# AUDITORY MODELS FOR EVALUATING ALGORITHMS

A Thesis
Presented to
The Academic Faculty

by

Abigail A. Kressner

In Partial Fulfillment
of the Requirements for the Degree
Master's degree in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
August 2011

# AUDITORY MODELS FOR EVALUATING ALGORITHMS

Approved by:

Professor Christopher J. Rozell, Advisor
School of Electrical and Computer Engineering
*Georgia Institute of Technology*

Professor David V. Anderson
School of Electrical and Computer Engineering
*Georgia Institute of Technology*

Professor Mark A. Clements
School of Electrical and Computer Engineering
*Georgia Institute of Technology*

Date Approved: 24 June 2011

# ACKNOWLEDGMENTS

I would like to thank Dr. Christopher Rozell for his role as an advisor and mentor. He always provides useful wisdom, experience, and patience, and his leadership is an example from which I hope to follow one day. Furthermore, I would like to thank Dr. David Anderson and Dr. Mark Clements for serving on my thesis committee and for their continued support. I wish to thank James Kates for sharing his work on HASQI and for providing helpful comments on parts of this text, Yi Hu for sharing his dataset, Philipos Loizou for his valuable textbook and corresponding MATLAB functions, and Vivienne Ming for sharing her dataset. Finally, I am eternally grateful to my family and friends for providing love, encouragement, and support.

# TABLE OF CONTENTS

# LIST OF FIGURES

# SUMMARY

Hearing aids are tasked with the undesirable job of compensating an impaired, highly-nonlinear auditory system. Historically, these devices have either employed linear processing or relatively unsophisticated, nonlinear processing techniques. With increasingly more accurate models of the auditory system, expanding computational power, and many more objective measures which utilize these models, we are at a turning point in hearing aid design.

Although subjective listener tests are often the most accepted methods for evaluating the quality and intelligibility of speech, they inherently treat the auditory system as a "black box." Conversely, model-based objective measures typically treat the auditory system as a cascade of physical processes. As a result, objective measures have the potential to provide more detailed information about how sound is processed and about where and why quality or intelligibility breaks down. Provided that we can generalize model-based objective measures, we can use the measures as tools for understanding how to best process degraded signals, and therefore, how to best design hearing aids.

However, generalizability is a key requirement. Since many of the well-known objective measures have been developed for normal-hearing listeners in the context of audio codecs, we are unsure about the generalizability of these measures to predicting quality and intelligibility for hearing-impaired listeners with "unknown" datasets (i.e. a set on which it was not trained) and distortions which are specific to hearing aids. Relatively recently, however, Kates and Arehart (Journal of the Audio Engineering Society, 2010) proposed the Hearing Aid Speech Quality Index (HASQI), which is a

model-based objective measure that predicts quality for normal-hearing and hearing-impaired listeners by taking into account many of the distortions which hearing aids introduce. HASQI solves many of our concerns of generalizability for predicting quality, but it still remains to test HASQI's ability to predict quality with datasets on which it was not trained. Thus, we explore the robustness of HASQI by testing its ability to predict quality for "unknown" de-noised speech, and we directly compare its performance to some other metrics in the literature.

# CHAPTER I

# INTRODUCTION

A simplified product development process for amplification devices starts with algorithm development and is followed by normal-hearing (NH) listener testing. Some number of iterations later, developers subsequently conduct hearing-impaired (HI) listener testing, and some number of iterations after that, algorithms are implemented in products. This process is flawed in several ways. First, HI listener testing can be costly and time-consuming, particularly for developers lacking direct access to testing facilities and subject pools. Second, the relationship between NH and HI listener preferences is unclear so initial testing with NH listeners may actually be prohibitive or misleading. Third, the feedback from listener testing is typically qualitative in nature. Put another way, the feedback does not tell developers directly where or why distortion is occurring, nor how to fix it.

Addressing the first two issues, developers have turned to objective measures of intelligibility and quality in place of expensive and time-consuming subjective measures. For many developers, clock cycles are faster and more economical than human ears. Furthermore, objective measures have the potential to provide quantitative feedback rather than just qualitative feedback. When using subjective listener testing, the auditory system is inherently treated as a "black box." Conversely, objective measures can be designed to treat the system as a cascade of physical processes. In this way, developers can look into the system to understand how sound is processed and to figure out exactly where and why intelligibility or quality succeeds or fails.

Section 1.1 provides a very basic primer to the auditory system, and Section 1.2 briefly describes methods for measuring hearing by discussing both subjective and objective types of measures. To conclude the chapter, Section 1.3 introduces the notion of the generalizability of objective measures while also presenting the contents of the remaining chapters.

## 1.1 Auditory system

The auditory system is often described as being made up of three compartments (Figure 1): the outer, middle, and inner ear. The outer ear consists of the pinna and external canal. Sound is reflected by the pinna into the canal and then travels through the canal to the tympanic membrane. The tympanic membrane, which commences the middle ear cavity, oscillates with the pressure waves from sound in the canal. Then, the oscillations are transferred to the ossicular chain, which is a chain of three small bones named the malleus, incus, and stapes.

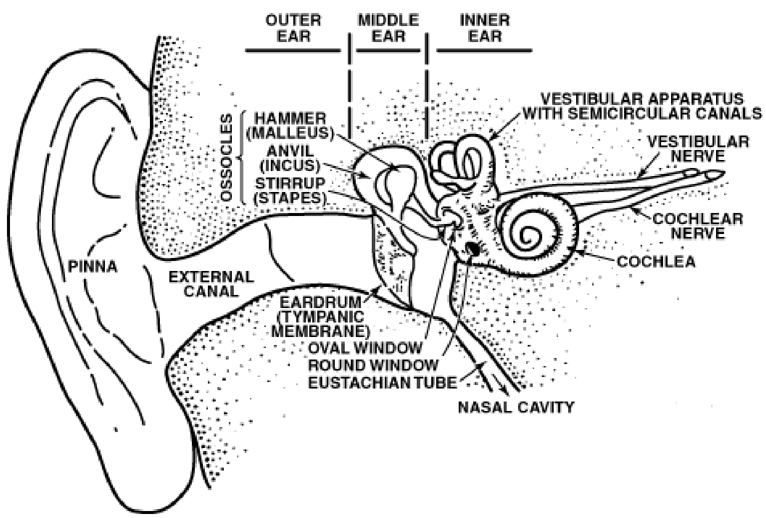


Figure 1: Schematic of the auditory system (not drawn to scale). From *J. L. Flanagan, Speech Analysis and Perception, Springer-Verlag, Berlin, 2nd edition, 1965.*

Oscillations in the stapes are transferred to the inner ear via the oval window. Movement of the fluid within the cochlea causes the basilar membrane to vibrate. Inner hair cells (IHCs) located along the basilar membrane, which are mechanosensors tuned to specific frequencies, produce action potentials when the local basilar membrane displacement is large enough. Then, afferent neurons innervating the IHCs transport the action potentials to the brain. Finally, feedback traveling along efferent neurons from the brain innervate the outer hair cells (OHCs), which essentially serve as pre-amplifiers and ultimately increase the dynamic range of hearing.

## *1.2 Measures of hearing*

Typically, hearing is tested clinically through the measurement of auditory thresholds. Listeners with thresholds in the range of 0 to 20 dB SPL are classified as having normal hearing. Conversely, listeners with thresholds that fall outside of this range are classified as having impaired hearing. Although threshold testing does provide some information, it is quite limited. To start, air-conducted thresholds alone cannot identify the source of hearing loss. Hearing loss caused by the blockage of conduction through the middle ear (often called conductive hearing loss) is drastically different than hearing loss caused by sensory or neural failure (often identified collectively as sensorineural hearing loss). Moreover, sensorineural hearing loss caused by damage to IHCs is different than hearing loss caused by damage to OHCs, and even more so, different than hearing loss caused by damage to the auditory pathway. Each of these types of hearing losses require different types of treatment.

