19th INTERNATIONAL CONGRESS ON ACOUSTICS
MADRID, 2-7 SEPTEMBER 2007

OPTIMIZING HEARING AIDS FOR MUSIC LISTENING

PACS: 43.66.Ts

Wessel, David[1]; Fitz, Kelly[2]; Battenberg, Eric[1]; Schmeder, Andrew[1]; Edwards, Brent[2]

[1] Center for New Music and Audio Technologies, Dept. of Music, University of California, Berkeley
Berkeley CA 94720 USA
{wessel,eric,andy}@cnmat.berkeley.edu

[2] Starkey Hearing Research Center, 2150 Shattuck Ave, Suite 408
Berkeley CA 94704 USA
{kelly_fitz,brent_edwards}@starkey.com

ABSTRACT

Modern digital hearing aid technology that is heavily optimized for speech intelligibility often makes music unlistenable, so that hearing aid wearers often prefer to remove their hearing aids when listening to music. The effects of hearing aid processing on musical signals and on the perception of music have received very little attention. There is no standard test of music perception, and to make the problem more difficult, different musical styles thrive in strikingly different acoustical environments. There have been some studies on the effect of reduced bandwidth on the perceived quality of music, but no systematic evaluation of the effects of dynamic range compression, the most ubiquitous form of gain compensation in digital hearing aids.

In this report we present a novel approach to hearing aid fitting applied to both individual differences in hearing impairment and differences among musical styles. The method uses a subjective space approach to reduce the dimensionality of the fitting problem and a non-linear regression technology to interpolate among hearing aid parameter settings. This listener-driven method provides not only a technique for optimal aid fitting, but also information on individual differences and the effects of gain compensation on different musical styles.

INTRODUCTION

Advances in modern digital hearing aid technology focus almost entirely on improving the intelligibility of speech in noisy environments. The effects of hearing aid processing on musical signals and on the perception of music receive very little attention, despite reports that hardness of hearing is the primary impediment to enjoyment of music in older listeners, and that hearing aid processing is frequently so damaging to musical signals that hearing aid wearers often prefer to remove their hearing aids when listening to music.

Though listeners and musicians who suffer hearing impairment are no less interested in music than normal-hearing listeners, there is evidence that the perception of fundamental aspects of (Western) musical signals, such as the relative consonance and dissonance of different musical intervals, is significantly altered by hearing impairment [14]. Measures such as the Articulation Index and the Speech Intelligibility Index [1] can be used to predict intelligibility from the audibility of speech cues across all frequencies, and a variety of objective tests of speech comprehension are used to measure hearing aid efficacy, but there is no standard metric for measuring a patient's perception of music. Moreover, hearing-impaired listeners are less consistent in their judgments about what they hear than are normal-hearing listeners [11], and individual differences in performance among listeners having similar audiometric thresholds make it difficult to predict the perceptual effects of hearing aid processing [4]. These factors, combined with the differences in the acoustical environments in which different styles of music are most often presented, underline the importance of individual preferences in any study of the effects of hearing aid processing on the perception of music.

There have been studies on the effect of reduced bandwidth on the perceived quality of music [8], but no systematic evaluation of the effects of dynamic range compression, the most ubiquitous form of gain compensation in digital hearing aids. This paper presents a subjective, listener-driven method for optimizing gain and compression parameters in a digital hearing aid. A radial basis function network is used as a regression method to interpolate a subspace of parameters. The listener navigates this subspace in real time using a two-dimensional graphical interface and is able to quickly converge on his or her personal optimal parameter set.

SIGNAL PROCESSING IN DIGITAL HEARING AIDS : MULTIBAND COMPRESSION

The primary signal processing operation in modern digital hearing aids is non-linear amplification in the form of multiband compression. Compression, a form of automatic gain control, amplifies low-level (quiet) signals but not high-level (loud) signals, reducing the overall dynamic range of the processed signal, and allowing quiet sounds to be made audible without loud sounds being made uncomfortably loud. Compressive amplification can compensate, in part, for damage to outer hair cells which, in a healthy ear, perform dynamic range compression. For patients suffering the reduced dynamic range that is typical of sensorineural hearing loss, compression can provide audibility and comfort over a wider dynamic range than linear amplification [6].

