and was used to keep their speech at similar levels across the different auditory conditions.

d. Procedure. The design of the experiment is schematically depicted in Fig. 2. Each experimental session consisted of two sessions. After the first session, subjects were given the opportunity to rest and drink water. However, both sessions occurred in a single stint that lasted less than an hour. Within each block, subjects made 10 productions of /tɑs/ in each of the following 15 blocks.

(1) The first block was a baseline condition in which subjects were recorded producing /tɑs/ without the prosthesis inserted into their mouth, and without any masking noise present. These initial utterances represented the normal /s/ productions for each subject and were later compared to other blocks to evaluate the progress of learning and the effects noise and the prosthesis had on production.

(2) The second block involved subjects producing utterances without the prosthesis in their mouth but in the presence of the white noise. This block controlled for the influence of masking noise on subjects' utterances in the absence of the vocal tract perturbation. It was used to establish the subjects' baseline productions in the absence of auditory feedback.

(3) In the next block, subjects produced utterances while wearing the prosthesis. Their auditory feedback was masked by noise and these utterances demonstrated the subjects' ability to compensate for the modification of the oral environment without the aid of auditory feedback.

(4) During the fourth block, subjects were given their first opportunity to practice saying /tɑs/ while wearing the prosthesis and receiving auditory feedback regarding the accuracy of their productions. Any differences observed between this block and the one previous can be attributed to the availability of the acoustic feedback (interacting with potential practice effects).

(5) Subjects were again asked to produce utterances in the presence of masking noise while wearing the prosthesis. These utterances were later compared to those made in the block previous to this one in order to test learning that may have occurred while receiving auditory feedback.

(6)–(13) These blocks were merely alternations of the availability of auditory feedback (block 4) and speaking in the presence of masking noise (block 5) to give speakers practice over a number of trials.

(14) During the second to last block, the subjects removed the prosthesis from their mouth and produced ten utterances in the presence of masking noise. These utterances were compared to those they made before the appliance was first placed in their mouth to determine if there was any evidence of carry-over effects that resulted from learning the new articulatory behavior.

(15) Subjects produced /tɑs/ in the absence of noise and without the prosthesis in their mouth.

2. Acoustic analyses

Although a number of techniques for parametrizing fricatives have been proposed, finding a good numerical method for characterization and classification of fricatives is still a problem. We determined the centroid (first moment) of the long-term average spectrum of each /s/ production using functions implemented in PRAAT (Boersma, 1993). The first moment or ''centroid'' is an index of the ''center of gravity'' of the spectrum for each fricative. Centroids have been found to correlate with the perceptual categorization of some fricatives and may therefore represent a perceptually salient features that speakers modify to alter the quality of their fricative productions (Forrest *et al.*, 1988). For example, /s/ sounds typically have higher centroid frequencies than /ʃ/ sounds produced by the same speaker (Nittrouer, Studdert-Kennedy, and McGowan, 1989).

During the production of an /s/, speakers modify their air pressure as well as their tongue blade and tip position in order to direct a jet of air at the teeth. This jet of air is directed to the surface of the teeth. The presence of unexpectedly long teeth would cause the normally small cavity in front of the constriction to be larger and make the resonance frequencies lower. This change would make speakers' initial productions more /ʃ/-like. Centroid values were therefore expected to be lower than normal until speakers learned to compensate.

In addition to the computation of the central moment, we also applied a technique first implemented by Evers, Reetz, and Lahiri (1998) to distinguish the acoustics of an /s/

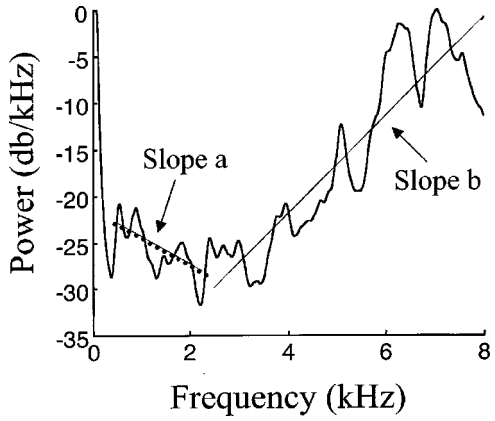

FIG. 3. The power spectrum of an unperturbed /s/ produced by a speaker in experiment 1. The linear regression lines between 0.5 and 2.5 kHz and 2.5 and 8 kHz are present for slope a and b, respectively.

from /ʃ/. Evers *et al.* (1998) compared the distribution of intensity over frequency of the spectra for /s/ and /ʃ/ in different languages. They noted that based on the slope of the spectral envelope below 2.5 kHz, and the slope between 2.5 and 8 kHz, one could visually distinguish between the two fricatives. The authors developed a reliable metric they called the ''steepness difference'' in which intensity values were regressed onto corresponding frequency values in these two frequency regions and subtracted. That is, the difference between the slope of a linear regression line (slope a) between 0 and 2.5 kHz and the slope of a linear regression line (slope b) between 2.5 and 8 kHz was used to reliably separate /s/ and /ʃ/. Figure 3 shows the power spectrum and the two regression lines for an /s/ produced by one of the speakers in this study. Evers *et al.* (1998) found that the spectral slope for frequencies up to 2.5 kHz quickly rises then above this frequency, abruptly levels off or declines for /ʃ/'s. For /s/'s, the slope is initially near zero or negative and then there is a slow rise through to the 8-kHz range. This distinction is reflected in a smaller difference in the rate of increase between the lower and higher frequencies for /s/ as compared to an /ʃ/.