Regardless of the source of the hearing loss though, the goal is usually to restore hearing back to "normal." One way of measuring the performance of the compensation is to measure aided-thresholds, which are the auditory thresholds obtained with the assistance of a communication device. However, other common methods for evaluating the performance of compensation techniques include looking at quality and intelligibility.

Quality and intelligibility are used in the evaluation of sound in many acoustic applications, both before and after processing. Currently, the most accurate method for measuring quality and intelligibility is through subjective listener tests. In the case of quality, this testing involves listeners rating on some pre-defined scale various aspects of the sound quality including, for example, signal distortion, background intrusiveness, or overall quality. In the case of intelligibility, this testing often involves scoring listeners' ability to identify words, sentences, or syllables.

Given that subjective listener tests are impractical in many situations and for many researchers, objective measures of quality and intelligibility are often used instead to predict subjective ratings. Considering the potential benefit of having good predictors of quality and intelligibility, it is no surprise that the number of measures which have been proposed is very extensive. Chapter 2 reviews many of the major quality and intelligibility objective measures.

## *1.3 Generalizability of objective measures*

A large number of objective measures have been developed in the context of audio codecs and communication channels, while others in the context of hearing aids or cochlear implants. Most are developed for NH listeners, while only a portion are developed for HI listeners. Moreover, some are developed from a single dataset, while others are more rigorously developed and tested for robustness across data. And still, some measures are purely phenomenological, while others are model-based.

Since many of the objective measures have been developed in the context of speech codecs and communication channels for NH listeners, there is often concern about the generalizability of these metrics to predict quality and intelligibility, first, for HI listeners, and second, for distortions which are introduced by hearing aid algorithms. Moreover, there is concern about the generalization of using objective measures for datasets and listener studies which are "unknown" to the measures. A metric which was proposed relatively recently, the Hearing Aid Speech Quality Index (HASQI), is a model-based objective measure of quality developed in the context of hearing aids for NH and HI listeners [35]. HASQI solves the first two generalization issues for those interested in using objective measures for the design and evaluation of hearing aid algorithms. However, HASQI has not been tested thus far on any datasets other than the one on which it was trained. Chapter 3 presents an exploration of the robustness of HASQI in predicting subjective quality. We use an “unknown” dataset of noisy speech processed by noise suppression algorithms to compare the performance of HASQI to the performance of several other objective measures. In closing, Chapter 4 presents conclusions and discusses future work.

# CHAPTER II

# OBJECTIVE MEASURES

Currently, the most accurate evaluation of speech quality and intelligibility is through subjective listener tests. However, listener tests can be time-consuming and expensive. Given the large number of people interested in measuring quality and intelligibility that do not have access to listener testing on a regular basis, it is no surprise that so much work has gone into developing objective measures. To give an impression of the vastness of the measures, we describe a selection of them.

Some of the earliest objective measures of quality were relatively simple measures which quantified the difference between a degraded signal and its corresponding clean version. Because of the relative lack of predict power with these simple measures, developers shifted to more of a model-based approach. Through the years, many new measures have been proposed to improve upon the ones before. Some of theses model-based measures are developed from scratch, but many are just extensions and combinations of earlier versions of measures.

## *2.1 Quality*

Quackenbush, Barnwell, and Clements [49] reviewed many of the early, simple objective measures of quality. Most of these measures were developed in the context of speech coding algorithms for NH listeners. Hu and Loizou re-reviewed some of these basic measures more recently [28]. Starting with an obvious choice, a metric based on the overall signal-to-noise ratio (SNR) was proposed, but after a few studies showed overall SNR to lack predictive power, frame-based segmental SNR (segSNR) was proposed in 1976 to take its place [46, 19]. Furthermore, frequency-weighted segmental SNR (fwsegSNR) was proposed as an extension to segmental SNR, where the SNR in each frame is measured as the weighted average of the power in the signal and noise [55]. In 1982, Klatt [36] proposed a weighted-slope spectral (WSS) distance, which is a metric based on spectral slope differences near the peaks in the critical-band spectra of a degraded signal and its reference signal. Other measures include the log-likelihood ratio (LLR) measure and the Itakura-Saito (IS) distortion measure, which both use linear prediction coefficients to predict quality [19, 28], and the cepstral distance measure, which uses cepstral coefficients [28]. Additionally, Wang [57] proposed a measure called Bark spectral distortion (BSD) which is the average squared Euclidean distance between spectral vectors of the original and coded utterances.

These basic measures predict quality fairly well, but in the search for better performance, many objectives have taken more of a modeling approach. In 1985, Karjalainen [32] developed a psychoacoustic model to estimate loudness, which in turn, could be used to predict quality. Kates [33] proposed a "central spectrum" model which incorporated a critical-band filterbank, steady-state representations of the averaged and synchronized firing rates, and temporal integration to predict the perception of coloration in filtered Gaussian noise. Beerends and Stemerdink [3] developed a metric they called the Perceptual Audio Quality Measure (PAQM) which used a model of the psychoacoustic internal representation of sound based on time-frequency spreading and level compression to determine subjective quality. Furthermore, Beerends and Stemerdink also developed the Perceptual Speech-Quality Measure (PSQM), which optimized level compression for music and speech coding [6]. PSQM was adopted in 1996 as the International Telecommunication Union's Recommendation ITU-T P.861. In 1998, ITU added an appendix to P.861 which included an alternate system based on Measuring Normalizing Blocks (MNB) [56].

Beerends and Stemerdink's PAQM model [3] was combined with several other models to form the Perceptual Evaluation of Audio Quality (PEAQ) method. PEAQ predicts wide-band audio by processing sound through an auditory filterbank and time and frequency masking. It then extracts features based on envelope modulation, the loudness of the noise, differences in the loudness, the noise-to-masker ratio, and the harmonic structure of the difference between the ref and degraded signals. These features are put into a neural network, which is trained to reproduce the listener quality judgements [54, 11]. PEAQ became ITU-R recommendation BS.1387 in 1998, which has mainly been used in the prediction of low-bit-rate coded audio signals.

Also in 1996, Dau, Püschel, and Kohlrausch were developing another family of models. They proposed a psychoacoustic model which produces an internal representation of the information available to higher neural stages of perception, and they proposed using this model to predict quality [14]. Hansen and Kollmeier use the same psychoacoustic model to compute the quality measure, $q_C$, which is the cross correlation of the envelopes of the internal representations [20, 22]. Dau et al.'s metric and $q_C$ differ only in the choice of template and normalization [22, 14]. Hansen and Kollmeier also developed a measure for the continuous prediction of quality [21, 22]. In 2000, Hansen and Kollmeier introduced a new objective measure for the transmission quality of low-bit-rate speech coding algorithms [23] to potentially replace PEAQ.

In 2000, Rix and Hollier extended Wang's BSD [57] to create the Perceptual Analysis Measurement System (PAMS) [50], the first to focus on end-to-end behavior. However, this focus on end-to-end behavior made the performance of existing measures seem poor. In response, PAMS and an updated and extended version of PSQM were combined to produce the Perceptual Evaluation of Speech Quality (PESQ) [2, 51, 5], which superseded ITU's recommendation of PSQM (ITU-T P.862). PESQ is the objective measure recommended by ITU-T for speech quality assessment of narrow-band handset telephony and narrow-band speech codecs. The basic components of PESQ include time alignment, a psychoacoustic model which maps signals into perceived loudness, disturbance processing, a cognitive model, aggregation of the disturbance in frequency and time, and finally, a mapping to the predicted subjective score [51]. In 2008, Beerends et al. [4] extended PESQ for HI listeners by adjusting for hearing loss and level variations.

