A NEW FEATURE SELECTION METHOD FOR VOLUME CONTROL IN DIRECT-LEARNING HEARING AID SYSTEMS

Jin Zhou[1], Hisham Othman[1], Hilmi Dajani[1], and Tyseer Aboulnasr[2]
[1] Dept. of Electrical Engineering, University of Ottawa, 800 King Edward Av., ON., Canada, K1N 6N5, tenniszj@163.com
[2] Faculty of Applied Science, University of British Columbia,5000-2332 Main Mall, BC., Canada, V6T 1Z4

1. INTRODUCTION

A desirable feature of modern hearing aids is the ability to automatically adjust its behavior in different acoustic environments. There are two kinds of approaches that can achieve this. One is based on "classification" of environments, another is "direct learning" of preferred hearing aid settings. Our focus is on the second approach, in which an artificial neural network (ANN) learns the preferred volume setting of the hearing aid user. The performance of such a direct-learning system strongly depends on the chosen signal features. While a large number of features have been derived for environment classification (Büchler, 2002; Nordqvist *et al*, 2004), we are now aware of any that have been derived specifically for "direct learning", which has different requirements. For example, environment classification should not in general be sensitive to the sound volume, whereas direct learning of volume setting not only depends on the volume, but also depends on how this volume affects speech intelligibility and user comfort. Moreover, the preferred volume setting depends on the hearing loss profile, and whether it is profound, severe, or moderate. The goal of this work is to derive suitable features that the ANN will use to set the volume such that it optimizes speech intelligibility. New features are proposed, which are based on measures of speech intelligibility, namely the Speech Intelligibility Index (SII) (ANSI S3.5-1997) and the Coherence SII (CSII) (Kates and Arehart, 2005). The performance of these features is then investigated using a simulator of a hearing aid user (SPOT, 2009).

2. METHOD

2.1 Perceptually-dependent Features

The new features are derived from the calculation of the CSII in third-octave bands (ANSI S3.5-1997; Kates *et al*, 2005), and they reflect the SNR and energy of the speech signal in these bands, weighted by the psychocacoustic characteristics of the listener. They are calculated as follows:

- The pure and distorted speech files, $x(n)$ and $y(n)$, respectively, are divided into 50% overlapping segments. Each segment is multiplied by a Hamming window, and then the segments $y_m(n)$ are categorized into three groups for high, medium and low energy segments.

In each energy group, the following steps are repeated to get the CSII in each band.

1. The spectra of $x_m(n)$ and $y_m(n)$, i.e. $X_m(k)$ and $Y_m(k)$ are obtained using FFT, and then the coherence measure for each group is estimated as follows:

$$|\gamma(k)|^2 = \frac{\left|\sum_{m=0}^{M-1} X_m(k) Y_m^*(k)\right|^2}{\sum_{m=0}^{M-1}|X_m(k)|^2 \sum_{m=0}^{M-1}|Y_m(k)|^2} \quad (1)$$

where m is the segment index, k is the FFT bin index and M is the number of segments.

2. Then, the perceptual weighting procedure in ANSI S3.5-1997 is used to calculate CSII in each third-octave band. The hearing loss profile is involved in this procedure. The difference of CSII and SII in the standard is that the SNR is replaced by the signal distortion ratio (SDR) (Kates and Arehart, 2005):

$$SDR(j) = \frac{\sum_{k=0}^{K} W_j(k)|\gamma(k)|^2 S_{yy}(k)}{\sum_{k=0}^{K} W_j(k)(1-|\gamma(k)|^2) S_{yy}(k)} \quad (2)$$

where $S_{yy}(k)$ is the estimated power spectral density of the distorted signal and j is the index of third-octave band. W is the simplified ro-ex filter for the band.

- For each third-octave band, one new feature is calculated by combining the weighted CSII for the three energy groups (high energy level, mid-energy level, low energy level) using the following function:

$$\begin{aligned} c(j) &= -3.47 + 1.84 CSII_{Low}(j) \\ &+ 9.99 CSII_{Mid}(j) + 0.0 CSII_{High}(j); \\ New\_feature(j) &= e^{-c(j)} \end{aligned} \quad (3)$$

In total, we have 18 perceptually-dependent features.