Compressive amplification is described by the ratio of input level (in decibels) to output level. A compression ratio of 2:1, for example, implies that a change in input level of 2 dB produces a 1 dB change in output level. Very high and very low-level signals typically receive linear amplification (1:1), and signals in between the extremes receive compressive amplification. In addition to the gain that is added to low-level signals, compression circuits are characterized by a pair of time constants that determine how quickly the gain is reduced when a sudden increase in signal level is detected (the *attack* time constant) and how quickly the gain is increased due to a sudden drop in signal level (the *release* time constant).

In *wideband compression*, gain is computed according to the overall signal level, and applied equally across all frequencies. This technique preserves the spectral shape of the processed signal, but it has the disadvantage that the gain computation is dominated by the region of the spectrum having the greatest energy. The presence of a strong, narrowband signal in one frequency region can thereby cause weaker signals at distant frequencies to be rendered inaudible. Moreover, many patients suffer hearing loss that is non-uniform in frequency. Wideband compressors offer no means of providing increased gain in frequency regions of greater hearing loss.

In *multiband compression*, the signal is filtered into several frequency bands, and compression is applied separately to the signal in each band. The bands overlap somewhat, and the width of the bands increases with frequency, reflecting the wider auditory filter bandwidth at higher frequencies. Multiband compression prevents a strong, narrowband signal from determining the gain at distant frequencies, and allows gain and compression to be prescribed differently in each band according to the patient's hearing loss. Typically 8 to 16 bands of compression are used. Compression ratios rarely exceed 2.5:1. Attack time constants are very short, almost instantaneous, to prevent a sudden loud sound being presented with painfully high gain to the hearing aid wearer, and release time constants are typically tens to hundreds of milliseconds. Time constants are often uniform across all bands, but need not be so.

Compressor settings are chosen to maximize speech intelligibility, and it is likely that these settings are not optimal for listening to music [3]. It is important to note that compression is not always detrimental to music signals. Multiband compression is routinely applied in broadcast and in music production. It is only the particular configuration of the compressor in digital hearing aids that is thought to be suboptimal for musical listening.

Additional Processing

An additional non-linear processor, a *peak limiter*, is employed in hearing aids to prevent very loud sounds being processed. A peak limiter is a compressor that provides no gain or attenuation at most input levels, but severe compression at very high signal levels, effectively enforcing a ceiling on the level of signals passing through the circuit, at the expense of significantly distorting signals that exceed that ceiling. Consistent with the bias toward speech intelligibility, the limiting level is typically set at about 85 dB SPL, since even the loudest speech sounds do not exceed this level. But many non-speech sounds, including music, regularly exceed 85 dB SPL, with the result that such intense signals are distorted (clipped) by the limiter. Elevated peak limiting levels have been shown to reduce distortion in hearing aid-processed music, and to improve the perceived quality of the processed music, but this parameter is often fixed for a particular hearing aid, so other workarounds have to be found for wearers of such aids [3].

Hearing aids commonly perform additional signal processing in the form of noise reduction, acoustic feedback suppression, and directional microphone processing. These algorithms may operate under the control of an automatic environment classifier. Hearing aids generally have several program modes, or presets, often including a “music” program that disables most of the processing except for the gain and limiting, because the other algorithms are tailored to speech intelligibility in specific (usually noisy) environments, and often perform poorly in musical listening situations.

Bandwidth Limitations

The bandwidth of most hearing aids is limited to 6 to 8 kHz. The loss of high frequencies may not be a problem for musical listening, because many impaired listeners have such severe losses at those frequencies that improvements afforded by extended bandwidth are inaudible to them. Hearing impaired listeners do not consistently prefer extended high frequency response for listening to music [8, 11].