Despite the relative success of PESQ, the search continues for a better, more intuitive, predictive measure, and many developers continue to propose new models or extensions of previous models. Moore and Tan propose using magnitude differences and spectral slope differences between the excitation patterns of degraded signals and their reference signals to predict quality for speech and music which has been subjected to bandwidth limitations, spectral slope, and spectral ripple [41]. Huber and Kollmeier propose an extension of $q_C$ [20] (called PEMO-Q), which is based on Dau et al.'s psychoacoustically validated, quantitative model of the "effective" peripheral auditory [14, 15]. PEMO-Q computes a Perceptual Similarity Measure (PSM) as the weighted average of the linear cross correlation coefficient across modulation channels, as well as an instantaneous version of PSM ($\mathrm{PSM}_t$) by computing PSM every 10ms. PEMO-Q is reported to be able to predict quality for narrow-band speech as well as wide-band audio signals [29]. Chen and Parsa [12] suggest an objective measure based on Moore and Glasberg's loudness model [44] and Bayesian modeling. In addition to these families of models, Kates and Arehart's group published quite a few papers which propose using coherence-based predictors of quality [1, 34, 37], whereas Tan and Moore's group were proposing normalized cross-correlation measures [52, 53]. Furthermore, Huo et al. focus on predicting, as they claim, the quality-relevant dimension "directness" and "frequency content" [30].

Not surprisingly, another extension of PEMO-Q, named PEMO-AQ, is reportedly being developed by Huber [9]. Concurrently, Beerends is developing the Perceptual Hearing Aid Quality Measure (PHAQM) based off of PAQM, PESQ, and PEAQ, and Haubold and Schmafußare developing the Model of auditory Comfort for Hearing Impaired persons—Revised version (MCHI-R) [9].

Two relatively recent measures specifically look at predicting quality in the context of hearing aid distortion. First, Parsa and Jamieson propose a measure based on an "auditory distance" parameter, which computes the distance between the hearing aid response and their model output [47]. Second, Kates and Arehart propose the Hearing Aid Speech Quality Index (HASQI) which is made up of two components. The first component captures the effects of noise and nonlinear distortion, and the second component, a modification of Moore and Tan's quality measure [41], captures the effects of linear filtering [35].

Kates and Arehart considered additive, multiplicative, and $L_\infty$ combinations of the two components of HASQI and found the multiplicative combination to be slightly more accurate [35]. Several other noteworthy measures looked at the method of combining components together. In contrast to Kates and Arehart, Moore et al. [42] found an additive combination to be more effective than a multiplicative combination. Beerends et al., on the other hand, propose $L_p$ averaging of their three components, with $p = 0.1$ [5]. Further yet, many other metrics have components which are combined using a neural network (e.g. [8, 11, 45]).

The plethora of measures presented here all require both the degraded signal and its corresponding clean version, and therefore, are classified as intrusive measures. The counterpart to this class is the non-intrusive class of measures. These measures do not require a reference signal at all. Since non-intrusive measures are primarily used in the framework of telecommunications, they are left out of the discussion here.

## *2.2 Intelligibility*

Objective measures of intelligibility have similarly been the object of much research. The Articulation Index (AI) and the Speech Transmission Index (STI) are the most widely used predictors of speech intelligibility. AI was first proposed by French and Steinberg [18]. At its core, AI computes a weighted average of the SNR within some number of frequency bands. Kryter validated many of the underlying principles of AI [39] and proposed methods for its use in practical settings [38]. AI became the ANSI standard in 1969 (ANSI S3.5-1969). Although many of the correction factors Kryter introduced increased AI's performance in predicting the effects of modulated noise, peak clipping, and reverberation, AI is known to perform poorly in the presence of some time-domain distortions [7].

Consequently, Houtgast and Steeneken proposed the Modulation Transfer Function (MTF) in 1973 to measure the loss of intelligibility due to echo and reverberation [26]. Steeneken and Houtgast later extended this approach to include the loss of intelligibility due to nonlinear distortions and branded the predictor as the Speech Transmission Index (STI). STI was adopted as an ANSI standard in 1997 as the Speech Intelligibility Index (SII), which is a revision of the AI standard (ANSI S3.5-1997) [48].

However, STI still fails to accurately account for all distortions. For example, it does not account for intraband masker nonlinearities or phase [7]. Based on the premise that the best physical variable we have of the perceptual variable of intelligibility is internal representations in the auditory pathway, many of the new objective measures are model-based predictors. In 1999, Chi et al. [13] first proposed the Spectro-temporal Modulation Index (STMI), which predicts intelligibility by employing a computational model of auditory processing. The model is composed of two basic stages: the "early auditory system" and the "central auditory system." The "early auditory system" models the periphery and computes an auditory spectrogram. The "central auditory system" models higher, more central auditory stages, primarily the primary auditory cortex, to estimate the spectral and temporal modulation content of the auditory spectrogram. Elhilali et al. has shown STMI to be a flexible metric which can be used in many different situations; furthermore, they have shown that STMI correlates well with subjective scores [17, 16].

Zilany and Bruce propose a modified version of STMI which uses their more physiologically accurate model of the auditory periphery [60] in place of Elhilali, Chi, and Shamma's simple periphery model in order to increase the accuracy of STMI. They show the modified STMI to be able to predict qualitatively the effects of presentation level and cochlear impairment on intelligibility [58]. More recently, Ibrahim and Bruce use the modified STMI to explore the interaction of envelope and temporal fine structure cues in speech perception [31].

Another proposed objective measure of intelligibility which uses the Carney and Bruce family of models [10, 60, 61, 59] is the Neural Articulation Index (NAI) [7]. NAI formulates a prediction of intelligibility in a manner which is similar to STI, but which replaces SNR with an estimate of neural distortion. Bondy et al. estimate the neural distortion by computing the distance between the auditory nerve response of a degraded signal and the auditory nerve response of its clean reference signal. A third, more recently developed approach by Hines and Harte, uses the Carney and Bruce family of models to compute neurograms for a degraded signal and its reference, where the "neurogram" is a representation of the discharge rates at the auditory nerve across all time and frequency [58, 25]. Hines and Harte employ the Mean Structured Similarity Index Measure (MSSIM) from the image processing community to compare neurograms and, in turn, predict intelligibility.

# CHAPTER III

# GENERALIZABILITY OF HASQI

Given the laundry list of objective measures presented in Chapter 2, it seems overwhelming to choose which measure to use when. Although many consider intelligibility to be the more important problem, listeners will refuse to use devices which yield bad quality even if the intelligibility improves. Thus, we need to be able to predict quality as accurately as possible if we are going to make any progress using objective measures to evaluate algorithms for the design of hearing aids.

Focusing on quality then, first we need an objective measure which is accessible. Second, we need to be able to predict quality for HI listeners. Third, we need an objective measure which can predict quality for the types of distortion which are introduced specifically by hearing aid algorithms. And finally, we need an objective measure which provides a good prediction of quality for datasets on which it was not trained. Many of the objective measures mentioned in Chapter 2 fail the first criteria since they are not available for us to use. Of the measures that are available to us, few actually predict quality for HI listeners and even fewer focus on distortions produced by hearing aids specifically.

Recently however, Kates and Arehart [35] developed an objective measure for evaluating distortions introduced specifically by hearing aids for both NH and HI listeners. Their metric, called the Hearing Aid Speech Quality Index (HASQI), aims to capture the aspects of quality deemed important for rating speech processed by hearing aids [1]. They report very high correlation, Pearson correlations of $r_p = 0.942$ and $r_p = 0.978$, between HASQI and subjective listener scores for NH and HI listeners, respectively [35]. Given such high correlations, HASQI is a promising candidate as an evaluation and design tool for hearing aid algorithms.

What remains then, is to address whether or not HASQI generalizes across studies. The objective is to explore how HASQI performs with an "unknown" set of speech samples and subjective ratings and to compare HASQI's performance directly with some other metrics in the literature, including segmental signal-to-noise ratio (segSNR), frequency-weighted segmental signal-to-noise ratio (fwsegSNR), weighted-slope spectral distance (WSS), Perceptual Evaluation of Speech Quality (PESQ), log-likelihood ratio (LLR), Itakura-Saito distance measure (IS), and a cepstral distance measure (CEP).

## *3.1 Methods*

Hu and Loizou evaluated an extensive collection of common objective measures of quality for NH listeners by comparing listener and predicted scores of the quality of noisy speech enhanced by noise suppression algorithms [28]. We use the same set of speech files, the same subjective scores, and the same analysis technique.

