VOICE FEEDBACK CONTROL WITH SUBJECT-IN-THE-LOOP

Dan Necsulescu, Lin Zhang, Elisha Pruner
University of Ottawa, Department of Mechanical Engineering, Ottawa, Ontario, Canada

Jerzy Sasiadek
Carleton University, Department of Mechanical and Aerospace Engineering, Ottawa, Ontario, Canada

Keywords: Signal Processing, Voice Feedback, Subject-in-the-Loop, Real-time Requirements, Signal Processing, Voice Training.

Abstract: A subject-in-the-loop feedback control system is composed of a bioengineering system including a subject, whose voice is received by a microphone, a computer that achieves the required signal processing of the sound signal by temporal and/or spectral computations and a speaker or earphones for auditory feedback to the subject of voice training. Frequency domain modifications of the signal are intended for voice training of subjects already familiar with traditional voice training. The objective of this paper is to present alternative methods for the implementation of subject-in-the-loop feedback control systems developed for voice training. The proposed feedback control scheme is an extension of the traditional control systems; feedback sensing and the control law are achieved by the human subject as a self-organizing controller. The experimental set-ups, developed for this purpose, contain programmable digital devices for real time modifications of the frequency content of the voice signal. The paper presents also a preliminary solution that satisfies the requirements for real-time operations, in particular that the subject does not perceive the delay between the sound generation and the auditory reception of the modified sound. The system performs spectral calculations for the analyses of the vocal sound signals. Preliminary experimental results illustrate the operation and the features of the proposed subject-in-the-loop real-time system for voice training.

1 INTRODUCTION

The purpose of this work is to develop a bioengineering experimental set-up with real-time capability for voice signal processing in a closed loop configuration. Acoustic loops refer to systems for signal amplitude increase or decrease of certain sound frequencies using digital manipulation of sound samples. The goal of this research is the design and construction of computer based modules for actors' voice training, as well as, for singers and public speakers who were already subjected to traditional voice training. Such an Audio-Formant Mobility Trainer is an adjunct to voice-training. The reason for previous training requirement results from the fact that the subject will have to produce different voice qualities for which it is necessary to have acquired a certain mobility of the bodily parts that produce speech. The device is intended to facilitate the production of new voice qualities by increasing the mobility of one's voice formants. The Audio-Formant Mobility Trainer is a module that can perform acoustical experiments with the subject's voice and band-pass filtering for each of the formants for the purpose of auditory-feedback. Any formant can hence be chosen to be manipulated in order to increase the ability of the subject to perceive it in his own voice. There are three types of formant manipulations:

1. Intensity: varying the relative intensity of the formant bandwidth ranging from filtering it out to increasing it above the spectral envelope.
2. Bandwidth: increasing and decreasing its width.
3. Pitch: increasing and decreasing its pitch.

Typically, the learner uses a microphone and headphones while singing or speaking. His voice is analyzed and processed by the computer. This training could be very useful for actors and singers who are called to produce different voice types. It

Necsulescu D., Zhang L., Pruner E. and Sasiadek J. (2010).
VOICE FEEDBACK CONTROL WITH SUBJECT-IN-THE-LOOP.
In *Proceedings of the 7th International Conference on Informatics in Control, Automation and Robotics*, pages 221-225
DOI: 10.5220/0002898502210225

may be also helpful for learners of new vowels using analysis called LTAS (Long Time Average Spectra), thereby determining the formants that characterize foreign language. Preliminary research work (Nesulescu and Weiss et al., 2006, 2005, 2008) led to the confirmation that complex acoustic phenomena can be simulated for the needs of designing acoustic hardware in the form of a closed loop experimental set-up for acoustics analysis. The presence of the subject in the control loop results in interesting new issues for the feedback control design.

