

# Nonconscious control of fundamental voice frequency

Honorata Zofia Hafke<sup>a)</sup>

*Institute of Acoustics, Adam Mickiewicz University, Umultowska 85, 61-114 Poznan, Poland*

(Received 4 December 2006; revised 19 September 2007; accepted 2 November 2007)

The aim of this paper is to answer the question whether “perception-action” dissociation, which is well documented in vision, may also be found in auditory information processing. Trained singers were asked to produce vowel sounds into a microphone. The sound that each singer produced was fed back to their ears via headphones. Two seconds after the sound production had begun, the auditory feedback was shifted in pitch by a certain degree (9, 19, 50, or 99 cents in either direction). In every set of sounds, instances without any pitch shifts also appeared. After each trial, participants reported whether they were aware of a pitch change or not. It was found that even though the participants were unaware of subtle pitch changes, the fundamental frequency of their vowel production was found to shift slightly in the opposite direction to the pitch shift. These results show that auditory information is processed by two separate systems: one for perception and one for action. They also show that the function of the auditory control system differs from the visual control system. The latter is used to control bodily movements while the function of the former is a nonconscious, instant control of vocalization. © 2008 Acoustical Society of America.

[DOI: 10.1121/1.2817357]

PACS number(s): 43.66.Hg, 43.70.Bk [DD]

Pages: 273–278

## I. INTRODUCTION

Milner and Goodale (1995) postulate two separate visual systems in the human brain. Evidence from studies of both humans and other primates has shown that there is a distinction between vision for perception and vision for action, which is reflected in the organization of the visual pathways in the cerebral cortex of primates (Aglioti *et al.*, 1995; Goodale *et al.*, 1991; Goodale and Milner, 1992; Goodale and Milner, 2004; Haffenden *et al.*, 2001; Kroliczak *et al.*, 2006). Each stream uses visual information in a different way. The ventral stream transforms visual information into perceptual representations of objects and their relations to each other. Such representations enable us to identify objects and attach various properties to them. In contrast, the transformations carried out by the dorsal stream provide moment-to-moment information concerning the location and disposition of objects, with respect to the effector being used, and thereby mediate the visual control of skilled actions directed toward those objects. In recent years, researchers have attempted to find a similar dissociation between action and perception in human audition (Repp, 2000, 2005). They tried to find an exact auditory analogue for the visual system which controls action. The idea that audition can mimic vision in controlling motor actions such as grasping or reaching objects is difficult to defend. Therefore, if we want to retain the general idea of two separate auditory systems: one for conscious perception and the other for nonconscious control of action, first we have to find out what kind of control could be effectively performed by the system that is responsible for the nonconscious processing of auditory information. The present author assumed that audition is mainly used to control different forms of vocal action. The experiments

were designed to test if this control is performed consciously or nonconsciously. If the experiments show that vocal utterances are controlled nonconsciously, then the hypothesis that there are two separate auditory systems will gain solid support.

Activities whose motor reactions are controlled by audition are for the most part based on the presence of acoustic feedback information. This information is picked up from sounds we generate ourselves, including our own voice, sounds from musical instruments, and sounds produced by other forms of our activity. The information obtained by the auditory system helps to correct the process of producing these sounds so as to meet our expectations. We are constantly informed of the results of our motor activities by acoustic signals generated during these activities. Acoustic waves intercepted by our auditory system allow the appropriate muscles to be activated in order to achieve the desired acoustic effect. The key role of feedback in the process of learning speech, mastering a musical instrument, and keeping a stable rhythm is well known and documented (Smith, 1975; Finney and Palmer, 2003; Repp, 2000; Pfordresher and Palmer, 2006). But, until now, authors have only focused on the functioning of our auditory control and the appropriate reactions to changes in feedback, thereby overlooking the nature of these reactions: they seem to assume that our actions are controlled consciously and engage the system of auditory perception.

The hypothesis tested in this paper is that fundamental voice frequency is tracked and controlled nonconsciously by an auditory system which controls vocal production. The effects of this control are used to correct vocal productions nonconsciously.