The speech samples were created using noisy sentences from the noisy speech corpus NOIZEUS [1] and 13 different noise suppression algorithms, including spectral subtractive, subspace, statistical-model-based, and Wiener-filtering type algorithms [28, 40, 27]. With this wide range of algorithms, the dataset contains a wide range of the distortions which are likely introduced during speech enhancement. The final set of 1792 files includes 16 different sentences (sp01-sp04, sp06-sp09, sp11-sp14, sp16-sp19), 14 algorithms (13 noise suppression algorithms plus the control case), four noise types (babble, car, street, and train), and two signal-to-noise ratios (5 dB SNR and 10 dB SNR).

The listener testing was conducted on the speech files by Dynastat, Inc. (Austin, TX) according to the ITU-T P.835 standard. A total of 32 NH subjects were recruited

[1]Available online: `http://www.utdallas.edu/~loizou/speech/noizeus/`

for the listening tests. Subjects were asked to focus on and to rate the speech files sequentially based on signal distortion, background intrusiveness, and overall quality [27, 28]. However, we focus only on overall quality. We compute an average subjective score for each combination of algorithm, noise type, and SNR by averaging across subjects and sentences. This averaging creates a total of 112 cases (13 noise suppression algorithms plus the control, four noise types, and two SNRs).

For all 1792 speech files, we predict quality scores with each objective measure (see Section 3.2). Then for each of the measures, we average across sentences to obtain average objective scores for each of the 112 cases.

We use three measures to evaluate performance. The first, Pearson's correlation coefficient, measures the linear dependence between the objective measures, $o$, and the subjective quality ratings, $s$, as

$$r_p = \frac{\sum_i (o_i - \bar{o})(s_i - \bar{s})}{\sqrt{\sum_i (o_i - \bar{o})^2)}\sqrt{\sum_i (s_i - \bar{s})^2)}}, \tag{1}$$

where $\bar{o}$ is the sample mean of $o$, and $\bar{s}$ is the sample mean of $s$. The second measure, an estimate of the standard deviation of the error when the objective measure is used in place of the subjective measure, is computed by

$$\hat{\sigma}_e = \hat{\sigma}_s \sqrt{1 - r_p^2}, \tag{2}$$

where $\hat{\sigma}_s$ is the standard deviation of the sample of the subjective scores. Since $\hat{\sigma}_e$ depends only on $r_p$ and $\hat{\sigma}_s$, we are not actually reporting any new information by including this measure. However, we are presenting the relationship between the objective and subjective scores from a different perspective.

The third measure is Spearman's rank correlation coefficient, $r_s$, which quantifies how well the relationship between two variables can be described by a monotonic function. The Spearman correlation coefficient is defined as the Pearson correlation coefficient between the ranked variables. In other words, the raw scores are converted to ranks in ascending or descending order and Spearman's rank correlation coefficient is computed from these ranked scores using Equation 1.

## *3.2 Objective measures*

### 3.2.1 Hearing Aid Quality Index (HASQI)

HASQI is the product of two independent indices. The first index, $Q_{\text{nonlin}}$, captures the effects of noise and nonlinear distortion by targeting short-time signal envelope behavior, and the second index, $Q_{\text{lin}}$, captures the effects of linear filtering and spectral changes by targeting differences in the long-term average spectra. Both indices are computed on outputs of an auditory model, and since the model can represent a normal or impaired periphery, the combined quality index serves as an objective measure for both NH and HI listeners.

#### *3.2.1.1 Auditory model*

Kates and Arehart use relatively simple models of the auditory periphery for the nonlinear and linear indices (Figures 2 and 3). In general, each model maps a sound waveform to steady-state neural firing rate. The basic components of each of the models are a middle ear filter, two parallel filterbank paths, compression, and a logarithmic conversion. In the figures, the dashed boxes indicate the components which are configured according to the health of the IHCs and OHCs and the thresholds of hearing for the periphery the model is attempting to represent. In this way, the model can represent a large number of hearing configurations.

The path through the first filterbank of each model, called the analysis filterbank, makes up the main pathway, and the path through the second filterbank, called the control filterbank, acts as a controller to the compression rule. The bandwidths of the analysis filters are inversely proportional to the condition of the outer hair cells (OHCs)—the more significant the hearing loss, the wider the bandwidths. Conversely,

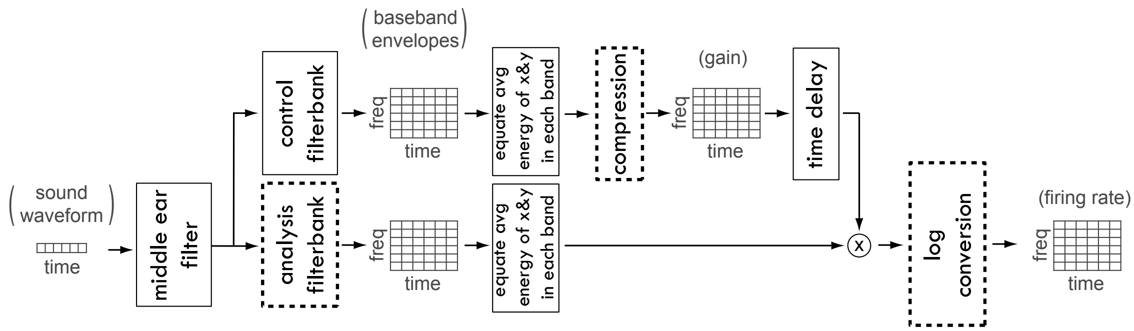


Figure 2: Schematic diagram of the auditory model for the computation of $Q_{\mathrm{nonlin}}$. The dashed boxes indicate those components which are configured for different types of hearing.

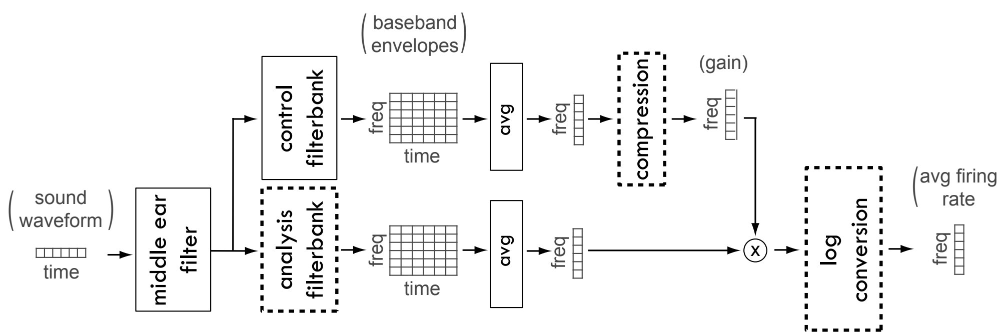


Figure 3: Schematic diagram of the auditory model for the computation of $Q_{\mathrm{lin}}$. The dashed boxes indicate those components which are configured for different types of hearing.

the bandwidths of the control filters are constant and are set to the same bandwidths as those of the analysis filters at maximum hearing loss.

The baseband envelopes at the output of the control filterbank become the input to the compression, which is characterized by a typical compression rule. OHC damage modifies the compression rule so as to shift auditory thresholds and reduce the compression ratio. Cochlear gain at the output of compression is applied, and then damage to the inner hair cells (IHCs) is incorporated via signal attenuation in each frequency band. Finally, the signal is logarithmically scaled to approximate the conversion from signal intensity to steady-state neural firing rate. Kates and Arehart originally claimed in [35] that the log conversion is an approximation of the conversion from signal intensity to loudness, but a better statement is that the log conversion is an approximate mapping of intensity to steady-state neural firing rate (J. Kates, personal communication, 17 May 2011).

The main difference between the two models is in their treatment of time. In the case of the nonlinear index, the output is the time-frequency representation of steady-state firing rate such that the model captures changes in the signal over time. Conversely, in the case of the linear index, the time-frequency representations are averaged across time so that the output is an estimate of the long-term average spectrum. For a more detailed description of the auditory models use, see Kates and Arehart [35].