Since the prosthesis speakers encountered created a slightly larger cavity within which the fricative noise would resonate, initial productions were expected to have more power in lower frequencies than productions without the prosthesis. Thus, the steepness difference may reflect changes in production as well or better than the centroid of the frequency distribution. We calculated the index in the same way as Evers *et al.* (1998) with two exceptions. First, Evers and his colleagues computed their slopes based on a 40-ms window placed in the middle of the fricative. We chose to calculate the power spectrum over the entire /s/ produced using Welch's method (Welch, 1967). In addition, our linear regression lines were calculated over the 0.5 to 2.5-kHz (slope a) and 2.5 to 8-kHz (slope b) frequency ranges. We expected the steepness difference to be initially larger for productions made with the prosthesis in the mouth. After a period of learning, this value was expected to decrease towards previously observed unperturbed values.

Apart from our spectral parametrizations, we were also interested in the relative intensity of the utterances. The presence of noise in a speaker's environment often causes them to produce utterances with higher amplitudes than environments without noise (Lane and Tranel, 1971). Amplitude does not affect the classification of sibilants (Behrens and Blumstein, 1988). However, to avoid any complication, we provided our speakers with a visual aid to help them make each production with the same amplitude. Nevertheless, the noise that was intended to mask the fricative sounds did not entirely mask the voiced portions of their utterances. Therefore, it is possible that speakers may have used the vowel portions to maintain their speaking level while fricative sounds remained affected by the masking noise. In order to test this notion we calculated the root mean squared (rms) for each /s/ production and evaluated the relative sound levels.

B. Results and discussion

The analysis of the relative intensity of the utterances showed that speakers' productions had a higher amplitude when the masking noise was present (61 dB) in comparison to when it was not (55.6 dB) [$F(1,5)=124.4$, $p<0.05$]. However, this difference was stable across the blocks [$F(1,5)=1.16$, $p>0.05$] and across the two sessions [$F(1,5)=2.35$, $p>0.05$], so any patterns observed across the sessions can be attributed to increased experience with the prosthesis and not the presence of the masking noise.

The centroid analysis showed that the presence of the tooth prosthesis affected the center of gravity of the distribution of energy over the spectrum of each subject's initial /s/ productions. The average centroid frequency values changed markedly after the prosthesis was inserted [$F(1,5)=56.71$, $p<0.01$]. The mean centroid frequency before the prosthesis was inserted into the subjects' mouths was 6171.8 Hz. After the prosthesis was inserted the mean centroid frequency dropped to 4482.0 Hz.

Apart from this initial difference, no other significant difference was observed in the centroid values between the auditory and masked conditions or across the sessions. There are at least two possible reasons for the null effects in the acoustic analyses. The most obvious explanation is that speakers were unable to learn to compensate for the dental prosthesis. Perhaps if speakers were given more extensive training, improvements in their productions might have been detectable with these statistical analyses. Notwithstanding the null finding in the acoustic analyses, the experimenters' subjective experience while listening to each subject was that the speakers' productions changed, if not improved, over the two experimental sessions. Thus, a second, alternative explanation for the null results is that the centroid was not an adequate measure for the evaluation of changes in the quality of the fricative in this particular context. It is clear from the literature that finding robust summary statistics that adequately characterize and distinguish between fricatives has been a difficult endeavor. Indeed, the reliability of such statistical measures seems to be dependent on the corpus used in a study (Evers *et al.*, 1998; Jesus and Shadle, 2002). Small, nonsignificant changes in the spectral distribution may be associated with significant changes in the perception of the /s/.

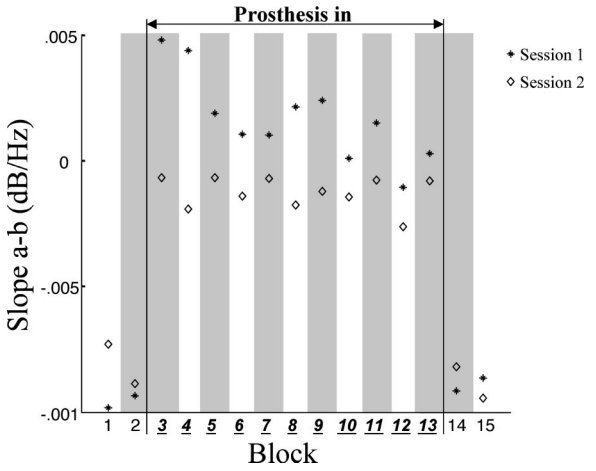

FIG. 4. The mean steepness difference (slope a−slope b) for the 15 blocks in sessions 1 and 2. Block numbers that are underlined indicate that the prosthesis was in the speakers' mouth during these blocks. The gray shading indicates that auditory feedback was masked during these blocks. The underlining of the block numbers between the two black vertical lines indicates that speakers had the prosthesis in their mouth for these productions.

Figure 4 shows the values for the steepness difference, the difference between slope a and b, over the course of the 15 blocks in each of the two sessions. A clear pattern is observable in that the values for the second session are closer to values measured before exposure to the prosthesis. The steepness difference values not only reflect a clear improvement between the two sessions, but also a linear trend toward normal values over the course of training in session 1. Both the difference between sessions 1 and 2 [$F(1,5)=11.25$, $p<0.05$] and the interaction between session and block [$F(1,5)=3.5$, $p<0.05$] are statistically significant. However, there was no statistically verifiable difference between the auditory and masked conditions.

In addition to our interest in the learning across blocks and sessions, we were also interested in differences in the learning within a block. Figure 5 shows the steepness difference for the first and last production within each of the 15 blocks across session 1 [see Fig. 5(a)] and session 2 [see Fig. 5(b)]. Even within these selected trials a significant improvement across the two sessions [$F(1,5)=8.23$, $p<0.05$] and across the blocks [$F(1,5)=2.88$, $p<0.05$] was observed. Again, no significant difference existed between the auditory and masked feedback conditions. Moreover, despite the pattern visible in the data from session 1 [Fig. 5(a)] there were no significant differences observed between the first and last trials within a block.

To summarize, the presence of the prosthesis caused centroid values to drop significantly. However, evidence that speakers were improving their productions over the experimental sessions was only observed in the spectral slope measure. Although normal production was never completely restored, the steepness difference values approached normal values gradually over blocks in session 1. This learning appears to have leveled off so that the improvement observed in session 1 is maintained during session 2.