It should be stressed that the proposed approach goes a step further than research on the role of the “what” and “where” pathways in the auditory domain (Rauschecker,

<sup>a)</sup>Electronic mail: honorata@ia.spl.amu.edu.pl.

1998; Rauschecker, 2000; Alain *et al.*, 2001; Arnott *et al.* 2004). The advocates of the latter approach claim that the “what” pathway, or the ventral stream, is responsible for the recognition of sound sources; while the “where” pathway, or dorsal stream, is responsible for the perception of location and of the arrangement of sound sources in space. The present author accepts Milner and Goodale’s (1995) hypothesis that the ventral stream processes information for perception while the dorsal processes information for action performance. However, this hypothesis had to be adapted to the sphere of audition by assuming that the dorsal stream in the auditory cortex is primarily responsible for the control of the production of sounds, and in particular for the control of phonation.

Some recent findings support the hypothesis that motor reactions are separated in the auditory modality. Repp (2000) showed that in a synchronized tapping task, subliminal pulse changes in a tone sequence were compensated for. This is an automatic, nonconscious process that is sensitive to the temporal order of information below the perceptual threshold. Action control is also insensitive to perceptual illusions (Repp, 2005). An auditory illusion, such as the influence of an intensity change on perceived timing, does not affect on-line action control. The results obtained in the work of Hickok and Poeppel (2003) suggest that in the case of speech perception a ventral stream is involved in mapping sound onto meaning, while a dorsal stream is involved in mapping sound onto articulatory-based representations.

To observe the dissociation between perception and motor control, a subliminal experimental situation was created. In a psychoacoustic experiment, values below the perceptual threshold were used, i.e., values which were not processed through the normal channels or apparatus of perception. The hypothesis was that a subliminal modification to auditory voice feedback would cause an appropriate correction as a response, even if this change was not actually perceived.

On the assumption that the auditory system functions in the same way as the visual one and processes the information vital for motor reactions in real time, a reaction that would compensate for such a modification should be expected. Two experimental conditions were used:  $F_0$  shifted up and  $F_0$  shifted down. Thanks to recent research (Burnett *et al.*, 1998; Larson *et al.*, 1996; Larson, 1998; Burnett and Larson, 2002; Natke *et al.* 2003; Yi Xu *et al.*, 2004; Sivasankar *et al.*, 2005), it has been established that during vocalization subjects react to perturbations in the pitch of voice feedback by changing their fundamental voice frequency ( $F_0$ ), usually to compensate for the pitch shift. Responses indicating a negative feedback system which stabilizes  $F_0$  occur with a latency of 100–160 ms (Burnett *et al.*, 1998; Hain *et al.*, 2000; Natke and Kalveram, 2001). The experiments did not focus on the way the listeners perceived the modifications. Previous research (Klatt and Zue, 1971; Pape and Mooshammer, 2006) determined the just-noticeable difference (JND) of fundamental voice frequency contours for digitally synthetic stimuli. Depending on the experimental conditions and the stimuli used (i.e., different vowels), it was found that pitch shifts ranging from 4 to 98 cents constituted the averaged JND. Under the experimental conditions of this study, and

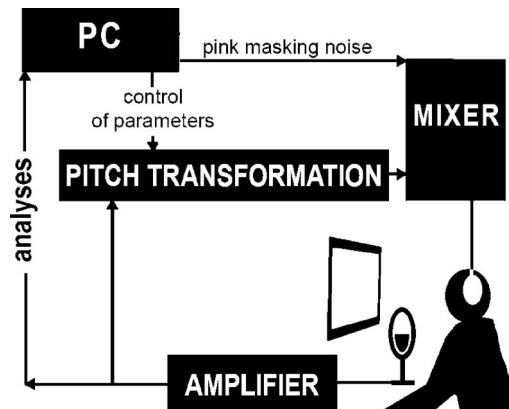


FIG. 1. Schema of the experimental acoustic feedback setup.

using the vowel “u,” the averaged threshold value below which pitch shifts could not be observed was 26 cents. Throughout all the trials, changes in the subjects’ fundamental voice frequency were examined. Reactions to subliminal pitch perturbations of voice feedback were also checked by an analysis of changes in fundamental voice frequency ( $F_0$ ).

