Speech distortion measures for hearing aids*

MALCOLM J. WILLIAMSON, M.A.,** KENNETH L. CUMMINS, Ph.D., and KURT E. HECOX, M.D., Ph.D.

*Portions of this article were presented at the Annual Meeting of the American Speech-Language-Hearing Association, November 1986.
**Address all correspondence to: Malcolm Williamson, Project Phoenix of Madison, Inc., P.O. Box 4287, Madison, WI 53711-0287.

INTRODUCTION

The effect of distortion on the intelligibility and quality of audio signals has interested acousticians for at least 150 years (5,6). The conceptual approach to distortion and the analytic tools brought to bear on its measurement have evolved considerably since that time. However, modern audioengineers still widely agree with the 1943 statement that “Distortion in its several forms is undoubtedly the most serious problem with which the engineer has to deal in the design of a satisfactory high fidelity sound reproducing system” (1).

As analog amplifier circuitry has evolved, the precision and complexity required of distortion measures have increased. Traditional measures such as harmonic, intermodulation, and transient distortion have been extended considerably. For example, there are at least six different schemes for specifying dynamic intermodulation distortion (7). With the introduction of digital technology the number of types of distortion will increase significantly. Both the process of data conversion and the signal processing transformations are potential sources of distortion. The distortions associated with digital technology are unfamiliar to most audioengineers and acousticians, and there is presently no consensus on measurement standards for digital devices.

The purpose of this paper is to propose a more general approach to the problem of measuring distortion so that the introduction of new processing techniques does not require the proliferation of new distortion measures. From this unifying approach, an overall measure of distortion can be obtained and displayed. By constraining the time-varying aspects of the linear model, it is possible to constrain the forms of nonlinearities that contribute to this overall measure. The resulting measure can then be correlated to perceptual judgments or subdivided into specific forms of distortions which can then be related to listener behavior.
The approach described in this note is put forth as a "working model" whose final form will require additional theoretical and empirical efforts. The need for more general methods of describing audio distortions is pressing because the introduction of digital hearing aids is imminent and because currently available methods of describing distortion do not adequately characterize the perceived performance of hearing instruments—whether digital or analog.

Following is a brief description of the model from which the approach is derived, the parameters involved in the implementation of the model, and typical results from the use of the model.

METHODS AND RESULTS

General Approach

The conceptual basis for our approach is outlined in Figure 1. The hearing aid is modeled as a linear filter with time varying characteristics. The time varying properties of the model permit the incorporation of hearing aid characteristics such as automatic gain control or compression. The filter properties are estimated from actual acoustic inputs and outputs of the aid. The output of the linear filter \( z(t) \) is then estimated and compared to the actual output of the hearing aid \( y(t) \). The difference \( d(t) \) between the hearing aid’s output and linear filter’s output represents the non-linear portion of the hearing aid’s performance. Our ultimate objective is to make the output of the filter, \( z(t) \) and \( d(t) \) map more directly to the perceived linear and distorted portions of the signal, respectively. Note that \( d(t) \) contains both the harmonic and intermodulation distortion, in addition to the less familiar forms of distortion.

Specific Methods

Figure 2 represents the experimental set-up. The test stimulus was the word “rapid”, which was digitized and stored on a Masscomp 531. The digitization rate was 20 kHz and the word was low-pass-filtered at 8 kHz prior to digitization. The word was delivered via the 12-bit digital-to-analog converters, filtered, amplified, and routed to an acoustically shielded sound chamber that enclosed the hearing aid, coupler and associated transducers. The output of the aid was amplified, filtered, digitized, and stored for analysis on a SUN workstation.

The volume setting for the hearing aid used to generate the figures in this note was set at a moderate position and the aid’s frequency-response curve was not atypical of many commercially available units. We will not dwell on the hearing aid characteristics since this report focuses attention on the measurement philosophy and its implementation.

The display format is shown in Figure 3 where the input signal is shown as a time, frequency, and amplitude plot. Frequency is displayed from low to high as you move from right to left, while time increases in 6.4 ms steps per sample as you move from front to back. The vertical (amplitude) units depend upon the specific dependent variable of interest (e.g., magnitude of gain, distortion, etc.). Division of the frequency scale is logarithmic, ranging from 125 Hz to 8 kHz.

The input signal was the word “rapid,” which was chosen only for reasons of convenience. Similar results are obtained with other speech input wave-
forms. These plots can be rotated to better visualize specific aspects of the waveform that are not easily seen from a particular display perspective.

The distortion measure $d(t)$ is dependent upon the linear filter, so the estimation of the linear filter is a critical part of the process. In our implementation we recomputed the filter every 6.4 milliseconds. It was restricted to a finite impulse response filter with 15 taps (.75 millisecond impulse response). The estimation used Levinson’s algorithm (2) on 12.8 milliseconds of data preprocessed with a Hanning window. Levinson’s algorithm finds an optimal filter that minimizes the energy of $d(t)$. From the filter coefficients we calculate the transfer function. This gives us the gain characteristics of the aid as a function of frequency. We also need to calculate the spectra of the speech and distortion waveforms. For this we used linear prediction spectral analysis based on the first 15 autocorrelations.