#### *3.2.1.2 Nonlinear Index*

To start, the reference signal, $x(t)$, and the signal under test, $y(t)$, are put through the auditory model. Each of the resulting time-frequency representations are windowed across time with an 8-ms raised-cosine (von Hann) window and 50% overlap. Each time frame then contains a short-time log-magnitude spectra on an auditory frequency scale. The inverse Fourier transform of the short-time log-magnitude spectra gives something similar to the mel cepstrum. To increase efficiency, the cepstral coefficients are computed with a cosine transform rather than the inverse Fourier transform [24]. Specifically, Kates and Arehart [35] compute the coefficients with a set of half-cosine basis functions, $\phi_j(k)$, where $j$ is the basis function number (or the $j^{th}$ "quefrency" band) from 0 to $J$-1 ($J = 6$ for Kates and Arehart) and $k$ is the gammatone filter index from 0 to $K$-1 ($K = 32$ for Kates and Arehart). To be clear, for a signal with $N$ total time frames after windowing, the $K$ by $N$ spectra, $X$ and $Y$, are transformed to $J$ by $N$ "cepstrograms", $C_x$ and $C_y$. Let $\Phi$ be the set of $\phi_j(k)$ in columns such that $\Phi$ is $K$ by $J$. Then,

$$C_x = \Phi^T X \tag{3}$$

$$C_y = \Phi^T Y. \tag{4}$$

Let $c_{x,j}(n)$ and $c_{y,j}(n)$ be the $j^{th}$ cepstral coefficient (the $j^{th}$ rows of $C_x$ and $C_y$) for all $N$ time frames. Then, $\hat{c}_{x,j}(n)$ and $\hat{c}_{y,j}(n)$ are $c_{x,j}(n)$ and $c_{y,j}(n)$ with the speech pauses and means removed. The normalized cross-correlation, $r(j)$, of $\hat{c}_{x,j}(n)$ and $\hat{c}_{y,j}(n)$ is computed for $j = 2$ through $j = J$.

Finally, the average cepstral correlation is computed as the average of the $2^{nd}$ to the $J^{th}$ correlation. The nonlinear index is a second-order regression fit on the average correlation, where different fits are reported in [35] for NH and HI listeners. Figure 4 provides a schematic diagram for computing $Q_{\text{nonlin}}$.

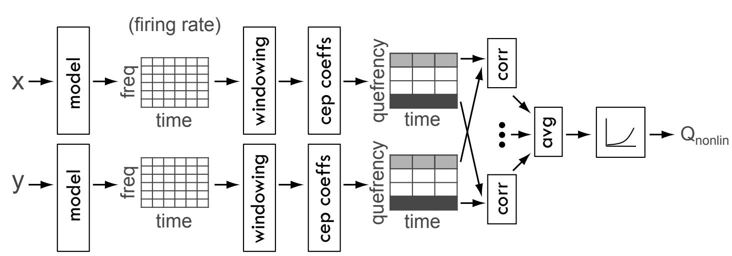


Figure 4: Schematic diagram for computing $Q_{\text{nonlin}}$.

Conceptually, $Q_{\text{nonlin}}$ can be thought of in two ways. First, each cepstral coefficient can be thought to describe the dynamics of a short-time spectrum through time. For example, the second cepstral coefficient is a half-cosine, and therefore, captures spectral tilt as a function of time. Furthermore, the correlation between each of the cepstral coefficients of the signal under test and the reference measures the degree to which the processing altered the dynamics of the short-term spectrum over time (J. Kates, personal communication, 11 May 2011).

The second way of viewing $Q_{\text{nonlin}}$ is to think of the cepstrograms as just efficient representations of the overall spectral shapes. By taking the average correlation then,

$Q_{\text{nonlin}}$ is just measuring how well the two cepstrograms match.

#### *3.2.1.3 Linear Index*

The linear index, $Q_{\text{lin}}$, is based on Moore and Tan's sound quality metric which focuses on predicting quality due to distortions from spectral modifications [41]. The Moore and Tan metric predicts quality based on the differences in excitation patterns and the differences in the slopes of the excitation patterns. Kates and Arehart take a similar approach, but replace the excitation pattern with the output of their auditory model.

The output of the auditory model for the linear index is an estimate of the long-term average spectra, $\bar{X}(k)$ and $\bar{Y}(k)$, of $x(t)$ and $y(t)$. Let $\hat{X}(k)$ and $\hat{Y}(k)$ be the normalized versions of $\bar{X}(k)$ and $\bar{Y}(k)$ (normalized to each have an RMS of one). If we define $d_1(k)$ as the difference in the spectra, we can estimate it as

$$d_1(k) = |\hat{Y}(k)| - |\hat{X}(k)|, \quad 0 \leq k \leq K-1, \tag{5}$$

Furthermore, if we define $d_2(k)$ as the difference in the spectral slopes, we can estimate it as

$$\begin{aligned} d_2(k) = & \left(|\hat{Y}(k)| - |\hat{Y}(k-1)|\right) - \ldots \\ & \left(|\hat{X}(k)| - |\hat{X}(k-1)|\right), \quad 1 \leq k \leq K-1. \end{aligned} \tag{6}$$

The standard deviation of both $d_1(k)$ and $d_2(k)$ is computed, and the linear index is a linear combination of the standard deviations. Note again that different fits are reported for the NH and HI listeners. Figure 5 provides a schematic diagram for computing $Q_{\text{lin}}$.

Conceptually, $Q_{\text{lin}}$ is much easier to understand than $Q_{\text{nonlin}}$. Basically, $Q_{\text{lin}}$ is capturing how large the majority of the differences are between the long-term average spectra of the signal under test and the reference.

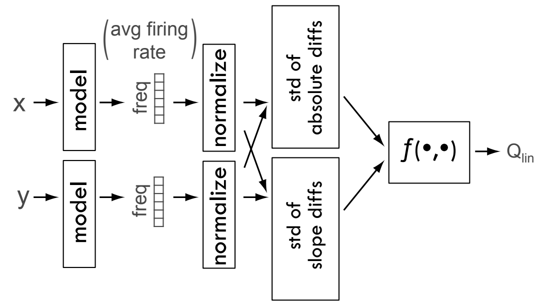


Figure 5: Schematic diagram for computing $Q_{\text{lin}}$.

### 3.2.2 Benchmarking objective measures

The benchmarking objective measures are reviewed briefly here. Refer to Hu and Loizou [28] and Loizou [40], and the respective references within, for more details. The time-domain segmental SNR (segSNR) is the average of the SNR across all time frames. Frequency-weighted segmental SNR (fwsegSNR) is the average of the SNR in each frame, where the SNR in each frame is computed as the weighted-average of the SNR in $K$ critical bands. The weighted-slope spectral distance (WSS) is computed as the weighted average of the square of the differences between the spectral slopes. The spectral slopes are estimated as the difference in the magnitudes between adjacent bands.

PESQ is the objective measure recommended by ITU-T for speech quality assessment of narrow-band handset telephony and narrow-band speech codecs (ITU-T P.862, 2001). The basic components of PESQ include time alignment, a psychoacoustic model which maps signals into perceived loudness, disturbance processing, cognitive modeling, aggregation of the disturbance in frequency and time, and finally, a mapping to the predicted subjective score [51].

