ENHANCEMENT OF SPECTRAL CONTRAST IN SPEECH FOR HEARING IMPAIRED LISTENERS

C. M. Aguilera Munoz, * Peggy B.Nelson, ** and Janet C. Rutledge**

*University of Malaga, Campus de Teatinos, Complejo Tecnologico, 29080-Malaga, (Spain)
**Division of Otolaryngology-HNS, Univ. of Maryland Medical School, Baltimore, MD (USA)
*e-mail: aguilera@ctima.uma.es

ABSTRACT

People with hearing loss of cochlear origin experience difficulties perceiving speech in noise, and there is much evidence that their difficulties are due, in part, to the reduction of frequency selectivity that usually accompanies the hearing loss. Enhancement of the spectral peaks of the spectrum of speech can compensate for this reduced frequency resolution. Presented here is a processing algorithm based on a sinusoidal speech model, which result is a sharpened speech with enhancement of spectral peaks. This sharpening algorithm is adaptive to the shape of the spectrum of the phoneme. Preliminary listening tests indicate that listeners with moderate and greater hearing losses showed benefit from the processing algorithm through enhancement of the valleys between formants.

1. INTRODUCTION

A number of studies incorporating a relatively large number of subjects have shown a significant relation between frequency resolution abilities and speech recognition scores (e.g., Dreschler *et. al*. [7]; Ching *et. al*. [6]; Nelson and Revoile [10]; Revoile [12]). The negative effects of poor frequency resolution could be eliminated through an enhancement of speech spectral contrast.

Several recent studies have studied the enhancement of spectral peaks in speech to improve speech intelligibility for hearing-impaired listeners: e.g., Boers [2]; Summerfield *et. al*. [13]; Bustamante and Braida [4]. Although these studies have not demonstrated improvements with natural speech, several factors encourage further evaluation of contrast enhancement techniques (e.g., Bunnell [3]). Spectral enhancement systems that combine spectral shaping with amplitude modification have so far shown little benefits (Bunnell [3]). Baer and Moore [1] studied a scheme using a mathematical optimization procedure to enhance spectral contrast in order to produce a normal excitation pattern in a impaired ear, but this scheme failed to produce statistically significant improvements in intelligibility.

None of these methods use sinusoidal modeling however, and this method has a greater potential for enhancement than the studies cited above because it can give an infinite peak-to-valley ratios. Kates [8] studied a new form of speech processing based on sinusoidal modeling for enhancing speech intelligibility in noise. This scheme chooses either 8 or 16 sinusoids for the reproduced speech. The 16 mean-peak-selection procedure takes the 16 highest peaks of the magnitude spectrum. The 8 mean-peak-selection procedure uses the smoothed spectrum provided by a pseudo-cepstral spectral substration to choose the frequency bands in which the formants are found. The experimental results in consonant recognition depended on the type of consonant. For example, among the worst scores using this algorithm compared to the original stimuli scores were the fricatives /f/ and /θ/. Among the best scores using this algorithm compared to the original was the stop /k/. These results suggest that the algorithm for mean-peak-selection must be more adaptive to the spectrum of the phoneme. More peaks are needed for the phonemes with a flat spectrum, and fewer peaks are necessary for a compact spectrum.

Presented in this paper is a processing algorithm to enhance the spectral peaks in natural speech for hearing impaired people. This algorithm is based on a sinusoidal model using an adaptive algorithm for choosing the peaks from the spectrum of the signal.

2. PROCESSING ALGORITHM

The sinusoidal model developed by Quateri and McAulay [9] is an analysis/synthesis procedure in which the signal is reproduced as the sum of sinusoids with various amplitudes, frequencies and phases. The frequencies of the sinusoids in frame k are chosen to correspond to the N(k) local maximum largest peaks in the magnitude of the short-time Fourier transform of the speech signal. This analysis/synthesis procedure matches the frequencies between two frames. If a frequency in the frame k+1 has no frequency close to it in the k frame, it is a sinusoidal *'birth*'. That sinusoid is ramped up smoothly from 0. If one frequency in frame k has no close frequency in frame k+1, it is a sinusoidal *'death'*. That sinusoid is ramped down smoothly to 0. If there is a frequency in the frame k close to another frequency in the frame k+1, they are matched. Then, the amplitudes, frequencies and phases are interpolated between these two frequencies.

When these peaks have been calculated, the mean peak-selection algorithm chooses the frequencies that are supposed to be the formant frequencies. These frequencies are the *main frequencies*. This algorithm adapts to the spectrum of the phoneme. If the spectrum is flat more peaks are chosen than when the spectrum is compact. This algorithm chooses the *main frequencies* presumed to be the formant frequencies as the largest peaks among the local maxima in the magnitude of the peaks. If the *main frequencies* have frequencies very close to them, they are chosen as well to provide normal formant pitch and more natural sound to the signal. All the chosen frequencies (main and lateral frequencies) are the *total frequencies*. This algorithm uses a maximum of 8 *main frequencies* and a maximum of 15 *total frequencies*.

An example of this mean peak-selection process is shown in Figure 1. In this case, the number of chosen *main frequencies* is 5, and the number of chosen *total frequencies* is 9. In this example, the number of chosen frequencies (*mean and total frequencies*) is lower than the maximun number because is a compact spectrum. For a flat spectrum this mean peak selection algorithm choose the maximun number of frequencies (*mean and total frequencies*) to synthesize the output signal. These results confirm that this algorithm for mean-peak-selection is adaptive to the spectrum of the phoneme. More peaks are chosen for the phonemes with a flat spectrum, and fewer peaks are selected for a compact spectrum.