2.2 Evaluation of Performance

To test the performance, we used a multi-layer perceptron with 1 hidden layer (14 neurons) and 1 output layer (1

neuron) (Demuth *et al.*, 2008). We used two feature sets: conventional features used in environment classification, which include CGAV, CGFS, Pitchvar, Delpitch, Onstem, Onsetv, Onsetc, Onseth, Beat, Width, Symmetry, Skewness, Kurtosis, Lower Half, MLFS, M1, M2, M3 (Büchler, 2002) and the new features we described in the previous section. The inputs of the neural network were the two feature sets, which were evaluated separately, and the output was the volume gain in dB. Mean square error (MSE) between the volume gain at the output of the ANN and the volume gain that optimizes the speech intelligibility in a simulation of a hearing aid user with profound hearing loss. (This work was done in a simulator to mimic the interactions between the acoustic environment (sound files) and the user behavior (settings) (SPOT, 2009)). One hundred repetitions of the testing and training procedures were used to evaluate the performance.

The dataset for the experiments was generated as follows: We chose 12 30-sec pure speech samples and scaled them to 65 dB SPL. We generated the distorted files by adding white noise to pure speech at different SNR (from -10 to 15dB) and then scaled the files to SPL=65dB. Consequently, we had 312 audio files, from which we randomly chose 200 for training and another 100 for testing. The randomization was done for each of the 100 experiment repetitions.

3. RESULTS

Figure 1 shows the performance of the ANN with the conventional and new feature sets over 100 experiment repetitions. The average MSE obtained from the conventional features is 1.3 (dB based), while that obtained from the features is 1.4 (dB based). With this dataset, we find that the performance of the new features is very similar to the performance of the conventional features.

4. DISCUSSION

Both the new and conventional feature sets performed very well. However, it should be noted that the dataset in this study used only white noise and one SPL level. In the future, we will repeat the experiment using a much more comprehensive dataset that mixes different types of noise, signal, and SNR levels, and with different hearing loss profiles. We expect that the new feature set will perform better in such a challenging situation. One disadvantage of the new feature set is that it requires an estimate of the SNR (or SDR), which may be difficult to obtain at times. We will therefore further investigate features, such as the temporal "modulation level" (Büchler, 2005), that would provide alternative information for the SNR, without the need to estimate it.

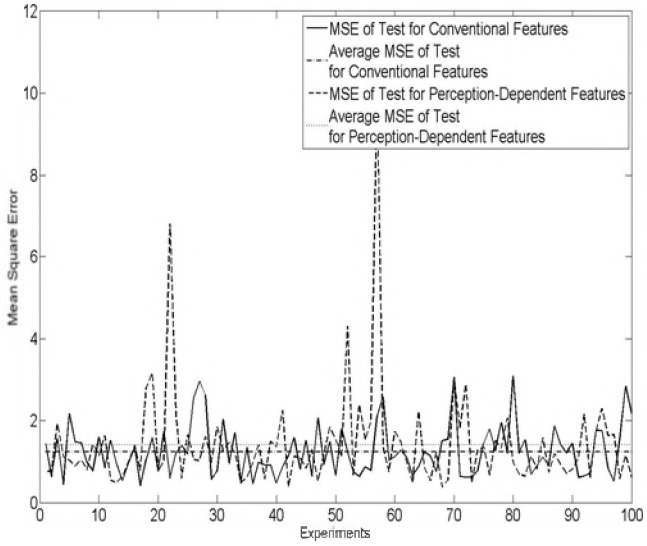

Figure. 1. The performance of the conventional and new features.

REFERENCES

ANSI S3.5-1997, (**1997**), Methods for calculation of the speech intelligibility index, American National Standards Institute, Inc., New York.

Büchler, M.C., (**2002**), Algorithms for Sound Classification in Hearing Instruments, PhD Thesis, Swiss Federal Institute of Technology Zurich, pp. 1-149.

Demuth, H.; Beale M.; and Hagan M., (**2008**), Neural Network Toolbox 6 (User's Guide). The Mathworks.

Kates, J.M.; Arehart, K.H., (**2005**), Coherence and the speech intelligibility index, Acoustical Society of America, vol. 117, pp. 2224-2236.

Nordqvist, P.; Leijon, A. (**2004**), "An efficient robust sound classification algorithm for hearing aids," The Journal of the Acoustical Society of America, vol. 115, pp. 3033-3041,.

SPOT (Signal Processing Oriented Technology) Group, (**2009**), University of Ottawa, Ottawa, Ontario, Canada, Report on Learning Systems, pp. 1-122.

ACKNOWLEDGEMENTS

The authors thank Professor Christian Giguère and Professor Wail Gueaieb for helpful comments and suggestion.