

Real-Time Combined Compression and Linear Gain Processing for Digital Hearing Aids

Juan Carlos Tejero-Calado, Janet C. Rutledge** and Peggy B. Nelson***

**University of Malaga, Campus de Teatinos-Complejo Tecnológico, 29080-Malaga (Spain)*

***Div. of Otolaryngology, Univ. of Maryland, 16 S. Eutaw St., Baltimore, MD 21201 (USA)*

Summary

Amplitude compression processing is used to reduce the amplitude level variations of speech to fit the reduced dynamic ranges of sensorineural impaired listeners. However this processing results in spectral smearing due in part to reduced peak-to-valley ratios. Presented here is a processing algorithm based on a sinusoidal speech model, implemented in real time, which preserves the important spectral peaks through a hybrid of compression and linear gain processing. Preliminary listening tests indicate that listeners with moderate hearing losses showed marked benefit from the processing algorithm through increased detectability of high-frequency information.

INTRODUCTION

Sensorineural hearing losses are characterized by a reduced dynamic range of hearing and reduced spectral resolution. Compensating for elevated thresholds involves amplification to raise speech above threshold, while using amplitude limiting to keep speech signals from exceeding the impaired listener's threshold of discomfort. Linear techniques apply gain directly to the amplitude of the incoming signal. Peak clipping and compression limiting are used to keep high level signals from exceeding the listener's threshold of discomfort.

Nonlinear techniques, such as amplitude compression, reduce the amplitude level variations of the signal to fit the listener's reduced dynamic range of hearing. Single-channel (wideband) systems process the entire speech signal on the basis of overall level. Multiband syllabic compression systems reduce the variation in speech level in each frequency band according to the subject's reduced dynamic range in that band. The wideband and multiband compression systems mostly use digital or analog filters along with equalization gain. In most cases, compression parameters remain the same over time.

Waveform parameterization models, such as wavelets or sinusoidal models, can be used in place of filter-based techniques to achieve linear or compression processing (1-3). These models allow greater flexibility in the range of compensation processing, and lend themselves to time-varying techniques. Compression processes commonly reduce spectral peak-to-valley ratios in speech. Resolution of spectral peaks is critical to speech perception. Tracking formants and then tailoring processing to maintain those formants is difficult and impractical. Because the sinusoidal model allows manipulation of individual frequency components in each frame, those peaks with the greatest energy in each band can be processed so that their relative shape is maintained. Presented here is a processing algorithm which combines linear and compression processing to maintain the best resolution possible for the critical peaks in each frame of speech.

The sinusoidal model developed by Quatieri and McAulay (4) represents speech as the sum of sinusoids with various amplitudes, frequencies and phases. This high quality model has parameters which are independent of voicing state and pitch period. The frequencies of the sinusoids in frame k are chosen to correspond to the $N(k)$ largest peaks in the magnitude of the short-time Fourier transform of the speech signal. The application here, which has been implemented on a TMS320C30 microprocessor, uses 7.5 msec analysis frames and 30 msec Hamming windows, leading to a 4 to 1 time overlap. A 256 point FFT is used to provide sufficient resolution for the speech sampled at 8 .025 kHz. Synthesis may be done by summing up the sinusoid components or by using an inverse FFT (2).

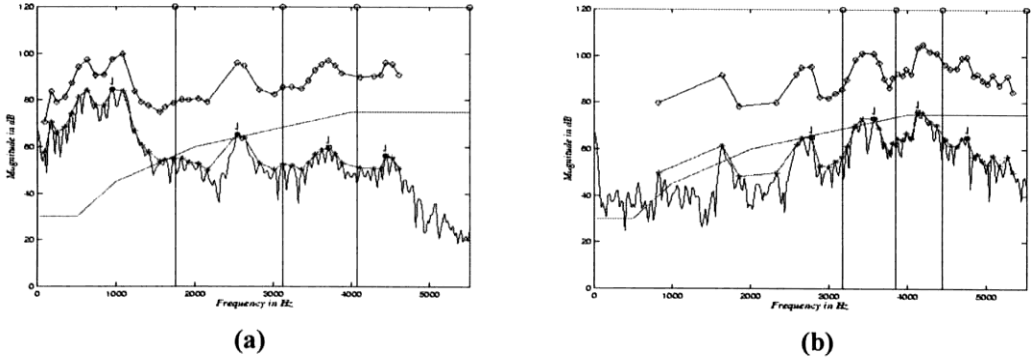


Figure 1: Speech spectrum indicating top four spectrally important peaks and their frequency bands before and after processing. (a) Vowel region. (b) Consonant region.