The log-likelihood ratio (LLR) distance at a given frame is defined as

$$d_{\text{LLR}} = \log \frac{\mathbf{a_y R_x a_y^T}}{\mathbf{a_x R_x a_x^T}}, \tag{7}$$

where $\mathbf{a_x}$ is the linear predictive coding (LPC) vector of the clean signal, $x(t)$, $\mathbf{a_y}$ is the LPC vector of the signal under test, $y(t)$, and $\mathbf{R_x}$ is the autocorrelation matrix of $x(t)$. The LLR objective measure is the mean of the smallest 95% of the LLR distances measured at each frame. Likewise, the Itakura-Saito measure (IS) is the mean of the following distance in each frame.

$$d_{\mathrm{IS}} = \frac{\sigma_x^2}{\sigma_y^2}\left(\frac{\mathbf{a_y R_x a_y^T}}{\mathbf{a_x R_x a_x^T}}\right) + \log\left(\frac{\sigma_x^2}{\sigma_y^2}\right) - 1, \tag{8}$$

where $\sigma_x^2$ and $\sigma_y^2$ are the LPC gains of $x(t)$ and $y(t)$, respectively, and $d_{\mathrm{IS}}$ is limited between zero and 100. Finally, the cepstral distance measure (CEP) is the mean of a mean-squared error calculation between cepstral coefficient vectors of $x(t)$ and $y(t)$ at each frame. The cepstral coefficient vectors are computed recursively from the LPC vectors and are limited between zero and ten.

We compute all metrics, except PESQ, by segmenting the sentences into 30-ms frames using Hamming windows with 75% overlap between adjacent frames. Furthermore, we compute the LPC-based objective measures (LLR, IS, and CEP) using tenth-order LPC analysis.

## *3.3 Results*

### 3.3.1 Pearson analysis

Figures 6 and 7 show the absolute value of the Pearson correlation and the estimate of the standard deviation of the error, respectively. Kates and Arehart report a correlation of $r_p = 0.942$ for NH listeners [35], whereas we report a correlation of only $r_p = 0.85$. With the exception of the cepstral distance measure (CEP), all scores for the benchmarking objective measures are within $\pm 0.01$ of those reported in Hu and Loizou [28].

HASQI falls short of only PESQ and LLR in terms of the absolute value of the Pearson correlation. With a Pearson correlation of $|r_p| = 0.85$ between an "unknown"

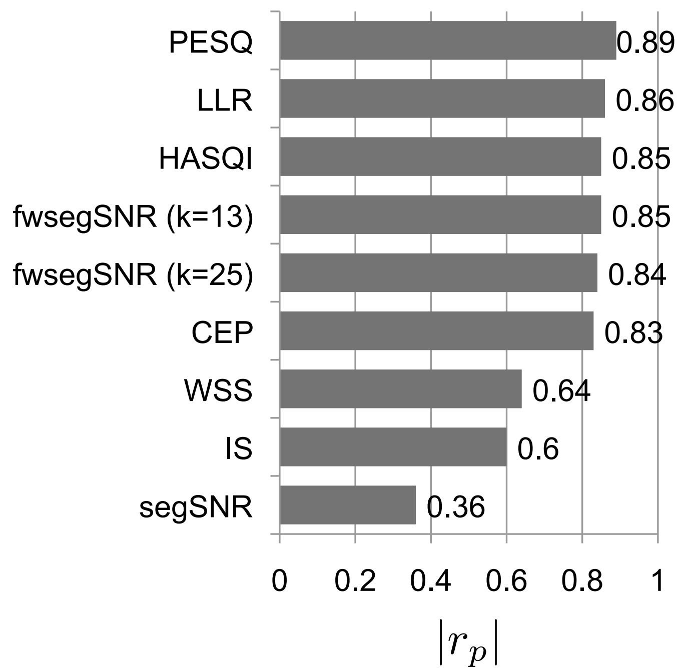


Figure 6: Absolute value of the Pearson correlation between objective and subjective scores. Objective measures are sorted in order of best performance from top to bottom.

dataset and listener scores, we have shown that HASQI is generalizable for NH listeners in the context of de-noised speech. Furthermore, the standard deviation of the error is small when HASQI is used in place of subjective scores ($\hat{\sigma}_e = 0.25$).

### 3.3.2 Spearman analysis

Figure 8 shows the absolute value of the Spearman rank correlation coefficient. A Spearman rank correlation coefficient of one would be ideal since correct ordering of the rank of perceptual quality would provide very valuable information when evaluating hearing aid algorithms. However, HASQI actually performs worse here relative to the other metrics than in the case of the Pearson analysis since it is only outperforming WSS, IS, and segSNR. Despite this initial appearance of breaking down, the absolute correlation is not bad; it is still above 0.8.

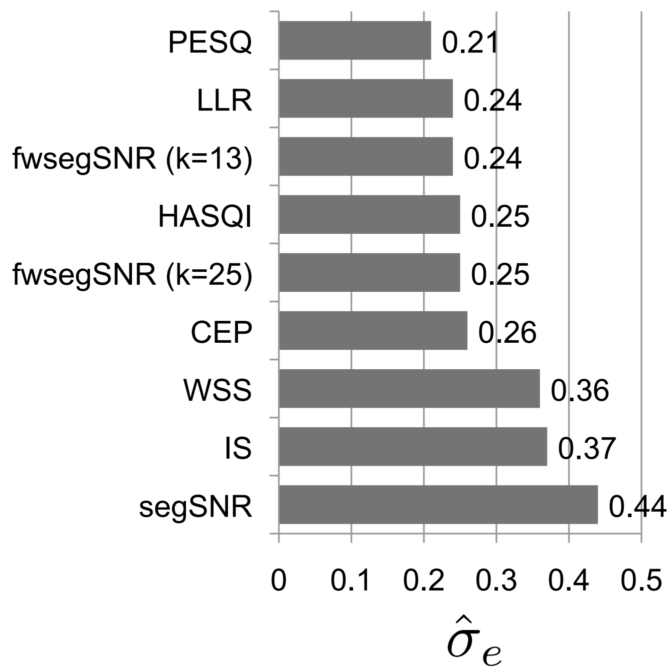


Figure 7: Estimate of the standard deviation of the error. Objective measures are sorted in order of best performance from top to bottom.

## *3.4 Discussion*

Although we see a decrease in performance between the Pearson correlation reported here and that which is reported by Kates and Arehart [35], we have shown that HASQI generalizable. It has performed reasonably well given that it was trained on only a single dataset. Furthermore, since HASQ has been validated for HI listeners, HASQI remains an attractive option for future use in evaluating hearing aid algorithms.

It should be noted though that in this study, we have looked only at the robustness of HASQI for NH listeners. Ideally we would have explored the robustness of HASQI for the HI population as well. However, because we have shown that HASQI can predict quality for “unknown” sets of speech samples and listener scores for NH listeners, we suspect that HASQI will also be able to predict quality for “unknown” sets of speech samples and listener scores for HI listeners.

It is noteworthy to mention that Kates and Arehart fit the second-order regression on the average cepstral correlation in the nonlinear index to a dataset which contained

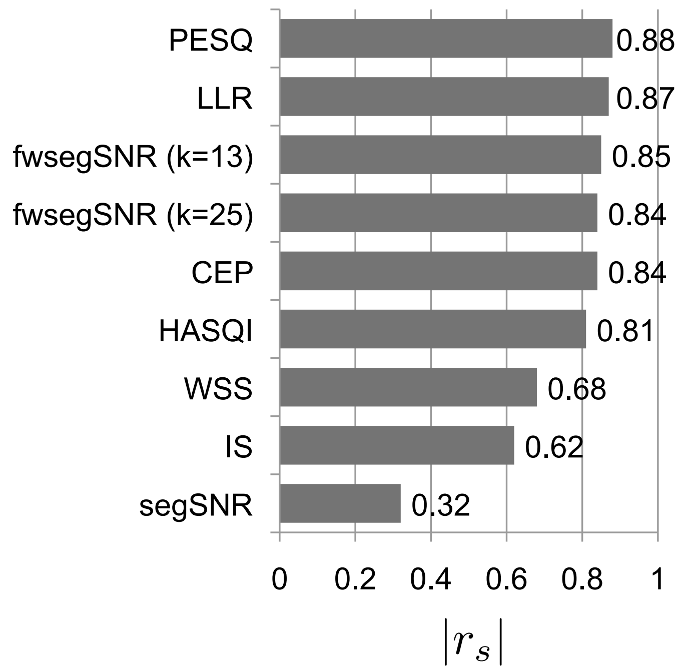


Figure 8: Absolute value of the Spearman rank correlation coefficient between objective and subjective scores. Objective measures are sorted in order of best performance from top to bottom.

average cepstral correlations only above about 0.5. As a result, Kates and Arehart assume a linear fit for average cepstral correlations below 0.5. For the dataset in this particular study, 72 of the 112 cases had average cepstral correlations below 0.5. Therefore, HASQI might benefit from training with datasets that contain samples with lower average correlations.

