Reprinted from APPLICATION OF SIGNAL PROCESSING CONCEPTS IN HEARING AIDS
Edited by Paul Yanick, Jr., M.A. and Stephen Freifeld, M.D., F.A.C.S.

Chapter 8

A REVIEW OF SIGNAL PROCESSING TECHNIQUES FOR THE HEARING IMPAIRED

H. Drucker, Ph.D.

Background

A large class of techniques exists in the speech processing field that can be used to increase the intelligibility of speech as perceived by a large class of sensorineurally hearing impaired persons, although many of these techniques originated independently of the hearing aid field. Speech, of course, is highly redundant so that a listener, normal hearing or not, in missing a phoneme or entire words because of poor pronunciation by the speaker or competing noise, can mentally reinsert the missing speech on hearing the rest of the text. A hearing impaired person misses more of the intelligence and must work harder to use the contextual clues. There comes a point when the competing noise is so great that the hearing impaired cannot function, while the normal hearing person may have little problem. The emphasis in signal processing is mainly on the high frequency sounds since the high frequency sounds are typically consonants and they carry more of the intelligibility than the low frequency vowels (1). Thus those persons who have some type of high frequency response loss would tend to show a greater intelligibility loss than those with a low frequency loss. In addition to the fact that the consonants tend to be at high frequencies, they have lower overall power level (2). This is not to say that instantaneously, (say in the stop sounds such as /p/ and /t/), the consonants do not have a large amount of energy, but the overall power level taken over the full period of the consonants is lower than that of the vowels.

One of the more classical techniques tested on the normal hearing was performed by Licklider and Pollack (3). In one experiment, normal speech was infinitely amplified and clipped so that only zero crossing information was retained. The speech was fairly intelligible but unnatural. In a second experiment, the speech was first differentiated and then clipped and then it was determined that the intelligibility of the clipped differentiated speech was higher than the clipped speech. The rationale for this was that differentiating speech was equivalent to high frequency emphasis of speech, raising the power level of the high frequency consonants. Clipped differentiated speech is extremely unnatural and fatiguing to listen to. First, all background noise is raised to high levels filling the silence between words. Secondly, the clipping process generates harmonics and intermodulation frequencies of the speech, many falling within the speech band. As an example, assume for simplification that a sound consists of an assymetrical 1000 Hz first formant and a 1500 Hz second formant. The clipping process would then generate 1000 Hz and multiples of 1000 Hz, i.e. 2000, 3000, 4000, ... Hz, multiples of 1500 Hz and then the intermodulation products which are the sum and difference frequencies, i.e. 500, 1000, 1500 ... Hz. Many of these frequencies are within the speech band.

Recent efforts made by Nelderjohn (4) have improved the naturalness of clipped speech. In a SSB (Single Side Band) modulation process, the speech band is translated to higher frequencies. Thus if the original speech band is considered to go from 500 Hz to 5000 Hz, a SSB process would (at a carrier frequency of 40 kHz) raise the speech band to a 40,500 to 45,000 Hz band. Clipping is then performed. However, the harmonics are now near 80 kHz and intermodulation products at low frequencies, outside the 40.5 kHz-45 kHz band so that when another SSB modulation is performed to bring the band down to 500-5000 kHz, no distortion terms are present in the output. The output sounds quite natural and intelligible, however its application to the hearing impaired has not yet been tested.

Although these techniques to raise the power of the consonants can help the moderately impaired, individuals with severe high frequency sensorineural loss cannot be helped because their dynamic range at high frequencies is severely limited. If, as an example, their dynamic range at high frequencies is 10 dB and the dynamic range of conversational speech is 36 dB (2), there is no linear amplifier or high frequency pre-emphasis that can alter this 36 dB dynamic range to fit within the 10 dB perceivable range. Clipping, a nonlinear process, although raising the overall power level, actually increases the dynamic range and would not be helpful. Sensorineural loss is also usually accompanied by recruitment, an abnormal perception of loudness. This causes vowel masking of following consonants and an inability to function in a competing speechlike noise environment. Techniques applicable to overcome the recruitment-dynamic range problem are frequency and amplitude compression.