The processing algorithm assumes there are up to six important peaks in each frame. The number of peaks selected is determined by the constraint that the peaks must be a minimum frequency spacing from each other. If two spectral peaks are close in frequency, it is assumed that they arise from a single formant. In the examples shown in Figure 1, the top four peaks were chosen. Frequency bands are created by selecting the midpoint bisector between each pair of peaks as the boundaries of the frequency bands. The processing is optimized for the main peak in each band. The compression ratio is calculated for each peak in each frame as the ratio between the impaired listener's dynamic range of hearing and the normal listener's dynamic range. These are represented in Figure 2 as Δ^* and Δ , respectively. The gain for a given peak is calculated such that the ratio of the peak sensation levels for normal versus impaired listeners is equal to the ratio of their respective dynamic ranges (1). In other words,

$$c = \Delta^* / \Delta = \delta^* / \delta, \quad (1)$$

where Δ is the normal dynamic range and Δ^* is impaired dynamic range at the frequency of the peak, δ is the sensation level for the sinusoid for normal listeners, and δ^* is the sensation level for the processed sinusoid for impaired listeners. The amplitude of the processed peak is

$$A^* = cA + T_{im} \quad (2)$$

where, T_{im} is the impaired threshold of hearing, and A is the amplitude of the original sinusoid peak. The resulting gain is the difference between A^* and A .

In order to preserve the relative peak-to-valley ratio of the main peak in each band, the gain applied to the peak is applied to all sinusoidal model amplitudes in that band. The result is a combination of compression and linear gain processing: compression is applied to the amplitude of the main peak, and the gain applied to all other peaks in that band is a constant linear gain. This processing leads to jump discontinuities at the boundaries of the frequency bands. To compensate for this an adjustment is made to the algorithm to linearly transition from the gain in one band to the gain in the next band in the valley region. The result is that the peak-to-valley ratio is not perfectly maintained. However it is evident from Figure 1 that the important peaks are clearly distinguishable in the processed speech and do not suffer from the smearing that accompanies conventional compression processing.

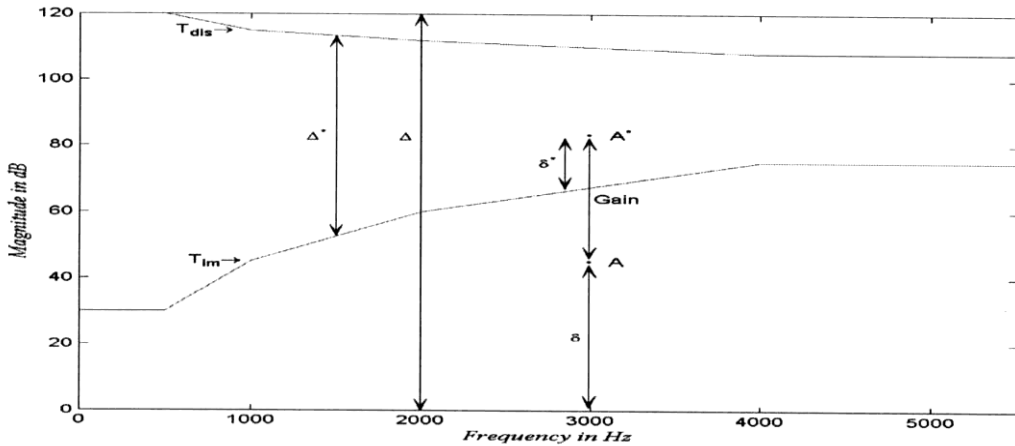


Figure 2: Calculation of gains based on the distance the sinusoidal model coefficient is above the threshold of hearing for the normal and impaired listener.