Since HASQI is based entirely on the signal envelope, distortions in the fine temporal structure are ignored. This may be especially significant for NH listeners, since they appear to make better use of temporal fine structure than HI listeners [35, 43]. Perhaps HASQI's predictive power would improve with the use of an auditory model which captures fine temporal structure.

As previously mentioned, all scores for the Pearson analysis were within $\pm 0.01$ of those reported by Hu and Loizou [28], with the exception of the cepstral distance (CEP). We report $|r_p| = 0.83$ and $\hat{\sigma}_e = 0.26$ here for CEP, whereas Hu and Loizou report $|r_p| = 0.79$ and $\hat{\sigma}_e = 0.29$. The most probable explanation for the improvement is that we used slightly different parameters than Hu and Loizou.

In conclusion, we have shown that HASQI is generalizable. Moreover, due to its validated extension for HI listeners, HASQI makes a useful tool for those developing algorithms specifically for the HI population.

# CHAPTER IV

# CONCLUSION

Hearing aids are tasked with the unenviable job of compensating an impaired, highly-nonlinear auditory system with severe power and size limitations. Historically, these devices have either employed linear processing or relatively unsophisticated, nonlinear processing techniques. With increasingly more accurate models of the auditory periphery, we are at a turning point in hearing aid design.

Model-based objective measures often treat the auditory system as a cascade of physical processes. As a result, they have the potential to provide more detailed information about how sound is processed and about where and why quality or intelligibility breaks down. Thus, provided that we can generalize model-based objective measures, we can use them as tools for understanding how to best process degraded signals, and therefore, how to best design hearing aids.

Many objective measures of quality and intelligibility have been proposed, and with increasing accuracy and computational power, we are capable now more than ever to utilize these measures as tools for the design and evaluation of algorithms. Given our own uncertainty about the generalizability of objective measures though, we conducted a large study on the generalizability of HASQI specifically. In doing so, we have established confidence that HASQI can provide useful information about quality. More generally though, this confidence translates to many other computational models and their corresponding objective measures.

Thus, we can look to use these computational models and objective measures for algorithm evaluation in the future. As Quackenbush, Barnwell, and Clements [49] point out, it is unlikely that any single objective measure will ever be able to accurately predict subjective quality across all possible distortions. Speech perception is too complex of a process to be modeled so simply. Despite this barrier, objective measures have still proven useful for predicting quality and evaluating algorithms. After all, modest performance from an objective measure is better than no objective measure at all.

# REFERENCES


[1] Arehart, K. H., Kates, J. M., Anderson, M. C., and Harvey, L. O. J., "Effects of noise and distortion on speech quality judgments in normal-hearing and hearing-impaired listeners," in *Journal of the Acoustical Society of America*, 2007.

[2] Beerends, J., Hekstra, A., Rix, A., and Hollier, M., "Perceptual evaluation of speech quality (PESQ) - The new ITU standard for end-to-end speech quality assessment - Part II - Psychoacoust model," *Journal of the Audio Engineering Society*, 2002.

[3] Beerends, J. and Stemerdink, J. A., "A Perceptual Audio Quality Measure Based on a Psychoacoustic Sound Representation," *Journal of the Audio Engineering Society*, 1992.

[4] Beerends, J., Eneman, K., Huber, R., Krebber, J., and Luts, H., "Speech Quality Measurement for the Hearing Impaired on the Basis of PESQ," in *Audio Engineering Society Convention 124*, 2008.

[5] Beerends, J. G., Busz, B., Oudshoorn, P., Van Vugt, J., Ahmed, K., and Niamut, O., "Degradation Decomposition of the Perceived Quality of Speech Signals on the Basis of a Perceptual Modeling Approach," *Journal of the Audio Engineering Society*, 2007.

[6] Beerends, J. G. and Stemerdink, J. A., "A Perceptual Speech-Quality Measure Based on a Psychoacoustic Sound Representation," *Journal of the Audio Engineering Society*, 1994.

[7] Bondy, J., Bruce, I. C., Becker, S., and Haykin, S., "Predicting Speech Intelligibility from a Population of Neurons," *NIPS 2003 Conference Proceedings: Advances in Neural Information Processing Systems*, 2004.

[8] Bramsløw, L., "An objective estimate of the perceived quality of reproduced sound in normal and impaired hearing," *Acta Acustica United with Acustica*, 2004.

[9] Bramsløw, L., "Validation of objective sound quality models for hearing aids," in *International Hearing Aid Conference*, 2008.

[10] Bruce, I. C., Sachs, M. B., and Young, E. D., "An auditory-periphery model of the effects of acoustic trauma on auditory nerve responses.," *The Journal of the Acoustical Society of America*, 2003.

[11] Campbell, D., Jones, E., and Glavin, M., "Audio quality assessment techniques-A review, and recent developments," *Signal Processing*, 2009.

[12] Chen, G. and Parsa, V., "Loudness pattern-based speech quality evaluation using Bayesian modeling and Markov chain Monte Carlo methods," *The Journal of the Acoustical Society of America*, 2007.

[13] Chi, T., Gao, Y., Guyton, M. C., Ru, P., and Shamma, S., "Spectro-temporal modulation transfer functions and speech intelligibility.," *The Journal of the Acoustical Society of America*, 1999.

[14] Dau, T., Püschel, D., and Kohlrausch, A., "A quantitative model of the "effective" signal processing in the auditory system. I. Model structure," *The Journal of the Acoustical Society of America*, 1996.

[15] Dau, T., Püschel, D., and Kohlrausch, A., "A quantitative model of the "effective" signal processing in the auditory system. II. Simulations and measurements," *The Journal of the Acoustical Society of America*, 1996.

[16] Elhilali, M. and Shamma, S., "Information-bearing components of speech intelligibility under babble-noise and bandlimiting distortions," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, 2008.

[17] Elhilali, M., Chi, T., and Shamma, S. A., "A spectro-temporal modulation index (STMI) for assessment of speech intelligibility," *Speech Communication*, 2003.

[18] French, N. R. and Steinberg, J., "Factors Governing the Intelligibility of Speech Sounds," *The Journal of the Acoustical Society of America*, 1947.

[19] Hansen, J. H. and Pellom, B. L., "An effective quality evaluation protocol for speech enhancement algorithms," in *Proc. Int. Conf. Spoken Lang. Process.*, 1998.

[20] Hansen, M. and Kollmeier, B., "Using a quantitative psychoacoustical signal representation for objective speech quality measurement," in *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, 1997.

[21] Hansen, M. and Kollmeier, B., "Continuous assessment of time-varying speech quality," *The Journal of the Acoustical Society of America*, 1999.

[22] Hansen, M., *Assessment and prediction of speech transmission quality with an auditory processing model.* PhD thesis, University of Oldenburg, 1998.

[23] Hansen, M. and Kollmeier, B., "Objective Modeling of Speech Quality with a Psychoacoustically Validated Auditory Model," *Journal of the Audio Engineering Society*, 2000.

[24] Hassanein, "On the use of discrete cosine transform in cepstral analysis," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 1984.

[25] Hines, A. and Harte, N., "Speech intelligibility from image processing," *Speech Communication*, 2010.

[26] Houtgast, T. and Steeneken, H. J., "Modulation Transfer-Function in Room Acoustics as a Predictor of Speech Intelligibility," *Acustica*, 1973.

[27] Hu, Y. and Loizou, P., "Subjective comparison and evaluation of speech enhancement algorithms," *Speech Communication*, 2007.