## II. METHOD

### A. Subjects

Nine adults between 21 and 28 years of age participated in this study. All the listeners qualified as having normal hearing, which was defined as the audiometric threshold of 20 dB hearing level, or better, for a range from 250 to 8000 Hz (ANSI, 1996). They reported no neurological defects and had no speech or voice disorders. All of them were trained singers. The subjects were seated in a sound-treated room.

### B. Stimuli

The subjects’ voices were recorded with a Shure SM 58 microphone (with a 6 cm mouth-to-mike distance), amplified with a Behringer mixer model 802A, and processed for auditory feedback pitch shifting through a DP2 Ensoniq ultraharmonizer. The pitch-shift processing introduced a small delay of 10 ms. The output of the harmonizer was mixed with pink masking noise [75 dB sound pressure level (SPL)] and presented to the subject over Sennheiser HD 600 headphones. The experiment schema is presented in Fig. 1.

Because previous studies (Burnett *et al.*, 1998) showed no relationship between voice  $F_0$  response and voice intensity, the intensity was not under strict control in this study.

### C. Procedure

In the experiment subjects were instructed to vocalize the vowel /u/ for 5 s. They were asked to maintain constant pitch and loudness. They heard their own voice via headphones. The voice signal was mixed with pink noise (70 dB SPL) to partially mask bone-conducted auditory feedback. The experiment schema is presented in Fig. 2.

At the beginning of the third second of each vocalization, voice feedback was pitch shifted by 9, 19, 50, or 100 cents, respectively, in both directions. In every set of sounds,

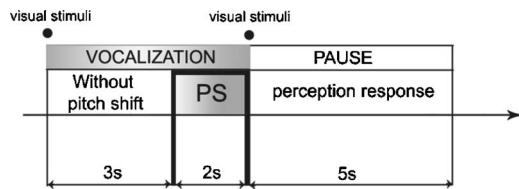


FIG. 2. Schematic illustration of stimulus presentation.

instances without any pitch shifts also appeared. Pitch shifts in the auditory feedback signal were controlled by a MIDI based control system. During a 5 s break between each vocalization, subjects were required to report whether the trial had contained a pitch shift and, if yes, what direction it was. Both the beginning and the ending of phonation were marked with visual stimuli. Pitch shift values were randomly ordered. Every pitch shift value was repeated 30 times.

At the same time, the listener's voice was digitally recorded and changes in their fundamental voice frequency were examined in order to check subject's actual reactions to pitch shifts.

In order to get information about threshold perception in pitch shift, the results from all the subjects had to be divided into two categories: results where subjects heard pitch changes during the vocalization, and results where subjects did not perceive any pitch changes during the vocalization. The results were presented in the form of psychometric curves, with a threshold value of 75% for correct detections. Additionally, the individual results were analyzed to see whether the direction of the fundamental voice frequency shift was appropriately detected. For each of the listeners, percentage values representing the number of correct answers to the pitch shift stimuli were calculated. The threshold was set at 50% correct answers, in order to intensify the criteria of the experiment and verify that the subject was actually able or not able to determine the pitch shift direction.

#### D. Analysis

The signal was low-pass filtered at 500 Hz in order to remove high frequency harmonics. The fundamental frequency of vocalization during each trial was calculated using an algorithm incorporated in the Praat software (Boersma, 1993). The data was then analyzed using a special application developed in the MATLAB environment. For each trial, an average value of  $F_0$  and a standard deviation were calculated for the period before the onset of the pitch shift. Subsequently, a reaction check was performed. A positive reaction was defined as a deviation in  $F_0$ , which had a latency of at least 60 ms, a magnitude of more than 2 SDs of the 1000 ms pre-stimulus mean, and a duration of at least 120 ms in a maximum of 800 ms after the pitch shift (Fig. 3). The application automatically determined the time when the averaged signal departed and reentered the 2 SD response criterion, and calculated the corresponding values in cents for valid responses only.