2 AUDITORY FEEDBACK AND VOICE PRODUCTION

Previous research showed that changing voice quality by altering the auditory perception of one's voice is, to a limited degree, possible. If a person's sound production possibilities are enlarged (through voice training), then altered auditory feedback might facilitate the generation of different voice qualities (Necsulescu, Weiss, and Pruner, 2008). The set-up consists in a subject hearing his voice through headphones while speaking into a microphone. However, the process allows a series of digital manipulations (temporal and spectral) designed to affect perception while examining the effects on vocal output. Whereas, the intensity feedback manipulations have been studied extensively (Purcell, and Munhall, Vol. 119 2006), (Purcell and Munhall 120, 2006), spectral changes effects on voice quality in auditory feedback and their relationship to voice production are still relatively unknown. Original proponents of the use of servo mechanical theory have claimed a direct effect on the vocal output when modified voice is fed back to the speaker. Essentially, according to this theory, if certain bandwidths of the voice spectra are modified in such a manner as to increase or decrease the energy in those regions, the person emitting those sounds will unconsciously react if the modified voice signal is fed back to his ears. The possibility of affecting voice output by auditory feedback remains a topic of intense interest for those involved in voice, speech and accent training (Necsulescu, Weiss and Pruner, 2008). This work has the long-term goal to carry out audio-vocal filtering experiments including subjects with or without vocal training in order to determine whether voice training could allow for vocal adjustments in conditions related to filtered auditory feedback. This paper describes the construction of computer based module for auditory feedback with no perceived temporal delay.

There are many teaching techniques in voice training, some auditory, some based on movement and some mixed. Independently of the technique, certain pedagogical approaches are often used. One of such techniques is bodily awareness through minimal movements (Purcell and Munhall, 2006). This objective of this approach is an effortless speech-motor learning system. A variable is introduced and the subject perceives it, plays with it, explores it, adjusts to it and integrates it in his own behaviour. This is the purpose of the Audio-Formant Mobility Trainer, an adjunct to voice-training when the learner has had already preliminary training with any traditional technique. The reason for the need for previous training is that the subject will have to produce different voice qualities for which it is necessary to have acquired a certain control of the mobility of the bodily parts that produce speech. The purpose of the device is expected to facilitate the production of new voice qualities by increasing the mobility of one's voice formants.

3 DESCRIPTION OF EXPERIMENT

The first experiment tries to ascertain whether it is possible to teach subjects to vary their fourth formant (F_4) at will. Previous research (Purcell, and Munhall, Vol. 119, 2006) has shown that subjects do it unconsciously when their auditory feedback is manipulated while uttering vowels. It is also known (Purcell and Munhall, 2006), that formant manipulation in pitch and bandwidth changes significantly the perceived voice quality

4 EXPERIMENTAL SET-UP

The main difficulty until recently was to achieve real-time capability in auditory feedback with programmable digital hardware. Some delay in auditory feedback cannot be avoided, but it is desired to reduce it, such that it will not be perceived.

The block diagram of the complete auditory feedback system is shown in Figure 1. Figure 2 shows this system in the traditional control system block diagram form. A human subject carries out in this case the feedback sensing, the comparator and the controller (regulator). Figure 3 shows the block diagram of the PC based auditory feedback system, while Figure 4 shows the Simulink^{R} diagram of the real-time system for signal acquisition from AI, filtering, FFT frequency analysis, display and headphone signal generation to AO (Necsulescu, Weiss and Pruner.

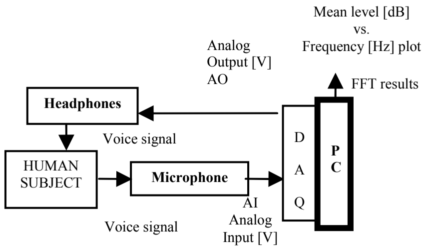

Figure 1: Block diagram of a generic auditory feedback system.

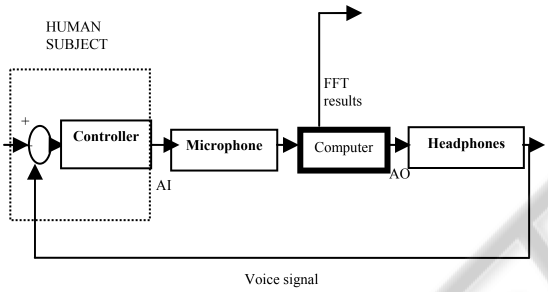

Figure 2: Block diagram of the auditory feedback system in feedback control (regulator) configuration.