At low frequencies, hearing aid bandwidth is limited by venting. Hearing aids that completely fill the ear canal usually include a vent, a small hole allowing some sound to pass in and out of the ear canal. This vent is needed to reduce the occlusion effect, the unpleasant sensation of the wearer’s own voice sounding “hollow” or “like talking in a barrel”. An unfortunate side effect of venting is that low frequency energy is dissipated through the vent, and it is not possible to provide enough gain to make up for the lost low frequency energy. For large vents, low frequency response may start to roll off as high as 750 Hz. This low frequency roll-off does not compromise speech intelligibility. In fact, low frequencies are sometimes deliberately attenuated in hearing aid fitting to improve speech clarity. However, the loss of low frequencies has a large, negative impact on the perception of music (middle C on a piano is approximately 262 HZ). Both normal-hearing and hearing-impaired listeners consistently prefer extended low frequency response for listening to music [8].

Though bandwidth limitations seem certain to play a significant role in the perception of hearing aid-processed music, the tradeoffs involved in extending hearing aid bandwidth are well-understood. In this paper, we will focus on the more difficult problem of optimizing the many parameters of a multiband compressor.

DIMENSIONALITY REDUCTION VIA A SUBJECTIVE SPACE APPROACH BASED ON PERCEPTUAL DISSIMILARITY

Characterizing perceptual dissimilarity as distance in a geometric representation has provided auditory researchers with a rich set of robust methods for studying the structure of perceptual attributes [13]. Examples include spaces for vowels and consonants [12], timbres of musical instruments, rhythmic patterns, and musical chords [10]. The most common method for generating a spatial representation is the *multidimensional scaling* (MDS) of pairwise dissimilarity judgments [2]. In this method, subjects typically rate the dissimilarity for all pairs in a set of stimuli. The stimuli are treated as points in a low-dimensional space, often two-dimensional, and the MDS method finds the spatial layout that maximizes the correlation between distances in the representation and subjective dissimilarity ratings among the stimuli. As an alternative to the MDS method we [10] and others [9] have found that directly arranging the stimuli in a subjectively meaningful spatial layout provides representations comparable in quality to MDS.

We have implemented, in the Max/MSP [5] audio programming environment, a user interface that provides for the placement of stimulus objects in a two dimensional space by clicking on an object to listen to it and then dragging it with the mouse to a subjectively appropriate position. With this method we have found that a large number of stimuli can be arranged much more rapidly than by the MDS method with pairwise dissimilarity judgments. With the addition of the interpolation technique described in the next section, this layout method has proven to be effective in providing for low-dimensional control of complex musical structures in real-time applications. We have applied this technique to the problem of adjusting the large number of parameters associated with multiband compressors common in digital hearing aids.

The Importance of Matching Loudness

Loudness plays a strong role in determining sound quality preference [7]. This phenomenon is well known to the loudspeaker salesman who plays the speaker system with which he wishes to impress his customer a bit louder than the alternatives. We found in preliminary experiments with our subjective space navigation system that listeners would tend to prefer compression parameter settings that sounded louder, regardless of the specifics of the parameter settings. In order to minimize, if not eliminate, this loudness preference effect, we require that listeners first match the loudness of each stimulus to a specified target setting by applying gain or attenuation stage via a method of adjustment. Loudness matching of compression settings is not without its difficulties. Compression alters the dynamic envelopes that, depending on the features to which the listener attends, influence the loudness impression. Though we have yet to explore the structure of loudness matching data, it should prove informative regarding compressor behavior and the specifics of a given listener's hearing impairment.

Differences Among Individuals and Musical Styles

Our listener-driven system provides for the personal tuning of parameters as well as the tailoring of parameters to different musical styles. The data sets provided by the method are rich in structure and will hopefully supply insight into individual differences in hearing impairment. The multidimensional scaling research community has provided a variety of methods for characterizing differences among individuals and stimulus conditions.

INTERPOLATION USING A RADIAL BASIS FUNCTION NETWORK

Interpolation within the subspace is performed using a radial basis function network composed of a radial basis hidden layer and a linear output layer as shown in Figure 1. This simple two layer design is very effective in accomplishing our goal of parameter interpolation.