Changes in the fundamental frequency were converted to cents using the following formula (Hain *et al.*, 2000):

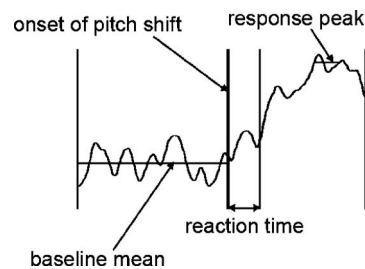


FIG. 3. Example of  $F_0$  response of a subject to downward pitch stimuli.

$$Cents = 100 * \left( \frac{10 * \text{Log}(P/M)}{\text{Log}(2)} \right),$$

where  $P$  is response peak  $F_0$  in hertz, and  $M$  is the base line mean  $F_0$  in hertz.

### III. RESULTS

More than 80% of the signals fully corresponded to the predefined criteria. Among the analyzed signals, in 89% of the answers there was compensation for the applied change (opposing responses), and in 11% there was no such compensation (following responses). In the display of results, the results of control trials (sets of sounds without pitch shifts) were omitted, because in the trials more than 70% of motor reactions were classified as invalid.

The results are discussed at two levels: as group data and individual data. The averaged results of the detection task and the averaged results of the motor response are presented in Figs. 4(a) and 4(b), respectively.

Figure 4(a) represents the averaged results for the perception task of the experiment. Percentage values represent the number of detections of the pitch shift. The average motor response values (in cents) are presented in Fig. 4(b). From the equation of the accumulative standard distribution curve, threshold values for the probability of 75% were calculated. The averaged threshold values were -21 cents for the downward pitch shift of fundamental voice frequency, and 30 cents for the upward pitch shift. Therefore the values -19, -9, 9, and 19 cents were under the 75% threshold, and were thus in the same range as those gathered in previous experiments of a similar nature (Klatt and Zue, 1971; Klatt 1973; Pape and Mooshammer, 2006). However, in the previous studies thresholds were determined in different experimental conditions.

The results were also presented on an individual level in order to show the correct detections of the direction of pitch shift. Individual results of the detection task, along with the corresponding motor reactions, are shown in Fig. 5. The column on the left represents the results of the perception part of the experiment. Percentage values represent the number of correct answers to the pitch shift. Only one subject perceived all the changes in voice pitch correctly. For seven subjects pitch shift values -19, -9, and 9 were under the 50% threshold. This means that, for these values, the listeners were unable to perceive the direction of the pitch shifts. Therefore, these values can be identified as subliminal for perception.