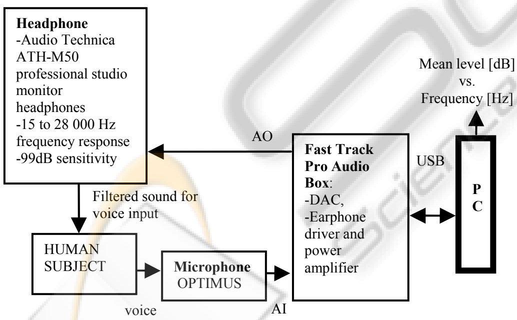

Figure 3: Block diagram of the PC based auditory feedback system.

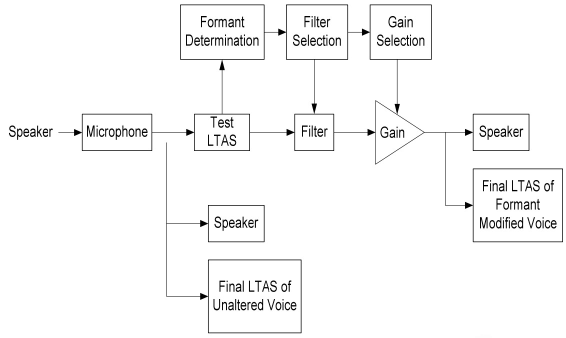

Figure 4: The diagram of the real-time system for signal acquisition from AI, filtering, FFT frequency analysis, display and headphone signal generation to AO.

5 PRELIMINARY EXPERIMENTAL SET-UP TESTING RESULT

The experimental setup was tested for verifying its performance. The current subject, used for experiments, has had extensive voice training. He sang for each audio-vocal filtering condition a 60 seconds French song using a neutral vowel. MATLAB representation of the amplitude versus time and the calculation of the Long Term Average Spectra (LTAS), permits the evaluation of the effects of voice signal processing (Purcell and Munhall, 2006). Figure 5 shows the frequency domain results of the sound signal in case of no headphones. These results are post-processed in frequency domain for the identification of formants

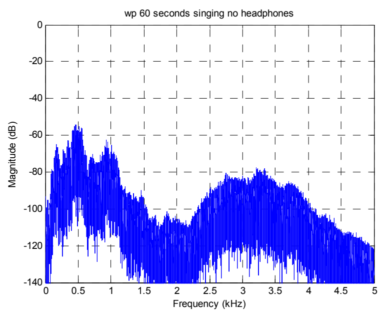

Figure 5: Results for LTAS of the sound signal with no headphones.

Figure 6 shows formant manipulation of the voice with LTAS for 500 Hz bandstop filtering. This figure shows what the subject heard following signal manipulation.

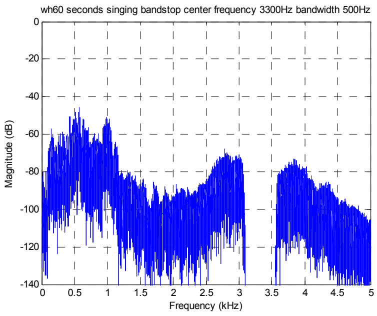

Figure 6: Formant manipulation of the voice with LTAS for 500 Hz bandstop filtering.

Bandstop hearing while singing produces a typical 3 zone spectrum: 1. The highest peaks, from fundamental frequency to 1100 Hz; 2. The second highest peaks, from 2500 to 3500-4000 Hz; 3. A Bowl, from 1100 to 2500 Hz.

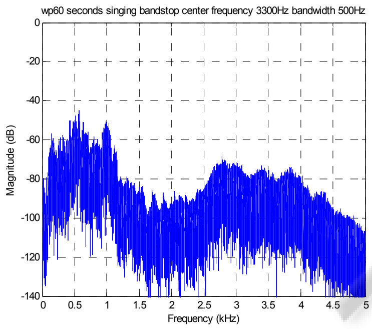

Figure 7: Recording the effects for the 500 Hz bandstop filtering.