FREQUENCY COMPRESSION TECHNIQUES

In frequency compression techniques, the high frequency information is moved to lower frequencies where the dynamic range is usually greater. One process that immediately comes to mind is to move, through the SSB modulation process, the 500-5000 Hz band to 0-4.5 kHz. This is an additive process in which all frequencies are moved equally. However, this does not work since it destroys the harmonic relationship between the pitch frequency and its harmonics. The resultant speech is completely unintelligible. What is needed is a multiplicative process. For instance, if we take an audio tape and slow it down by a factor of .8 then all frequencies are shifted by a factor .8 and thus the 500-5 kHz band becomes 400 to 4 kHz, compressing the original bandwidth by .8 and preserving the harmonic relationship. Unfortunately, the total message time increases by a factor of 1.25 and thus we do not have a real time technique. Since the speech waveform is so highly redundant, it is possible to discard some of the slowed-time segments to achieve the frequency compression in real time. The problem is in selectively removing segments without destroying the intelligibility. In a series of experiments (5) conducted using this technique on the moderately hearing impaired, intelligibility increased as the frequency band was reduced up to 80% of its initial value. After that point, intelligibility decreased. The conclusion was that moderate frequency compression would aid the moderately impaired. Other frequency compression techniques are reviewed in (5).

AMPLITUDE COMPRESSION TECHNIQUES

Linear amplification with high frequency preemphasis is effective if the dynamic range of the listener is greater than the dynamic range of speech at high frequencies. For the severe hearing loss cases, however, some nonlinear technique must be used. The necessary compression technique is discussed in detail in a later chapter and other articles (6,7,8,9,10). Basically, in the most effective technique, the speech band is broken into two smaller bands using 1500 kHz low pass and high pass filters to separate speech into (approximately) the consonants and vowels. Amplitude compression is performed on both bands to reduce the dynamic range, the two channels are recombined and equalization performed. The most critical aspect necessary to increase the intelligibility is the compression amplifier design, which is a variable gain device. The compression amplifiers compress over most of the dynamic range but below a threshold, 30-36 dB below the maximum input amplitudes, the amplifier expands. This causes low level background noise to be suppressed. The amplifier must respond to the sudden onset of large amplitudes within 2 msec for a 40 dB increase in input value. For a 40 dB decrease in the input value, the output must reach a steady state gain condition within 20 msec. The release time is critical-- too short and the speech sounds unnatural, too long and information bearing transients are not amplified to perceivable levels. The maximum output level for each band should be the same so that the consonants have energy levels higher than they would have in ordinary speech. A series of experiments (7,8) using experienced hearing aid users showed a significant increase in performance in the presence of human-type noise with the two channel compression system.

DIGITAL PROCESSING TECHNIQUES

With the increasing availability of low cost digital processors, there are now available a host of techniques for analyzing speech. Filtering and amplification can now be done completely using digital computer programs rather than hardware (11). Therefore, changes in procedures can readily be accomplished by changing a computer program rather than making expensive hardware changes. However, any type of complex filtering cannot be done in real time by a programmed computer and so the digital processing techniques can only be used to show the feasibility of some hypothesized speech processing technique. To operate in real time, special purpose digital hardware or analog circuitry must be constructed. In general, analog circuitry is still less expensive than digital hardware. Because the military and the telephone companies have been historically interested in bandwidth compression techniques so that more conversations can be transmitted over a given communication channel, extensive digital hardware does exist for this purpose. However, these techniques cannot be used by the hearing impaired since they are in such a format that is unintelligible until a reverse process has been performed which restores the speech to its original bandwidth. Thus, at the present time, completely digital techniques are useful for analyzing speech and simulating analog systems but not for real time processing. A hybrid scheme that consists of digital and analog hardware is described in the chapter entitled, "A Microprocessor-Based Speech Processor", and allows more flexibility than the analog circuit forms.