III. EXPERIMENT 2

The acoustic analyses of speaker productions indicated that speakers were altering the acoustics of their productions

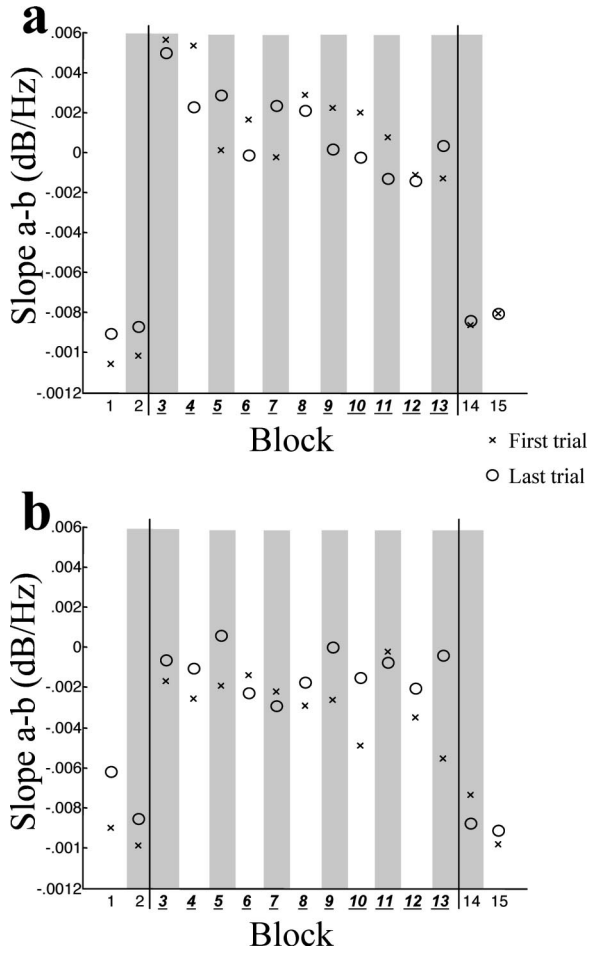

FIG. 5. The mean steepness difference (slope a−slope b) for the first and last utterances produced in the 15 blocks in sessions 1 (panel a) and 2 (panel b). Block numbers that are underlined indicate that the prosthesis was in the speakers' mouth during these blocks. The gray shading indicates that auditory feedback was masked during these blocks. The underlining of the block numbers between the two black vertical lines indicates that speakers had the prosthesis in their mouth for these productions.

as a function of experience wearing the prosthesis. However, there were no statistically verifiable differences between productions produced at the beginning as opposed to the end of a particular block. Neither was a difference between the two auditory conditions observed. As previously mentioned, steepness difference has been shown valuable for separating /ʃ/ from /s/ sounds (Evers *et al.*, 1998). However, the index is a simple and relatively crude representation of the power spectrum of a fricative. Listeners, and therefore speakers, are likely sensitive to smaller changes in the shape of the power spectrum. We therefore obtained listener judgments to see if the results we found in the acoustic analysis were comparable to the perceived quality of the /s/ productions.

As a result of the experimental design, a large quantity of data was collected. For the perceptual experiment, we therefore focused on a subset of these data. Only productions from the first and last trials of key blocks were presented to listeners. Specifically, listeners heard the first and last trials of blocks 3 and 4 which were the first masked and auditory feedback blocks after the prosthesis was inserted into the subjects' mouth (see Fig. 2). We presented the first and last trials of blocks 12 and 13, the last masked and auditory feed-

back blocks prior to the removal of the prosthesis. Finally, the subjects rated the first and last trials of blocks 14 and 15, which were the masked and auditory feedback blocks immediately after the prosthesis was removed from the subjects' mouth. The /s/ productions from both sessions for the blocks and trials above were rated by listeners.

A. Method

1. Subjects

Sixteen listeners (13 women and 3 men) between 20 and 25 years of age (mean age 21.4 years) made judgments regarding a subset of the /s/ productions made by the speakers. The listeners were native speakers of Canadian English and reported having no history of hearing, speech, or language difficulties or disorders.

2. Stimuli

The stimuli consisted of the subset of the segmented /s/ productions analyzed using the acoustical analyses described in experiment 1. The fifth utterance produced during the baseline blocks 1 and 2 (without the prosthesis present, with or without noise) of the first session were selected as comparison exemplars of each speaker's normal /s/ production. The fifth or middle production of these blocks was chosen because subjects were most likely to be acclimated to the speaking condition by this trial. Productions made with and without the presence of the masking noise were chosen to control for any differences that may have been solely caused by the presence of the masker.

The test utterances were the first and last utterances produced during blocks 3, 4, 12, and 13. Blocks 3 and 4 were the first masked and auditory conditions during which speakers wore the prosthesis; blocks 12 and 13 were the last auditory and masked conditions during which speakers wore the prosthesis. Each comparison exemplar was paired with all the test conditions. This design meant that exemplars that were produced during the masking condition and exemplars produced during the feedback condition were both paired with test stimuli that were produced with and without feedback. This procedure allowed us to test for any differences that could be attributed to the presence of noise, and not merely the result of the presence or absence of feedback. Only the first and last productions from these blocks were presented in order to reduce the number of trials listeners had to judge. Testing these trials also allowed us to track the effects of learning within the blocks.

The exemplars were also paired with the first and last productions of blocks 14 and 15 in each session. These blocks occurred after the removal of the prosthesis and, respectively, with and without the presence of the masking noise. Asking listeners to judge utterances from these blocks allowed us to evaluate the effects noise had on normal production. In all, each listener made 192 judgments.