A number of other indices used for the above described calculations could also be displayed to the user, if desired (e.g., $S_{xx}$, $S_{xy}$, $S_{yy}$, etc.).

The smoothness of the input and output plots is a result of the spectral estimation procedure that was used—autoregressive with 15 lags.

**RESULTS**

A speech input waveform ("rapid") is shown in Figure 3. The first formant is well visualized and
the time varying aspect of the amplitude easily appreciated. The advantage of rotating the display is shown in Figure 4, where a relatively minor change in display perspective makes visualization of the second formant easier.

A typical output waveform is shown in Figure 5. Note the relative enhancement of the second formant, reflecting the high-frequency gain of the device. Again the time varying aspect of the absolute and relative amplitudes are apparent.

A clearer picture of the system’s time dependent gain can be obtained by a plot such as that shown in Figure 6. Both the high-frequency emphasis of the gain and its temporal dependence are easily visualized. A visual comparison of Figure 6 and Figure 3 reveals the input level dependence of the gain. The level dependence of the gain helps distinguish linear from compressive systems.

The vertical scale for Figure 6 is dB, whereas the scale for other plots was rms. We chose different scales (dB for the gain and rms for signal levels) to enhance the relevant information.

Figure 7 shows a plot of d(t). Note that the frequency content of the distortion is spread across frequencies; the energy is widely distributed rather than at simple harmonics as it would be for a pure-tone input. It is spread more evenly across the spectrum than the input speech, and follows the amplitude of the speech signal. As expected, at low input levels the aid operates more linearly and as the input level increases so does the distortion.

This is a qualitative view of the distortion. This type of plot can be used to determine the degree of distortion at the beginning of words, or whether the consonants are affected more than the vowels. In addition, data can be collected for comparison. For example, it is possible to average along the time dimension and determine how distortion is spread across frequencies, or we can replot data to determine how the distortion level varies with the speech level. These data can be presented as distortion percentages and the distortion analyzed statistically.

**DISCUSSION**

The chief advantages of the proposed method are: 1) the generality of the technique; 2) the ease of visualization and interpretation of the results; 3) the ability to capture the time varying aspects of a system’s performance; 4) the ability to accommodate both simple or complex inputs with equal ease; 5) the ability to exclude some forms of distortion which do not contribute to perceived distortion; and 6) the fact that it makes available a consolidated measure of multiple types of distortion for the first time. The computational complexity of the approach is not inconsistent with implementation on a microprocessor-based system. Finally, the outcome of this approach should be interpretable by clinicians regardless of their engineering or mathematics background.

One of the most persistent challenges in the area of acoustic distortions has been to define the relationship between the amount of distortion and its impact on listener behavior. Significant effort has been expended in defining the intensity at which
Section IV. Issues of Measurement: Williamson et al.

Figure 6. Hearing aid gain as a function of time is shown. Scale units are the same here as for earlier figures, except amplitude is plotted in dB instead of rms.

Figure 7. Hearing aid distortion as a function of time. The amplitude scale has been exaggerated compared to earlier plots to enhance the visualization of the distortion energy.

distortion becomes audible, while less is known with respect to the consequences of distortion on intelligibility or sound quality—especially for hearing aids. Some of the earliest efforts in relating distortions to their perceptual consequences was reported in work by Licklider and colleagues (3,4) and by Ian Thomas (8,9).

In general, studies by these investigators emphasized the remarkable immunity of speech signals to degradation by distortion—at least with respect to intelligibility. Specifically, the lack of correlation between measures of harmonic distortion and speech intelligibility has been established for years, yet harmonic distortion remains the most commonly used measure for hearing aids.

Implementation of the distortion measures described herein requires the selection of a number of parameter values. The rules by which the “best” parameter values are determined are largely unknown, although parameter-setting decisions of this nature are historically based on engineering criteria. Current measures of distortion do not “weight” various forms of distortion to reflect their impact on either intelligibility or sound quality. The selection of parameters in a fashion that takes into account the perceptual salience of a particular distortion would be of tremendous advantage to the clinician. Review of the literature suggests that the behavioral data needed to establish a perceptually derived distortion index does not exist. It is especially in this regard that we must consider the approach to distortion analysis described in this note as a working model.

In addition to serving as an approximate measure of sound quality, distortion analysis can help direct a manufacturer or service engineer in determining the presence or nature of a system malfunction. It should be recognized that the most sensitive or specific measure of distortion (from the standpoint of monitoring the electroacoustic performance of a circuit) may not be the most relevant distortion from the perceptual standpoint. In the past, the needs of the engineering/manufacturing community for distortion measurement have not been clearly separated from the needs of a clinician in selecting or fitting a hearing device. The introduction of digital hearing aids is likely to underscore the fact that a single distortion measure will not fulfill the needs of both groups. We believe that the approach outlined in this paper can serve as a foundation from which a more perceptually relevant set of distortion measures can be derived and applied to the characterization of digital or analog hearing aids.

ACKNOWLEDGMENTS

The authors would like to acknowledge the assistance of Tom Worrall in the implementation of graphics software, Joyce Tikalsky for her editing and review efforts, and Jan Habich for assistance in manuscript preparation.
REFERENCES