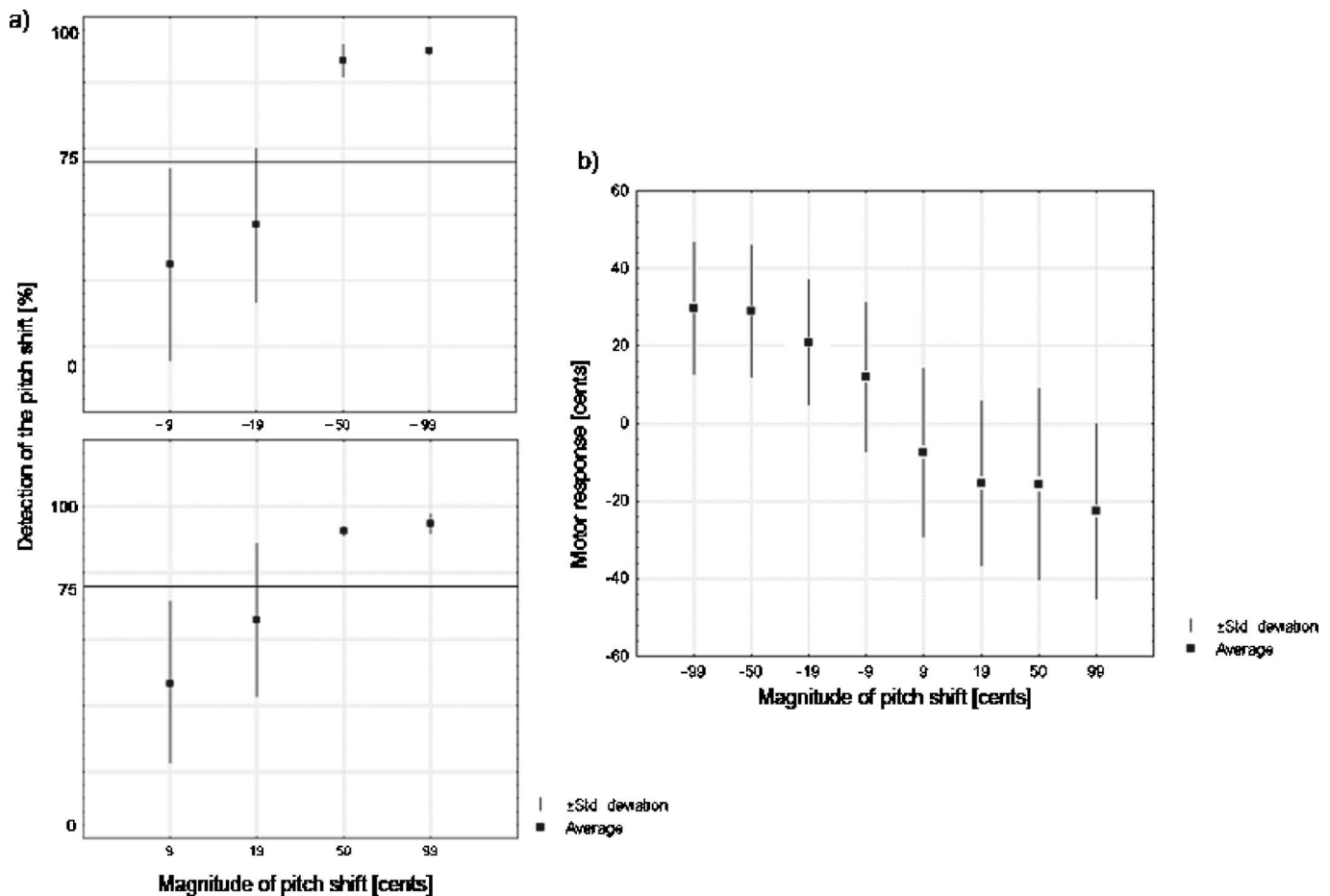


FIG. 4. The averaged results of the detection part of the experiment. Percentage values show the number of detections of the pitch shift (a). The averaged results of the motor part of the experiment (b).

An interesting asymmetry of the results can be observed: it was easier to perceive upward pitch shifts than downward pitch shifts.

The individual motor response values (in cents) for each subject are presented in the right-hand column. It can be clearly seen that when the pitch feedback is shifted down, the subjects raise their pitch, which contrasts with the pitch feedback being shifted up in order to compensate for the pitch change in the auditory feedback. The same response can also be observed for shifts which are subliminal to perception. The post-hoc T Tukey (HSD) test was carried out for stimuli that were not noticed perceptually. The test showed the statistical significance of differences between reaction values to upward and downward pitch shifts for the majority of conditions in all cases.

#### IV. DISCUSSION

Recent studies (Burnett *et al.*, 1998; Larson *et al.*, 1996; Larson, 1998; Burnett and Larson, 2002; Hain *et al.*, 2000; Jones and Munhall, 2000; Natke *et al.*, 2003; Yi Xu *et al.*, 2004; Sivasankar *et al.*, 2005) in the field of fundamental voice frequency control have clearly shown that the audio-vocal system controls the stability of fundamental voice frequency by compensating for the changes applied to feedback. It has been proven that the so-called pitch shift reflex

process is of automatic nature (Burnett *et al.*, 1998), although its neural mechanism is still poorly understood.