3. Procedure

Perceptual judgment sessions took place in the soundproof booth previously used to record the speakers. The digitized auditory stimuli were equally amplified relative to the original, recorded level (NAD Electronics, model 3020I) and presented over headphones (Sennheiser, HD 265 Linear). Each speaker's tokens were presented within a single block with the order of the six different speakers randomized across listeners. The presentation of the tokens within each of the six-speaker blocks was also randomized. On each trial, subjects first heard an exemplar (a baseline /s/ from blocks 1 and 2 that was produced with or without the presence of noise) and then an /s/ production that had been produced in the presence or absence of masking noise while the speaker wore the dental prosthesis.[1] Subjects were asked to consider the first stimulus to be a normal /s/ production for that speaker. They then rated the quality of the second /s/ production on a scale of 1 through 5, with 5 representing a perfect /s/ production and 1 representing a very poor quality production. Subjects made responses by pressing appropriately labeled keys on a keyboard.

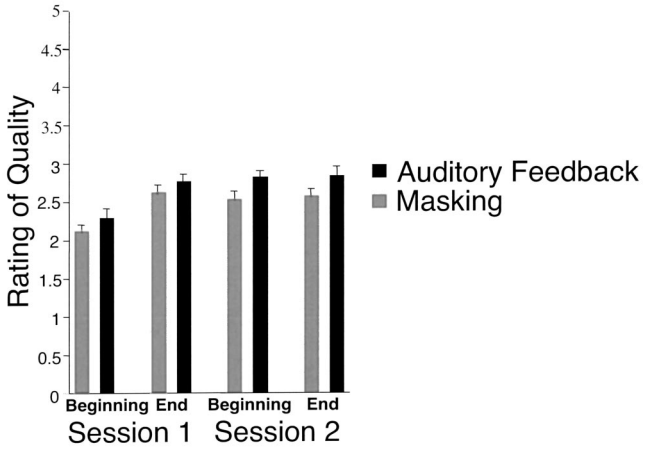

FIG. 6. The mean and standard errors of listeners' ratings of the quality of /s/ productions during the auditory and masked feedback conditions at the beginning and end of session 1 and 2.

B. Results

The results of our perceptual study showed that utterances produced while speakers could hear their own feedback were rated by listeners to be higher quality /s/'s than the productions that occurred while speakers' feedback was masked by noise. The study also showed that speakers' utterances improved with increased practice. In addition to evaluating the perceptions of listeners, we also wished to acoustically quantify the differences between the utterances produced during the two auditory conditions across the two sessions.

Figure 6 shows the mean and standard errors of listeners' perceptual ratings of the auditory stimuli speakers produced during the four blocks from the first and second sessions. A 5-way ANOVA [session×position in session (beginning versus end of each of the two sessions)×auditory feedback (feedback versus masked)×exemplar (tokens produced in the presences of feedback versus those in masking noise)×trial (first versus last trial within a block)] was used to analyze the subjects' responses. The ANOVA revealed an overall main effect for session [$F(1,15)=24.98$, $p<0.01$]. As can be seen in Fig. 6, productions made during the second

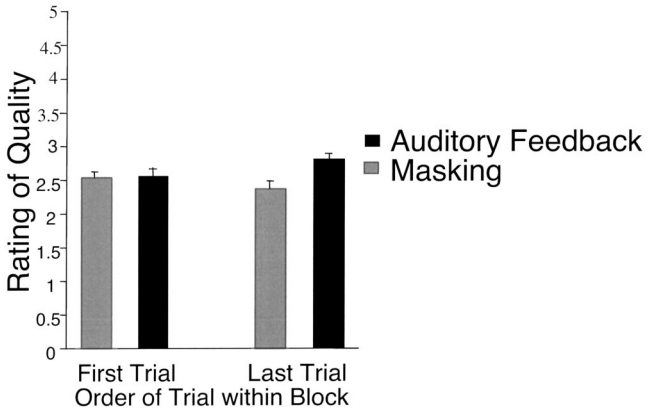

FIG. 7. The mean and standard error of listener ratings for the first and last utterances produced in the auditory feedback and masked conditions averaged across the first and second session.

session were judged to be better /s/'s than those produced during the first session. In addition, there was no interaction between session and auditory condition. Thus, speakers learned to produce better /s/'s in both the auditory feedback and masked conditions.

A similar main effect of practice was observed within each session. Utterances produced at the beginning of each session were rated poorer in quality than those produced at the end of each session across both the auditory feedback and masked conditions [$F(1,15)=49.5$, $p<0.01$]. As can be seen in Fig. 6, the amount of improvement observed in the ratings from the beginning of a session to the end was greatest in the first session for both the auditory feedback and masked conditions. This difference in improvement generated a significant interaction between session and the order of conditions [$F(1,15)=39.72$, $p<0.01$].

The availability of auditory feedback during production had a significant effect on the listener's ratings of the speakers' utterances. When auditory feedback was available during productions, utterances were judged to be of significantly higher quality than those produced in the presence of the masking noise [$F(1,15)=13.93$, $p<0.01$]. The difference in ratings between the auditory conditions was greatest during the second session, but the interaction effect did not quite reach statistical significance [$F(1,15)=2.39$, $p>0.05$] (see Fig. 6).

In addition to the main effect for auditory condition, there was also a significant interaction between auditory condition and trial [$F(1,15)=25.4$, $p<0.01$]. Figure 7 shows the mean ratings for the first and last of the ten utterances produced in the auditory feedback and masked conditions averaged across the first and second session. In only this comparison were the results complicated by the use of exemplars from different auditory feedback conditions. When test stimuli produced in the presence of auditory feedback were paired with exemplars produced in either feedback condition, last utterances were rated more ''/s/-like'' than first utterances [exemplar produced with feedback; $F(1,15)=8.92$, $p<0.01$; exemplar produced in noise, $F(1,15)=11.31$, $p<0.01$]. However, the opposite pattern of results was observed when test stimuli produced in noise were paired with exemplars that were also produced in noise; the first utterances were rated more /s/-like than the last utterances. On the other hand, the first and final utterances of test stimuli produced in noise received equivalent ratings when paired with exemplars that speakers produced in the presence of auditory feedback [$F(1,15)=0.13$, $p>0.05$]. In either case, however, there was no improvement in /s/ quality over each block of ten trials when there was no auditory feedback.