The 500 Hz bandstop produces a clear peak between 3500-4000 Hz. This indicates that the bigger the gap the bigger the compensation. After hearing the signal shown in Figure 6, the resulting spectral form from Figure 7 appear similar in this case to the one produced while singing without headphones, shown in Figure 5. Figure 8 shows the results of the intensified amplitude of a 500 Hz about 3300 Hz. After hearing the signal shown in Figure 8, the subject produces the signal with the spectral content shown in Figure 9. The spectral form from Figure 9 differs significantly from the results from Figure 6, for the case of no headphones. The spectrum is flattened when compared to the three-zone spectrum of no headphones condition. This confirms that the real time signal was reduced to the desired frequency domain and the audio test based on the system shown in Figure 1 confirmed subjectively the validity of this result. This preliminary confirmation of the significant effects of signal processing on the subject is an encouraging result for the prospect of using it in voice training. The subject produced different voice qualities unconsciously when given different auditory feedback conditions. The approach seems useful for training the voice for different voice qualities.

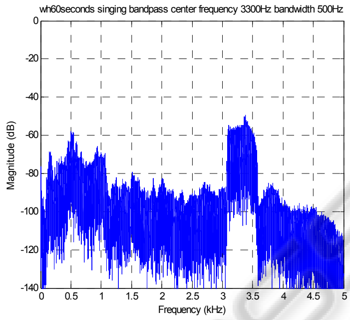

Figure 8: Formant manipulation of the voice with LTAS for 500 Hz frequency band amplification.

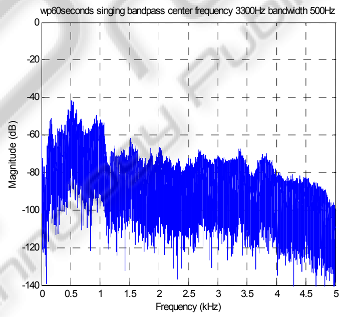

Figure 9: Recording of the effects of the 500 Hz frequency band amplification.

6 CONCLUSIONS

The experimental setup presented in this paper is useful for the voice training experiments in selected subjects. It includes:

- Long term average spectra under several conditions;
- Processing and storing data in the computer;
- Comparison of spectra with filtering and, without filtering;
- Comparison of experimental [trained] and non-trained groups.

The proposed auditory feedback experimental set-up proved to satisfy the requirements of acquiring signals with a microphone and filtering procedure, as well as, displaying and transmitting the modified signals to the earphones in real time. Moreover, auditory signals were FFT processed for the successful identification of each particular subject formants. This experimental set-up was deemed appropriate for purposes of voice training for selected subjects.

REFERENCES

Necsulescu, D., Zhang, W., Weiss, W., Sasiadek J., “Room Acoustics Measurement System Design using Simulation”, *IMTC 2006 - Instrumentation and Measurement Technology Conference, Sorrento*, Italy 24-27 Apr 2006.

Necsulescu, D., Advanced Mechatronics, *World Scientific*, 2009.

Necsulescu, D, Weiss, W., Zhang, W., “Issues Regarding Hardware in the Loop Experimental Set-up for Room Acoustics”, Montreal, McGill University, *20th Canadian Congress of Applied Mechanics*, 2005.

Necsulescu, D., Weiss, W., Pruner, E., “Acousto-Mechatronic System for Voice Training”, *Bul. I. P. Iasi, Tom LIV (LVIII), Fasc. X*, 2008, pp. 481 - 487.

Purcell, D. W. and K. Munhall. “Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation”. *Journal of the Acoustical Society of America 120*(2), (2006), pp. 966-977.

Purcell, D. W. and Munhall, K. “Compensation following real-time manipulation of formants in isolated vowels”, *J. Acoust. Soc. Am. Vol. 119*, Apr. 2006, 2288–2297.