The results of the experiment suggest that the stream utilized to control fundamental voice frequency could be both independent of the perception stream and cooperate with it at the same time. In the results of the experiment presented in this study there are vast differences between reported perceptions and motor reactions. In many cases, listeners reacted appropriately to voice pitch shifts, although they did not perceive the change. This fact suggests that a certain part of the information contained in the acoustic signal is also processed subconsciously for calculations necessary for appropriate motor reactions. The present experiment was carried out on a group that consisted only of trained singers. The decision to choose this type of group was based on the fact that, because of the subjects' training in singing, random changes in voice pitch before the onset of pitch shift were not significant to the results of the experiment. It can be presumed, however, that in the case of people not trained in singing, the same mechanism of fundamental voice frequency control operates. The results of previous studies (Hain, 2000) revealed that the process of controlling fundamental voice frequency could be independent of our will. It also suggests a possible division into perception and motor streams.

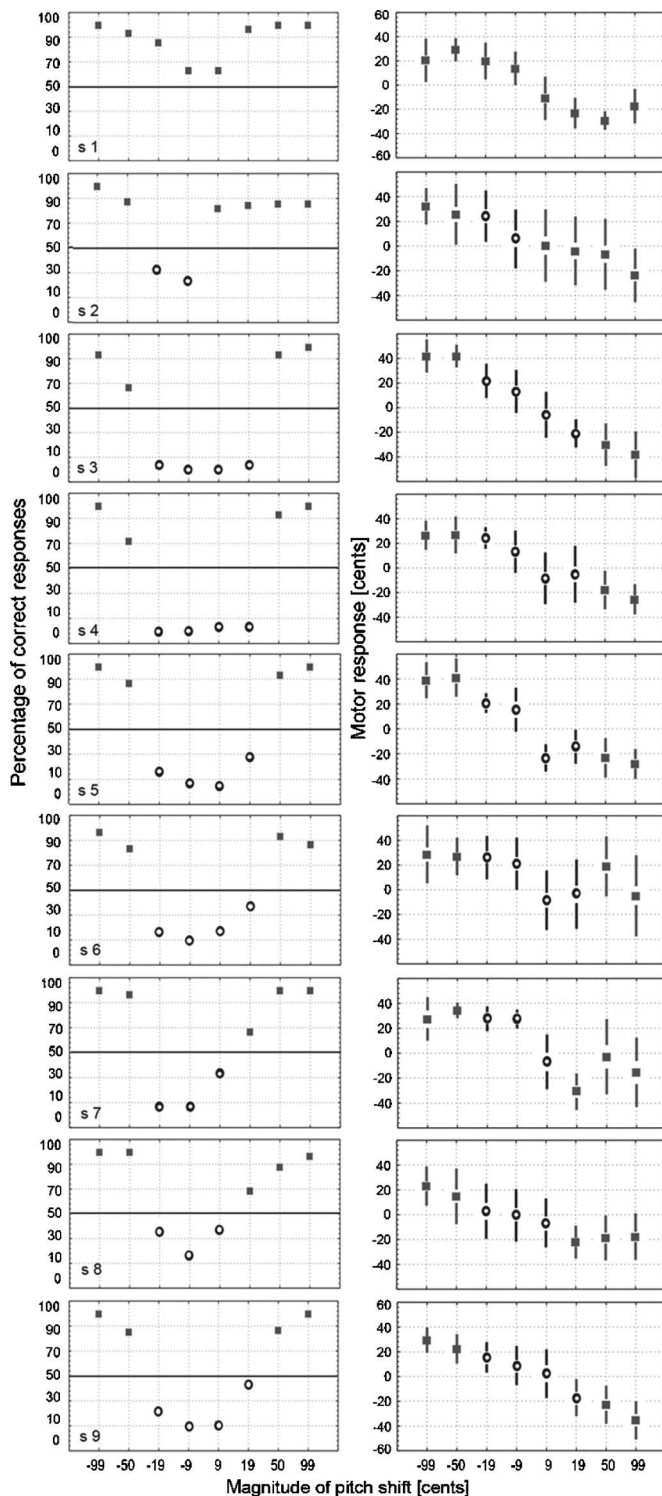


FIG. 5. Results with appropriate standard deviations for all subjects. The column on the left represents the results of the perception part of the experiment. Percentage values show the number of correct answers to the pitch shift. Circles aligned under the line corresponding to the perception threshold stand for values that were not registered perceptually. The column on the right represents the results of the motor part of the experiment. Reactions that were not registered in the perception part of the experiment were singled out.