Finally, we examined the ratings assigned to blocks 14 and 15 of each session. These blocks occurred immediately after the dental prosthesis had been removed from the subjects' mouths. A separate ANOVA did not reveal a significant difference between utterances produced in the presence or absence of noise during these blocks [$F(1,15)=0.84$, $p>0.05$]. However, there was a significant difference between these trials and the training trials at the end of the two sessions [Tukey honest significance test, $p<0.01$; mean rating for the final two blocks of unperturbed trials was 3.6; mean rating for the final two blocks of training trials was 2.7].

In summary, the results of the perceptual study confirm and extend the acoustic analyses of the speakers' /s/ productions. Listeners judged productions made at the end of the first session to be better than those produced at the beginning. This improvement was maintained throughout the second session. Listeners also rated productions made while auditory feedback was available to be of higher quality than those produced while the feedback was masked with noise. In addition to the learning that occurred with increased experience across the blocks of session 1, utterances produced at the end of individual blocks were more highly rated than those produced first in a block. However, this pattern was only observed for utterances produced in the auditory feedback condition.

IV. GENERAL DISCUSSION

Subjects were asked to learn to say the word /tAs/ wearing a prosthesis that elongated their maxillary incisor teeth while receiving intermittent auditory feedback. When speakers produce the word in the auditory feedback condition, a naive group of listeners judged their final utterance to be of higher quality than their initial utterances in the condition. The opposite trend was observed when speakers' feedback was masked; their initial utterances were judged to be higher in quality than their final utterances. This difference suggests that speakers were able to use auditory information to adjust their articulations and compensate online and the lack of auditory feedback led to degraded performance over trials.

In addition to the immediate effects caused by the presence of auditory feedback, speakers also gradually improved their productions with increased exposure to the novel vocal-tract configuration. The learning curve resembled patterns from other skill acquisitions studies (e.g., Rosenbaum, Carlson, and Gilmore, 2001). Large gains are made initially and performance slowly asymptotes, producing an exponential learning curve (Heathcote, Brown, and Mewhort, 2000). Within and across each session, productions were judged to be higher in quality with increasing amounts of practice. This effect was also observed for tokens produced in the presence of the masking noise. These observations suggest that the

learning that occurred while feedback was available to the speakers transferred to utterances produced in the absence of feedback.

The multidimensional nature of speech perception means that identifying acoustic correlates is often difficult. The learning effects we observed were partially supported by acoustical analysis of the data. The size of the steepness difference moved toward unperturbed values as speakers gained more experience during the first session. This improvement plateaued but was maintained through session 2. However, in contrast to listener judgments, no statistically significant difference was found between the auditory and masked conditions. Additionally, no differences were observed between the first and the last trials within a block. We believe that the null effects in the acoustic analysis reflect a lack of sensitivity rather than the absence of effects. The steepness difference reflected the larger differences that occurred over the course of the experiment, but we must rely on listener perceptions for evaluation of smaller changes in the spectra of the speakers' productions.

Speakers were asked to make productions of similar amplitude. Nevertheless, an analysis of the intensity of the utterances showed that productions made in the masked feedback condition were higher in amplitude than those made when feedback was available. Although undesirable, the difference in intensity between the feedback conditions does not complicate interpretation of the learning effects we observed because this difference was constant within and across the sessions. The effect of the masking noise was equivalent across the entire experiment and not confounded with learning. In any case, the amplitude of fricative production does not affect classification (Behrens and Blumstein, 1988).

Our finding that the overall quality of the speech sound improved with increased practice with our novel vocal-tract arrangement is not surprising and replicates the observations of a number of other researchers. For example, Baum and McFarland (1997) found comparable results when they asked subjects to speak with an artificial palate in their mouth. Subjects read /s/-laden passages in order to promote adaptation. Every 15 min over the course of an hour, subjects produced the consonant–vowel /sa/ a number of times. The results showed that subjects gradually improved their /s/ productions with increased exposure to the altered vocal tract. Thus, even short periods of exposure can lead to significant improvements in speech.

However, others have found that it can take speakers from several hours to weeks of practice with an artificial palate before a speaker regains the high quality of their original speech categories (Hamlet and Stone, 1976; Hamlet *et al.,* 1978). Although our speakers improved over the course of the experimental session, they did not fully compensate for their artificially elongated teeth. Listeners on average judged the speakers' productions at the end of the session to be much lower in quality than utterances produced after the prosthesis was removed.

Even within this short experiment there is evidence that longer-term learning took place. Subjects' performance in the second session of training showed benefits from experience in the previous session. Similar benefits have been reported in other speech and motor learning studies. In one study, subjects quickly readapted to the presence of an artificial palate even though the original training with the palate had occurred months earlier (Hamlet *et al.,* 1978). This facility for rapid adaptation during subsequent exposure to an artificial palate suggests that new speech motor programs can be learned and then later recalled for the appropriate context (McFarland *et al.,* 1996). These findings are strikingly similar to observations that subjects reaching for visual targets adapt to visual perturbations (e.g., McGonigle and Flook, 1978) and dynamic perturbations (Brashers-Krug, Shadmehr, and Bizzi, 1996) faster if they have previously experienced the unusual sensorimotor conditions.

The improvements that we observed resulted from the opportunity for speakers to practice in the presence of auditory feedback. This finding that auditory feedback provides information used to compensate for altered vocal tracts is contrary to prior observations in adults (Garber *et al.,* 1980b; Honda and Kaburagi, 2000) and even young children (Garber, Speidel, and Siegel, 1980a). The null effects observed in these other studies, however, could be related to measurement sensitivity or the task that subjects were asked to perform. For example, Honda and Kaburagi (2000) tracked compensations made to dynamical structural perturbations of the palate shape while we imposed a static perturbation. Recovery from other static perturbations such as the restriction of articulator movement with a bite block is enhanced by the presence of auditory feedback (e.g., Hoole, 1987; Flege *et al.*, 1988; McFarland and Baum, 1995; Baum, McFarland, and Diab, 1996; McFarland *et al.,* 1996).