To reduce ambient background noise in silent regions, if the maximum peak in a band is below 30 dB, it is only given the gain that would be applied to a 30-dB peak. It is important to apply processing in these silent regions in order to avoid audible gaps in the processed signal.

RESULTS

The resulting processed speech has high quality sound both in quiet and in the presence of background speech babble. Figure 3 illustrates how the spectral structure is maintained.

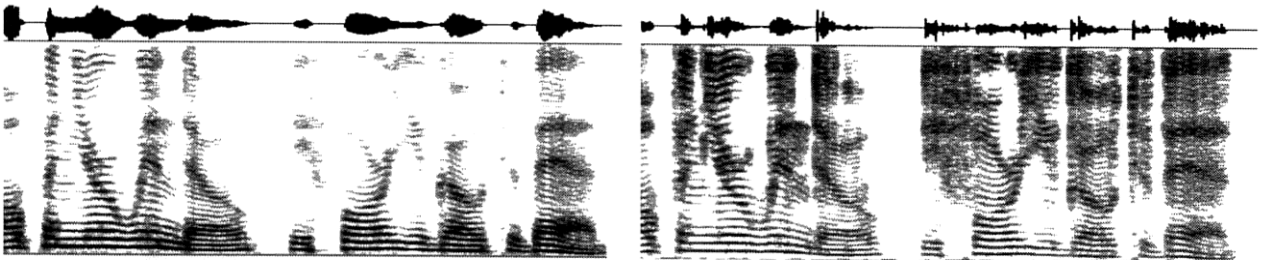


Figure 3. Spectrograms of original sentence (left panel) and processed sentence (right panel.)

Table 1: Results of subject testing. Mean % correct identification for 3 listeners with moderate hearing loss.

Stimuli	Easy noise condition	Moderate noise condition	Difficult noise condition
Original	68	58	41
Processed	91	62	54

Three young adult listeners with moderate, gently-sloping hearing losses participated in preliminary evaluations of the processing. Stimuli consisted of lists of ten CID Everyday Sentences, presented at varying signal-to-noise ratios. Sentences were presented monaurally to each listener's better ear at the listener's self-selected MCL via TDH-49 headphones. Attenuation levels for each sentence were adjusted individually based on the RMS amplitude of the sentence and the target presentation level. Lists consisted of sets of 10 sentences, and were randomized and counterbalanced to avoid list bias. Listeners wrote the sentences on a blank answer form. The answers were scored for number of key words correct by an experimenter who was blind to the conditions. Lax criteria were used for the scoring; errors of tense or number were scored as correct responses if the root word was correct. Listeners with moderate losses showed marked benefit from the processing for both sets of sentences. Results are shown in Table 1 for processed and unprocessed test sentences.

DISCUSSION AND CONCLUSIONS

The processing algorithm, which uses the sinusoidal model with a combination of compression and linear gain processing, produces high quality output speech that is within the residual dynamic range of the impaired listener and has strong resolution of the important spectral peaks. This hybrid processing operates on a time-varying basis to adjust to the characteristics of the speech in each frame. Pilot testing indicates improved performance by impaired listeners. Future plans call for extensive testing under a variety of listening conditions using the real-time system.

ACKNOWLEDGMENTS

This work was supported by the NIDCD (K08 DC00124), NSF, and the University of Malaga.

REFERENCES

1. Drake, L.A., Rutledge, J.C. and Cohen, J., *IEEE Trans. Sig. Proc.* **41**, pp. 3306-3312 (1993).
2. Rutledge, J.C. and Clements, M.A., "Compensating for recruitment of loudness using a sinusoidal model of speech." In *Proc. Int. Conf. Acoust., Spch. and Sig. Proc.*, 1991.
3. Rutledge, J.C., *Time Frequency and Wavelets in Biomedical Signal Processing*, (M. Akay, Ed). New York: IEEE Press, 1997, ch. 13.
4. McAulay, R.J. and Quatieri, T.F., *IEEE Trans. Acoust., Spch. and Sig. Proc.*, **34**, pp. 744-754, (1986).