PREDICTING THE EFFECTS OF SIGNAL PROCESSING

The function of signal processing is to transform the signal parameters so that to the hearing impaired, the signal received at the cochlea is the same as an unprocessed signal would be to a normal hearing listener. Thus, the signal processing ideally reverses the distortion present due to sensorineural impairment. If the individual has high frequency hearing loss, we use high frequency pre-emphasis. For dynamic range loss, we use amplitude compression. In designing a hearing aid, or any speech processing system, the engineer uses amplifiers and filters. The behavior of these two devices can be predicted to great accuracy and corrections obtained to the inadequate frequency response. The frequency response of the microphone and receiver must also be known and through curves obtained from the manufacturer, the designer knows whether these devices are adequate. The frequency response of the better electret microphone is flat within the speech frequency band, however, the frequency response of the receiver may not be so. The frequency response may be inadequate, i.e., high frequencies or low frequencies may be attenuated, or there may be valleys or peaks in the response. In any of these cases, the amplifier frequency response can be changed to correct for any deficiencies in the reciever. All this is straightforward. However to completely correct for the listener's frequency response, the designer must take into account that the output of the receiver is presented to the ear through an acoustical tube fitted into an earmold. The frequency response of the total system will then also depend on the dimensions of the ear canal and the earmold configeration. Correction of the frequency response involves the total system from microphone to ear canal. The use of a standard coupler to simulate the ear canal is helpful but ideally, every hearing aid should reflect that individual's physical characteristics. Until recently, this could not be predicted without in situ measurements of the frequency response. However, Egolf (12) has shown how to predict the overall frequency response from the electrical parameters of the hearing aid, the earmold configeration, and ear canal measurements. A computer program is used to make these predictions, the input to the program being electrical characteristics and physical measurements. As pointed out previously, the digital computer can be used to simulate the hearing aid with the parameters of the simulated hearing aid easily variable to optimize the individual's speech perception. Finally, the computer can be used to design the electrical network. Although all three types of programs presently exist; prediction of frequency response, digital processing of signals, and electronic circuit synthesis, no one as of this date has tied these three types of programs together.

REFERENCES

1. French NR, Steinberg JC: Factors governing the intelligibility of speech. J Acoust Soc Am 19, 1949
2. Dunn HK, White SD: Statistical measurements on conversational speech. J Acoust Soc Am 11:278-298, 1938
3. Licklider JCR, Pollack I: Effects of differentiation, integration, and infinite peak clipping upon the intelligibility of speech. J Acoust Soc Am 20:42-51, 1947
4. Thomas IB, Nelderjohn RJ: The intelligibility of filtered-clipped speech in noise. IEEE Trans on ASSP 18:15-18, No 3, 1970
5. Mazor M, Simon H, Schemberg J, Levitt H: Moderate frequency compression for the moderately hearing impaired. J Acoust Soc Am 62:1273-1278, No 5, 1977
6. Yanick P: Improvement on speech discrimination with compression vs. linear amplification. J Audit Res 13:333-338, 1973
7. Yanick P: The Effects of signal processing on the intelligibility of speech in noise for subjects possessing sensorineural hearing loss. J Amer Aud Soc 2:5, 1976
8. Yanick P, Drucker H: Signal processing to improve intelligibility in the presence of noise for persons with ski-slope hearing loss. IEEE Trans on ASSP 24:507-512, No 6, 1976

9. Yanick P: The effects of signal processing on consonant identification for persons with sensorineural hearing loss. J Soc Medical Aud 4, No 4 1977
10. Villchur E: Signal processing to improve speech intelligibility in perceptive deafness. J Acoust Soc Am 53:1646-1657, 1973
11. Rabiner LR, Rader CM (editors): Digital signal processing. IEEE Press, 1972
12. Egolf D: A mathematic scheme for predicting the electro-acoustic frequency response of hearing aid receiver-earmold-ear systems. Ph.D. dissertation, Purdue University, W. Lafayette, Ind., August, 1976