Even in the absence of vocal-tract modifications, auditory feedback has been shown to increase the precision with which speech categories are produced. For instance, studies of cochlear implant patients for whom feedback can be directly manipulated by turning the implanted device on and off have shown rapid modifications in speaking level, F_0, and vowel formants (Svirsky and Tobey, 1991). Small differences have also been observed in fricatives (Perkell *et al.*, 2000b).

Furthermore, larger effects tend to occur when the implant is turned on compared to when it is suddenly turned off (Perkell *et al.,* 2000a). That is, the improvements observed when deaf speakers receive auditory feedback after a period of time without it are larger than the degradations that appear immediately after feedback is removed. These observations indicate that the speakers maintained the parameters necessary for normal speech production for a period of time after the feedback was removed. This result parallels our own observations that utterances produced while speakers' auditory feedback was masked, improved as a function of their previous practice while feedback was available. In essence, we ''turned off'' the feedback received by subjects and found that the new articulations they learned while feedback was present persisted to some extent when feedback was removed. Thus, the improvements we observed were not strictly due to feedback control but were also a function of learning: The auditory feedback was used by the speech mo-

tor control system to modify an underlying representation mapping vocal gestures to their acoustic consequences.

Although the quality of /s/ productions improved during the masked condition, the new articulatory movements that were learned during the auditory feedback conditions did not completely transfer to production in the absence of feedback. Utterances produced in the presence of feedback consistently received higher quality ratings from listeners. There are a number of possible reasons for this effect. First, it is possible that auditory feedback provides information that the speech motor system can use to adjust ongoing articulation, and over time, these compensatory modifications are learned so that the new speech gestures can be reproduced in the absence of feedback. There is evidence that auditory feedback is used for both online compensation and long-term adaptation. For instance, Houde (1997) asked speakers to whisper one-syllable words while hearing altered auditory feedback. His speakers heard the formants of their vowel productions gradually shifted enough over successive utterances to eventually change the vowels' phonetic identity. Speakers spontaneously compensated for the formant transformations. Houde intermittently tested the speakers' productions in noise and found that speakers slowly adjusted their speech in the same direction as their compensation, suggesting that a modification of the mapping between their vocal-tract productions and their acoustic feedback occurred.

A second possibility is that somatosensory feedback was solely responsible for the improvements in production observed with practice during the masked conditions. There is a large amount of information available to the speech motor control system from proprioceptive and cutaneous receptors in the vocal tract (Gracco, 1995; Kent, Martin, and Sufit, 1990). Vocal-tract manipulations such as the insertion of artificial palates reduce tactile cues regarding, for example, tongue contact against the palate. However, in our study, the prosthesis that extended the subjects' teeth did not reduce sensation in any way. Information regarding tongue–palate contact and cues indicating airflow were unaffected. One might therefore assume that the improvement we observed in the masking condition was due to the fact that the motor system used somatosensory feedback to modify productions in the noise conditions, and not evidence that learning was dependent on auditory information acquired during the feedback conditions. However, our observation that final utterances were judged to be lower in quality than initial productions does not support this conclusion. Nevertheless, the superior quality of production observed when auditory feedback was available to speakers could have been the result of the use of tactile and auditory feedback in combination. Only careful manipulation of both the presence and the absence of auditory feedback in conjunction with manipulations of tactile feedback can satisfactorily resolve the particular contributions of the two modalities (see Hoole, 1987; Gammon *et al.*, 1971; Ringel and Steer, 1963; Scott and Ringel, 1971).

Finally, it should be noted that our data represent the average across speakers. Individual differences in the magnitude of the perturbation and subsequent learning were observed; these differences can have a number of origins. For example, people differ in their response or strategy to vocal-tract perturbations. In addition, although the lengthening of the teeth was relatively equivalent (6 mm), the effective perturbation differed depending on the shape of a speaker's alveolar ridge, the position of their teeth, or how they normally produce an /s/.

V. CONCLUSIONS

The nature of the learning that takes place when adults encounter modified vocal tracts is still a matter of debate. The merit of the unique vocal-tract modification is that all normal tactile information remains intact, allowing a more direct evaluation of the role feedback plays. The results of our investigation show that the availability of auditory feedback can help speakers compensate for structural modifications of their vocal tract. Indeed, the learning we observed only occurred when speakers could hear their speech. In addition, our data also suggest that auditory feedback provides information necessary for long-term modification of a subject's productions.

ACKNOWLEDGMENTS

This work was funded by NIH Grant No. DC-00594 from the National Institute of Deafness and other Communications Disorders, NSERC and CRL Keihanna Human Info-Communication Research Center. We wish to thank Dr. Craig Morison, who made the dental prostheses, and Amanda Rothwell for her help with the collection and preparation of the data in the perceptual study. Thanks also to Christine Shadle for helpful discussions regarding the acoustic analysis of fricatives. Finally, thanks to Yoh'ichi Tohkura for originally suggesting the study.

[1]Because the speakers were exposed to noise through headphones, the utterances they produced both in presence and absence of the masking noise were recorded under the same noise-free conditions. Therefore, listeners were not aware of the acoustic conditions under which the talkers spoke.

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

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

Behrens, S. J., and Blumstein, S. E. (**1988**). "Acoustic characteristics of English voiceless fricatives," J. Phonetics **16**, 295–298.

Benjamin, B. J. (**1997**). "Speech production of normally aging adults," Semin Speech Lang **18**, 135–141.

Binnie, C. A., Daniloff, R. G., and Buckingham, H. W. (**1982**). "Phonetic disintegration in a five-year-old following sudden hearing loss," J. Speech Hear Disord. **47**, 181–189.