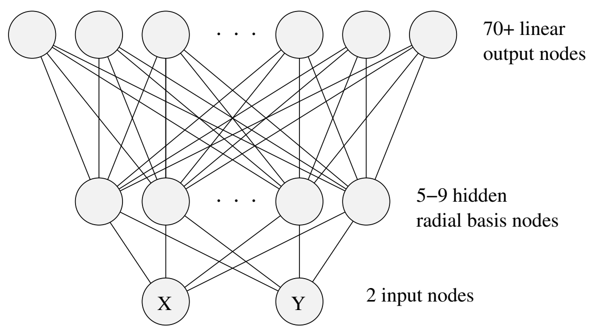

Figure 1: *Neural Network Architecture*

The specifics of the system are shown in Figure 2. To begin, the neural network takes the two dimensional input vector and measures its distance from each of the q preset locations which are stored as the columns of a matrix $\mathbf{L}$. The output of this distance measure is a q-dimensional vector which is then scaled by a constant a and then passed through the Gaussian radial basis function. The constant a affects the spread of the Gaussian function and ultimately controls the smoothness of the interpolation space. The output of the radial basis function is a q- dimensional vector of preset weights. For example, if the input location corresponds to one of the preset locations, then the weight corresponding to that preset would be 1. The radial basis weight vector is now the input to the linear output layer.

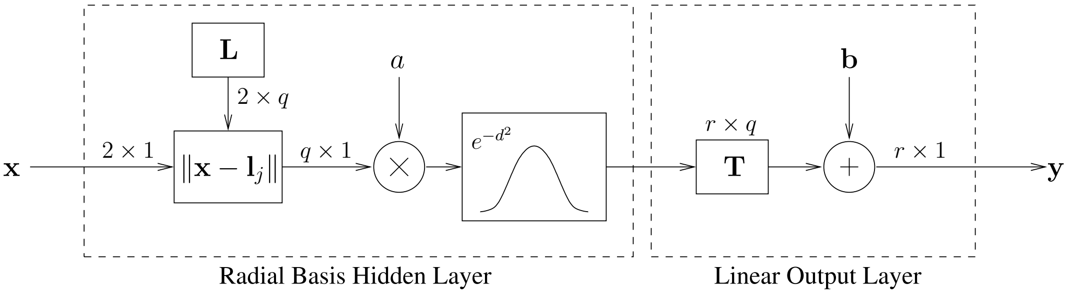

Figure 2: *System Diagram*

The linear layer consists of a mapping from the q-dimensional weight vector to the r-dimensional parameter space. This linear transformation is carried out using a matrix $\mathbf{T}$ that left-multiplies the weight vector and a constant vector $\mathbf{b}$ which is summed with the resulting matrix product. If $\mathbf{w}$ is the weight vector and $\mathbf{y}$ the output vector, we have

$$\mathbf{y} = \mathbf{T}\mathbf{w} + \mathbf{b}. \qquad \text{(Eq. 1)}$$

The training of the network is fairly simple and does not require any complex iterative algorithms. This allows the network to be retrained in real-time, so that the user can instantly experience the effects of moving presets within the space. The network is trained so that each preset location elicits an output equal to the exact parameter set corresponding to that preset.

The values that must be determined by training are the preset location matrix $\mathbf{L}$, the linear transformation matrix $\mathbf{T}$, and the vector $\mathbf{b}$. The matrix $\mathbf{L}$ is trivially constructed by placing each two-dimensional preset location in a separate column of the matrix. The matrix $\mathbf{T}$ and vector $\mathbf{b}$ are chosen so that if the input location lies directly on a preset, then the output will be the parameters corresponding to that preset. To solve for these, we can set up a linear system of equations. We can place $\mathbf{T}$ and $\mathbf{b}$ together in a matrix

$$\mathbf{T}' = [\mathbf{T}|\mathbf{b}]\,. \qquad \text{(Eq. 2)}$$

Then we place the weight vectors corresponding to each preset location into a matrix $\mathbf{W}$ and append a row vector of ones, $\mathbf{1}_{1\times q}$, so that

$$\mathbf{W}' = \left[\frac{\mathbf{W}}{\mathbf{1}_{1\times q}}\right]. \qquad \text{(Eq. 3)}$$