In their neurological studies, [Hickok and Poeppel \(2003\)](#) discovered the division of two separate streams in cortical processing. One of those streams attached a certain meaning to the sound event, while the other articulated representations

of the sound itself. The latest studies ([Purcell and Munhall, 2006](#)) show that auditory feedback during speech production plays a significant role in the maintenance of accurate articulation. Any real time changes that modified the frequency of the first formant caused a reaction which compensated for those changes by adjusting the parameters of the vocal track. Such a reaction is strictly a motor reaction. All the experiments show that acoustic feedback is necessary for the regulation of the parameters of produced sounds, as it triggers appropriate motor reactions. In this study, by arranging a specific experimental situation, we were able to separate the functioning of the two streams. The use of stimuli situated beneath the threshold of perception allowed us to isolate the motor stream in order to observe its functioning. A similar situation was arranged by Repp in his study of the perception of rhythm with simultaneous finger tapping. In his study he showed that listeners are capable of reacting to subliminal changes in a rhythmic sequence, even though such changes are not registered by their perception. The intent of our study was to verify the possibility of there being motor reactions beyond conscious control which are used to control fundamental voice frequency. The experiment dedicated to this purpose verified that this is the case for subliminal changes in acoustic feedback. Such a reaction also took place with distinctive changes, but according to the previous studies ([Burnett et al., 1998](#); [Larson et al., 1996](#); [Larson, 1998](#), [Burnett and Larson, 2002](#); [Hain et al., 2000](#)) it did not fully compensate for the changes introduced into feedback.

## V. CONCLUSION

1. The results of the experiment carried out suggest that the motor control of fundamental voice frequency functions properly, even when the pitch shifts are not being consciously perceived.
2. The obtained results suggest that in the case of the motor control of phonation, a dissociation between perception and action for auditory information takes place.

## ACKNOWLEDGMENTS

Grateful thanks to Professor Anna Preis and Professor Andrzej Klawiter for extensive assistance, many helpful comments and discussions.

Alain, C., Arnott, S. R., Hevenor, S. J., Graham, S., and Grady, C. L. (2001). "What' and 'where' in the human auditory system," *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12301–12306.

ANSI. ANSI S3.6-1996, (1996). Specifications for Audiometers, American National Standards Institute, New York.

Aglioti, S., DeSouza, J., and Goodale, M. A. (1995). "Size-contrast illusions deceive the eyes but not the hand," *Curr. Biol.* **5**, 679–685.

Arnott, S. R., Binns, M. A., Grady, C. L., and Alain, C. (2004). "Assessing the auditory dual-pathway model in humans," *Neuroimage* **22**, 401–408.

Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (1998). "Voice F0 responses to manipulations in pitch feedback," *J. Acoust. Soc. Am.* **103**(6), 3153–3161.

Burnett, T. A., and Larson, C. R. (2002). "Early pitch shift response is active in both steady and dynamic voice pitch control," *J. Acoust. Soc. Am.* **112**, 1058–1063.

Boersma, P. (1993). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," *Proceedings of the Inst. of Phonetic Sciences* **17**, 97–110.