Boersma, P. and Weenink, D. (**1996**). Praat, a System for doing Phonetics by Computer, version 3.4. Institute of Phonetic Sciences of the University of Amsterdam, Report 132.

Borden, G. (**1979**). "An interpretation of research on feedback interruption during speech," Brain Lang **7**, 302–319.

Brashers-Krug, T., Shadmehr, R., and Bizzi, E. (**1996**). "Consolidation in human motor memory," Nature (London) **382**, 252–255.

Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (**1998**). "Voice F_0 responses to manipulations in pitch feedback," J. Acoust. Soc. Am. **103**, 3153–3161.

Cowie, R., and Douglas-Cowie, E. (**1992**). "Postlingually acquired deafness," in *Trends in Linguistics, Studies and Monographs, 62* (Mouton de Gruyter, New York).

Donath, T. M., Natke, U., and Kalveram, K. T. (**2002**). ''Effects of frequency-shifted auditory feedback on voice *F*0 contours in syllables,'' J. Acoust. Soc. Am. **111**, 357–366.
Elman, J. L. (**1981**). ''Effects of frequency-shifted feedback on the pitch of vocal productions,'' J. Acoust. Soc. Am. **70**, 45–50.
Evers, H., Reetz, H., and Lahiri, A. (**1998**). ''Cross-linguistic acoustic categorization of sibilants independent of phonological status,'' J. Phonetics **26**, 345–370.
Flege, J. E., Fletcher, S. G., 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.
Fletcher, S. G., and Newman, D. G. (**1991**). ''[s] and [ʃ] as a function of linguapalatal contact place and sibilant groove width,'' J. Acoust. Soc. Am. **89**, 850–858.
Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. (**1988**). ''Statistical analysis of word-initial voiceless obstruents: Preliminary data,'' J. Acoust. Soc. Am. **84**, 115–123.
Fowler, C. A., and Turvey, M. T. (**1980**). ''Immediate compensation in bite-block speech,'' Phonetica **37**, 306–326.
Gammon, S. A., Smith, P. J., Daniloff, R. G., and Kim, C. W. (**1971**). ''Articulation and stress/juncture production under oral anesthetization and masking,'' J. Speech Hear. Res. **14**, 271–282.
Garber, S. R., and Moller, K. T. (**1979**). ''The effects of feedback filtering on nasalization in normal and hypernasal speakers,'' J. Speech Hear. Res. **22**, 321–333.
Garber, S. R., Speidel, T. M., and Siegel, G. M. (**1980a**). ''The effects of noise and palatal appliances on the speech of five-year-old children,'' J. Speech Hear. Res. **23**, 853–862.
Garber, S. R., Speidel, T. M., Siegel, G. M., Miller, E., and Glass, L. (**1980b**). ''The effects of presentation of noise and dental appliances on speech,'' J. Speech Hear. Res. **23**, 838–852.
Gay, T., Lindblom, B. E. F., and Lubker, J. (**1981**). ''Productions of bite-block vowels: Acoustical equivalence by selective compensation,'' J. Acoust. Soc. Am. **69**, 802–810.
Gracco, V. L. (**1995**). ''Central and peripheral components in the control of speech movements,'' in *Producing Speech: Contemporary Issues*, edited by F. Bell-Berti and L. J. Raphael (American Institute of Physics, New York).
Hamlet, S. L. (**1973**). ''Speech adaptation to dental appliances: Theoretical considerations,'' J. Baltimore Coll. Dent. Surg. **28**, 51–63.
Hamlet, S. L., Cullison, B. L., and Stone, M. L. (**1979**). ''Physiological control of sibilant duration: Insights afforded by speech compensation to dental prostheses,'' J. Acoust. Soc. Am. **65**, 1276–1285.
Hamlet, S., and Stone, M. (**1976**). ''Compensatory vowel characteristics resulting from the presence of different types of experimental dental prostheses,'' J. Phonetics **4**, 199–218.
Hamlet, S., and Stone, M. (**1978**). ''Compensatory alveolar consonant production induced by wearing a dental prosthesis,'' J. Phonetics **6**, 227–248.
Hamlet, S., Stone, M., and McCarty, T. (**1978**). ''Conditioning dentures viewed from the standpoint of speech adaptation,'' J. Prosthet. Dent. **40**, 60–66.
Heathcote, A., Brown, S., and Mewhort, D. J. K. (**2000**). ''The Power law repealed: The case for an exponential law of practice,'' Psych. Bull. Rev. **7**, 185–207.
Honda, M., and Kaburagi, T. (**2000**). ''Speech compensation to dynamical structural perturbation of the palate shape,'' in Proceedings of the 5th Seminar on Speech Production: Models and Data, 21–24.
Hoole, P. (**1987**). ''Bite-block speech in the absence of oral sensibility,'' in Proceedings of the XIth International Congress of Phonetic Sciences, **4**, 16–19.
Houde, J. F. (**1997**). ''Sensorimotor adaptation in speech production,'' unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.
Houde, J. F., and Jordan, M. I. (**1998**). ''Sensorimotor adaptation in speech production,'' Science **279**, 1213–1216.
Jesus, L. M. T., and Shadle, C. H. (**2002**). ''A parametric study of the spectral characteristics of European Portuguese fricatives,'' J. Phonetics **30**, 437–464.
Jones, J. A., and Munhall, K. G. (**2002**). ''The role of auditory feedback during phonation: Studies of Mandarin tone production,'' J. Phonetics **30**, 303–320.
Jones, J. A., and Munhall, K. G. (**2000**). ''Perceptual calibration of F_0 production: Evidence from feedback perturbation,'' J. Acoust. Soc. Am. **108**, 1246–1251.
Kawahara, H. (**1995a**). ''Transformed auditory feedback: The collection of data from 1993.1 to 1994.12 by a new set of analysis procedures,'' ATR Technical Report, TR-H-120.