Let the matrix $\mathbf{P}$ be the target matrix composed of columns of the parameters corresponding to each preset. Now our linear system of equations can be represented by the single matrix equation

$$\mathbf{T}'\mathbf{W}' = \mathbf{P} \qquad \text{(Eq. 4)}$$

Because there are more degrees of freedom in the system than constraints, the system is underdetermined and has infinitely many solutions. We choose the solution, $\mathbf{T}'$ with the lowest norm by right multiplying by the pseudo-inverse of $\mathbf{W}'$. The solution with lowest norm was chosen to prevent the system from displaying erratic behavior and to keep any one weight from dominating the output. After we have solved for $\mathbf{T}$ and $\mathbf{b}$, the training is complete. Compared to other neural network training procedures, such as back propagation, this method is extremely fast and still produces the desired results.

CONCLUSION

We have implemented a prototype listener-driven interactive system for adjusting the high dimensional parameter space of multiband compressors for digital hearing aids. The system has three components. The first allows listeners to match the loudness of various parameter settings. The second has listeners lay out a two dimensional space of parameter settings so that the relative distances in the layout correspond to the subjective dissimilarities among the settings. The third performs a nonlinear regression between the coordinates in the subjective space and the underlying parameter settings thus reducing the dimensionality of the parameter adjustment problem. This regression is performed by a radial basis function neural network that trains rapidly with a few matrix operations. The neural network provides for smooth real-time interpolation among the parameter settings.

The system is intuitive for the user. It provides real-time interactivity and affords non-tedious exploration of high dimensional parameter spaces such as those associated with multiband compressors. The system captures rich data structures from its users that can be used for understanding individual differences in hearing impairment as well as the appropriateness of parameter settings to differing musical styles.

References

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

[2] I. Borg, P. J. F. Groenen. *Modern Multidimensional Scaling: Theory and Applications*. Springer, New York, NY (2005)

[3] M. Chasin, F. A. Russo. Hearing Aids and Music. Trends in Amplification **8 (2004), no. 2** 35–47

[4] C. C. Crandell. Individual Differences in Speech Recognition Ability: Implications for Hearing Aid Selection. Ear and Hearing **12 (1991), no. 6 Supplement** 100S–108S

[5] Cycling ’74. Max/MSP. `http://www.cycling74.com/products/max/msp/html`

[6] B. Edwards. Hearing aids and hearing impairment. In S. Greenberg, W. Ainsworth, A. N. Popper, R. R. Fay, eds., *Speech Processing in the Auditory System*. Springer-Verlag, New York, NY (2004) 339 – 421

[7] H. Fastl. Psycho-acoustics and sound quality. In J. Blauert, ed., *Communication Acoustics (Signals and Communication Technology)*. Springer, Berlin, Germany (2005) 139–162

[8] J. R. Franks. Judgments of Hearing Aid Processed Music. Ear and Hearing **3 (1982), no. 1** 18–23

[9] R. L. Goldstone. An efficient method for obtaining similarity data. Behavior Research Methods, Instruments, & Computers **26 (1994), no. 4** 381–386

[10] A. Momeni, D. Wessel. Characterizing and controlling musical material intuitively with geometric models. In *Proceedings of the 2003 Conference on New Interfaces for Musical Expression*. Montreal, Canada (2003) 54–62

[11] J. L. Punch. Quality judgments of hearing aid-processed speech and music by normal and otopathologic listeners. Journal of the American Audiology Society **3 (1978), no. 4** 179–188

[12] R. N. Shepard. Psychological Representation of Speech Sounds. In E. David, P. B. Denes, eds., *Human Communication a Unified View*. McGraw-Hill, New York, NY (1972) 67–113

[13] R. N. Shepard. Multidimensional Scaling, Tree-Fitting, and Clustering. Science **210 (1980), no. 4468** 390 – 398

[14] J. B. Tufts, M. R. Molis, M. R. Leek. Perception of dissonance by people with normal hearing and sensorineural hearing loss. Acoustical Society of America Journal **118 (2005)** 955–967