[28] Hu, Y. and Loizou, P., "Evaluation of Objective Quality Measures for Speech Enhancement," *Audio, Speech, and Language Processing, IEEE Transactions on*, 2008.

[29] Huber, R. and Kollmeier, B., "PEMO-Q: A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception," *Audio, Speech, and Language Processing, IEEE Transactions on*, 2006.

[30] Huo, L., Waltermann, M., Scholz, K., Raake, A., Heute, U., and Moller, S., "Estimation Model for the Speech-Quality Dimension "Directness / Frequency Content"," *Applications of Signal Processing to Audio and Acoustics, 2007 IEEE Workshop on*, 2007.

[31] Ibrahim, R. A. and Bruce, I. C., "The Interaction of Envelope and Temporal Fine Structure Cues in Speech Perception," in *Abstracts of the 34th ARO Midwinter Research Meeting*, 2011.

[32] Karjalainen, "A new auditory model for the evaluation of sound quality of audio systems," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '85*, 1985.

[33] Kates, J. M., "A central spectrum model for the perception of coloration in filtered Gaussian noise," *The Journal of the Acoustical Society of America*, 1985.

[34] Kates, J. M., "Quality ratings for frequency-shaped peak-clipped speech," *The Journal of the Acoustical Society of America*, 1994.

[35] Kates, J. M. and Arehart, K. H., "The Hearing-Aid Speech Quality Index (HASQI)," *Journal of the Audio Engineering Society*, 2010.

[36] Klatt, "Prediction of perceived phonetic distance from critical-band spectra: A first step," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '82*, 1982.

[37] Kozma-Spytek, L., Kates, J. M., and Revoile, S. G., "Quality ratings for frequency-shaped peak-clipped speech: results for listeners with hearing loss.," *Journal of speech and hearing research*, 1996.

[38] Kryter, K. D., “Methods for the Calculation and Use of the Articulation Index,” *The Journal of the Acoustical Society of America*, 1962.

[39] Kryter, K. D., “Validation of the Articulation Index,” *The Journal of the Acoustical Society of America*, 1962.

[40] Loizou, P. C., *Speech Enhancement: Theory and Practice (Signal Processing and Communications).* CRC Press, 2007.

[41] Moore, B. and Tan, C., “Development and validation of a method for predicting the perceived naturalness of sounds subjected to spectral distortion,” *Journal of the Audio Engineering Society*, 2004.

[42] Moore, B., Tan, C., Zacharov, N., and Mattila, V., “Measuring and predicting the perceived quality of music and speech subjected to combined linear and nonlinear distortion,” *Journal of the Audio Engineering Society*, 2004.

[43] Moore, B. C. J., “The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people.,” *Journal of the Association for Research in Otolaryngology : JARO*, 2008.

[44] Moore, B. C. J., Glasberg, B. R., and Baer, T., “A Model for the Prediction of Thresholds, Loudness, and Partial Loudness,” *Journal of the Audio Engineering Society*, 1997.

[45] Nielsen, L. B., “Objective Scaling of Sound Quality for Normal-Hearing and Hearing-Impaired Listeners,” in *Audio Engineering Society Convention 98*, 1995.

[46] Noll, P., “Adaptive quantization in speech coding systems,” in *lnt. Zurich Seminar on Digital Communication (IEEE)*, 1976.

[47] Parsa, V. and Jamieson, D. G., “Hearing Aid Distortion Measurement Using the Auditory Distance Parameter,” in *Audio Engineering Society Convention 111*, 2001.

[48] Pavlovic, C., “The speech intelligibility index standard and its relationship to the articulation index, and the speech transmission index,” *The Journal of the Acoustical Society of America*, 2006.

[49] Quackenbush, S. R., Barnwell III, T. P., and Clements, M. A., *Objective Measures of Speech Quality.* Prentice Hall, 1988.

[50] Rix, A. W. and Hollier, M. P., “The perceptual analysis measurement system for robust end-to-end speech quality assessment,” in *Acoustics, Speech, and Signal Processing, 2000. ICASSP ’00. Proceedings. 2000 IEEE International Conference on*, 2000.

[51] Rix, A. W., Beerends, J. G., Hollier, M. P., and Hekstra, A. P., "Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs," in *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on*, 2001.

[52] Tan, C.-T. and Moore, B. C. J., "Perception of nonlinear distortion by hearing-impaired people," *International Journal of Audiology*, 2008.

[53] Tan, C., Moore, B., Zacharov, N., and Mattila, V., "Predicting the perceived quality of nonlinearly distorted music and speech signals," *Journal of the Audio Engineering Society*, 2004.

[54] Thiede, T., Treurniet, W., Bitto, R., Schmidmer, C., Sporer, T., Beerends, J., Colomes, C., Keyhl, M., Stoll, G., Brandenburg, K., and Feiten, B., "PEAQ - The ITU standard for objective measurement of perceived audio quality," *Journal of the Audio Engineering Society*, 2000.

[55] Tribolet, "A study of complexity and quality of speech waveform coders," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '78*, 1978.

[56] Voran, S., "Objective estimation of perceived speech quality. I. Development of the measuring normalizing block technique," *Speech and Audio Processing, IEEE Transactions on*, 1999.

[57] Wang, "An objective measure for predicting subjective quality of speech coders," *Selected Areas in Communications, IEEE Journal on*, 1992.

[58] Zilany, M. S. A. and Bruce, I. C., "Predictions of Speech Intelligibility with a Model of the Normal and Impaired Auditory-periphery," in *Neural Engineering, 2007. CNE '07. 3rd International IEEE/EMBS Conference on*, 2007.

[59] Zilany, M. S. A. and Carney, L. H., "Power-Law Dynamics in an Auditory-Nerve Model Can Account for Neural Adaptation to Sound-Level Statistics," *Journal of Neuroscience*, 2010.

[60] Zilany, M. S. A. and Bruce, I. C., "Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery.," *The Journal of the Acoustical Society of America*, 2006.

[61] Zilany, M. S. A. and Bruce, I. C., "Representation of the vowel /epsilon/ in normal and impaired auditory nerve fibers: model predictions of responses in cats.," *The Journal of the Acoustical Society of America*, 2007.

# VITA

## Education

Attended Vanderbilt University School of Medicine, 2008-2009

Graduate studies in Hearing and Speech Sciences

Attended Washington University in St. Louis, 2004-2007

Received Bachelor of Science in Biomedical Engineering, 2007

## Academic honors and awards

First place in 21st Annual SAIC Student Paper Competition, 2010

National Defense Science & Engineering Graduate Fellowship, 2010-2013

National Science Foundation Graduate Research Fellowship, 2010-2015

President's Fellowship - Georgia Institute of Technology, 2009-2013

Department of Education training grant - Vanderbilt University, 2008-2009

Elected to Tau Beta Pi Engineering Honor Society, 2006

Jeffrey & Nancy Balter Biomedical Engineering Scholar, 2004-2007

Gustav Kurt Mesmer Scholar, 2004-2005

Society of Women Engineers Chicago Chapter Scholar, 2004-2005

James R. Hoffa Memorial Scholar, 2004-2005

## Publications

Kressner, A.A., Anderson, D.V., and Rozell, C.J. Robustness of the Hearing Aid Speech Quality Index (HASQI). In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, October 2011. Submitted.

Kressner, A.A., Anderson, D.V., and Rozell, C.J. Computational auditory models validate the intelligibility benefits of efficient filters. In *International Symposium on Auditory and Audiological Research (ISAAR)*, Nyborg, Denmark, August 2011. Accepted.

Charles, A., Kressner, A.A., and Rozell, C.J. A Causal Locally Competitive Algorithm for the Sparse Decomposition of Audio Signals. In *Proceedings of the IEEE Digital Signal Processing (DSP) Workshop*, Sedona, AZ, January 2011.

Kressner, A.A., Rozell, C.J., and Anderson, D.V. Predicting speech quality using a computational auditory model. In *International Hearing Aid Research Conference (IHCON)*, Lake Tahoe, CA, 11-15 August 2010.