Kawahara, H. (**1995b**). ''Hearing voice: Transformed auditory feedback effects on voice pitch control,'' in Proceedings of the International Joint Conference on Artificial Intelligence Workshop on Computational Auditory Scene Analysis, pp. 143–148.
Kelso, J. A. S., and Tuller, B. (**1983**). ''Compensatory articulation under conditions of reduced afferent information: A dynamic formulation,'' J. Speech Hear. Res. **26**, 217–224.
Kent, R. D., Martin, R. E., and Sufit, R. L. (**1990**). ''Oral sensation: A review and clinical prospective,'' in *Human Communication and its Disorders*, edited by H. Winitz (Ablex, Norwood, NJ).
Lane, H., and Tranel, B. (**1971**). ''The Lombard sign and the role of hearing in speech,'' J. Speech Hear. Res. **14**, 677–709.
Lane, H., and Webster, J. W. (**1991**). ''Speech deterioration in postlingually deafened adults,'' J. Acoust. Soc. Am. **89**, 859–866.
Lindblom, B., Lubker, J., and Gay, T. (**1979**). ''Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation,'' J. Phonetics **7**, 147–161.
Lindblom, B. E. F., and Sundberg, J. (**1971**). ''Acoustical consequences of lip, tongue, jaw, and larynx movement,'' J. Acoust. Soc. Am. **50**, 1166–1179.
McFarland, D. H., and Baum, S. R. (**1995**). ''Incomplete compensation to articulatory perturbation,'' J. Acoust. Soc. Am. **97**, 1865–1873.
McFarland, D. H., Baum, S. R., and Chabot, C. (**1996**). ''Speech compensation to structural modifications of the oral cavity,'' J. Acoust. Soc. Am. **100**, 1093–104.
McGonigle, B. O., and Flook, J. P. (**1978**). ''Long-term retention of single and multistate prismatic adaptation by humans,'' Nature (London) **272**, 364–366.
Narayanan, S. S., Alwan, A. A., and Haker, K. (**1995**). ''An articulatory study of fricative consonants using magnetic resonance imaging,'' J. Acoust. Soc. Am. **98**, 1325–1347.
Nittrouer, S., Studdert-Kennedy, M., and McGowan, R. S. (**1989**). ''The emergence of phonetic segments: Evidence from the spectral structure of fricative–vowel syllables spoken by children and adults,'' J. Speech Hear. Res. **32**, 120–132.
Oller, D., and Eilers, R. (**1988**). ''The role of audition in infant babbling,'' Child Dev. **59**, 441–449.
Osberger, M. J., and McGarr, N. S. (**1982**). ''Speech production characteristics of the hearing impaired,'' in *Speech and Language: Advances in Basic Research and Practice*, edited by N. J. Lass (Academic, New York).
Perkell, J. S., Manzella, J., Wozniak, J., Matthies, M. L., Lane, H., Svirsky, M. A., Guiod, P., Delhorne, L., Short, P., MacCollin, M., and Mitchell, C. (**1995**). ''Changes in speech production following hearing loss due to bilateral acoustic neuromas,'' in Proceedings of the XIIIth International Congress of Phonetic Sciences, **3**, 194–197.
Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M. L., Perrier, P., Vick, J., Wilhelms-Tricarico, R., and Zandipour, M. (**2000a**). ''A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss,'' J. Phonetics **28**, 233–272.
Perkell, J. S., Zandipour, M., Vick, J., Matthies, M. L., Lane, H., Guenther, F. H., and Gould, J. (**2000b**). ''Rapid changes in speech production parameters in response to a change in hearing,'' in Proceeding of the 5th Seminar on Speech Producton: Models and Data, 245–248.
Ringel, R. L., and Steer, M. (**1963**). ''Some effects of tactile and auditory alterations on speech output,'' J. Speech Hear. Res. **6**, 369–378.
Rivers, C., and Rastatter, M. P. (**1985**). ''The effects of multitalker and masker noise on fundamental frequency variability during spontaneous speech for children and adults,'' J. Aud Res. **25**, 37–45.
Rosenbaum, D., Carlson, R., and Gilmore, R. (**2001**). ''Acquisition of intellectual and perceptual-motor skills,'' Annu. Rev. Psychol. **52**, 453–470.
Scott, C. M., and Ringel, R. L. (**1971**). ''Articulation without oral sensory control,'' J. Speech Hear. Res. **14**, 804–818.
Shadle, C. (**1985**). ''The acoustics of fricative consonants,'' in *Research Laboratory of Electronics, Technical Report 506* (Massachusetts Institute of Technology, Cambridge, MA).
Shadle, C. H. (**1991**). ''The effect of geometry on source mechanisms of fricative consonants,'' J. Phonetics **19**, 409–424.
Smith, C. (**1975**). ''Residual hearing and speech production in deaf children,'' J. Speech Hear. Res. **18**, 795–811.
Stoica, P., and Moses, R. (**1997**). *Introduction to Spectral Analysis* (Prentice Hall, Upper Saddle River, NJ).

Stone, M., Faber, A., Raphael, L. J., and Shawker, T. H. (**1992**). ''Cross-sectional tongue shapes and linguopalatal contact patterns in [s], [ʃ], and [l],'' J. Phonetics **20**, 253–270.

Svirsky, M., and Tobey, E. (**1991**). ''Effect of different types of auditory stimulation on vowel formant frequencies in multichannel cochlear implant users,'' J. Acoust. Soc. Am. **89**, 2895–2904.

Ternström, S., Sundberg, J., and Collden, A. (**1988**). ''Articulatory F_0 perturbations and auditory feedback,'' J. Speech Hear. Res. **31**, 187–192.

Waldstein, R. (**1990**). ''Effects of postlingual deafness on speech production: Implications for the role of auditory feedback,'' J. Acoust. Soc. Am. **88**, 2099–2114.

Welch, P. D. (**1967**). ''The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms,'' IEEE Trans. Audio Electroacoust. **AU-15**, 70–73.