Finney, S. A., and Palmer, C. (2003). "Auditory feedback and memory for

- music performance: Sound evidence for an encoding effect," *Mem. Cognit.* **31**, 51–64.
- Goodale, M. A., Milner, A. D., Jakobson, L. S., and Carey, D. P. (1991). "A neurological dissociation between perceiving objects and grasping them," *Nature (London)* **349**, 154–156.
- Goodale, M. A., and Milner, A. D. (1992). "Separate visual pathways for perception and action," *Trends Neurosci.* **15**, 20–25.
- Goodale, M. A., and Milner, M. A. (2004). *Sight Unseen: An Exploration of Conscious and Nonconscious Vision*, Oxford: (Oxford University Press, Oxford), p. 140.
- Haffenden, A. M., Schiff, K. C., and Goodale, M. A. (2001). "The dissociation between perception and action in the Ebbinghaus illusion: Nonillusory effects of pictorial cues on grasp," *Curr. Biol.* **11**, 177–181.
- Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., and Kenney, M. K. (2000). "Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex," *Exp. Brain Res.* **130**, 133–141.
- Hickok, G., and Poeppel, D. (2003). "Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language," *Cognition* **92**, 67–99.
- Jones, J. A., and Munhall, K. G. (2000). "Perceptual calibration of F0 production: Evidence from feedback perturbation," *J. Acoust. Soc. Am.* **108**(3), 1246–1251.
- Klatt, D. H., and Zue, V. W. (1971). "Just—noticeable differences for selected aspects of synthetic fundamental—frequency contours," *J. Acoust. Soc. Am.* **49**, 86–87.
- Klatt, D. H. (1973). "Discrimination of fundamental frequency contours in synthetic speech: Implications for models of pitch perception," *J. Acoust. Soc. Am.* **53**, 8–16.
- Kroliczak, G., Heard, P., Goodale, M. A., and Gregory, R. L. (2006). "Dissociation of perception and action unmasked by the hollow-face illusion," *Brain Res.* **1080**, 9–16.
- Larson, C. R., White, J. P., Freedland, M. B., and Burnett, T. A. (1996). "Interactions between voluntary modulations and pitch-shifted feedback signals: Implications for neural control of voice pitch," in *Vocal Fold Physiology: Controlling Complexity and Chaos*, edited by P. J. Davis and N. H. Fletcher (Singular, San Diego), pp. 279–289.
- Larson, C. R. (1998). "Cross-modality influences in speech motor control: The use of pitch shifting for the study of F0 control," *J. Commun. Disord.* **31**, 489–503.
- Milner, A. D., and Goodale, M. A. (1995). *The Visual Brain in Action* Oxford (Oxford University Press, Oxford).
- Natke, U., Donath, T. M., and Kalveram, K. T. (2003). "Control of voice fundamental frequency in speaking versus singing," *J. Acoust. Soc. Am.* **113**(3), 1587–1593.
- Natke, U., and Kalveram, K. T. (2001). "Fundamental frequency under frequency shifted auditory feedback of long stressed and unstressed syllables," *J. Speech Lang. Hear. Res.* **44**, 577–584.
- Pape, D., and Mooshammer, C. (2006). "Is intrinsic pitch language-dependent?—evidence from a cross-linguistic vowel pitch experiment (with additional screening of the listeners' DL for music and speech)," in *MULTILING-2006*, paper No. 018.
- Pfordresher, P. Q., and Palmer, C. (2006). "Effects of hearing the past, present, or future during music performance," *Percept. Psychophys.* **68**, 362–376.
- Purcell, W. D., and Munhall, G. K. (2006). "Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation," *J. Acoust. Soc. Am.* **120**, 966–977.
- Rauschecker, J. P. (1998). "Cortical processing of complex sounds," *Curr. Opin. Neurobiol.* **8**(4), 516–521.
- Rauschecker, J. P., and Biao, T. (2000). "Mechanisms and streams for processing of "what" and "where" in auditory cortex," *Proc. Natl. Acad. Sci. U.S.A.* **97**(22), 11800–11806.
- Repp, B. H. (2000). "Compensation for subliminal timing perturbations in perceptual-motor synchronization," *Psychol. Res.* **63**, 106–128.
- Repp, B. H. (2005). "Does an auditory perceptual illusion affect on-line auditory action control? The case of (de)accentuation and synchronization," *Exp. Brain Res.* **168**, 493–504.
- Sivasankar, M., Bauer, J. J., Babu, T., and Larson, Ch. R. (2005). "Voice responses to changes in pitch of voice or tone auditory feedback," *J. Acoust. Soc. Am.* **117**(2), 850–857.
- Smith, C. R. (1975). "Residual hearing and speech production in deaf children," *J. Speech Hear. Res.* **18**, 795–811.
- Yi, Xu, Larson, C. R., Bauer, J. J., and Hain, T. C. (2004). "Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences," *J. Acoust. Soc. Am.* **116**(2), 1168–1178.