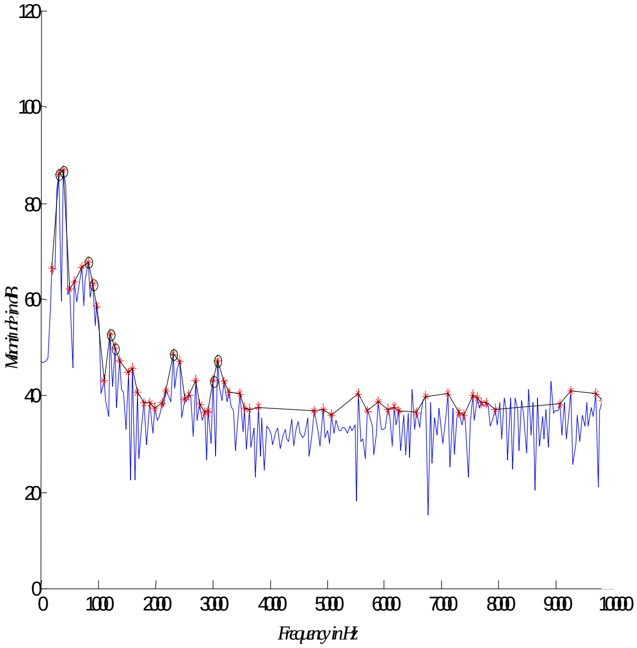


Figure 1: Speech spectrum indicating all the peaks (open symbols) and the mean peaks and the closer peak (filled symbols) for a compact spectrum.

The application here uses 7.5 msec analysis frames and 25.6 msec Hamming windows. A 512 point FFT is used to provide sufficient resolution for the speech sampled at 20 KHz. The initial number of peaks used is 60.

3. RESULTS

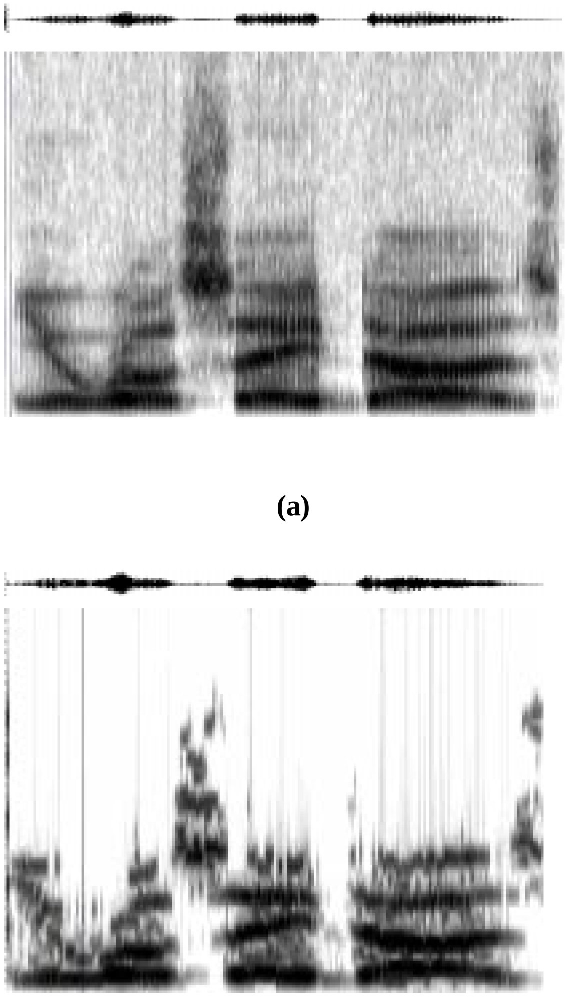
The resulting processed speech maintained the fundamental spectral and temporal pattern with a good sound quality. In Figure 2 is shown the spectrogram and time domain of the sharpened and the original signal. This illustrates how clearly the spectral structure is maintained in the processed speech.

Two listeners with severe, relatively flat hearing losses participated in the study. They were young adult and had losses of presumed cochlear origin. Stimuli consisted of the following list of 16 sentences each:

1. CID W-22 words in sentences, male talker, in quiet.
2. CID W-22 words in sentences, male talker speech, sharpened signal with a maximum bandwidth of 200Hz, with 10 dB S/N.
3. CID W-22 words in sentences, male talker speech, sharpened signal with a maximum bandwidth of 350Hz, with 10 dB S/N.

Sentences were presented 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. List consisted of sets of 16 sentences, and were randomized and counterbalanced to avoid list bias. The listeners were seated in a sound-treated chamber. They indicated the word they heard at the end of the sentence from a list of 8 choices. They indicated their choice via touch screen, and indicated their judgment of the quality of the signal (1 to 4). The answers were scored for number of key words correct by an experimenter who was blind to the conditions.

The results indicate that sharpening improved these listeners' understanding of voiceless consonants, while providing no significant change in identification of glide or fricative consonants. Subjects' rating of the quality of the sharpened stimuli indicate good acceptance of the algorithm. The better results were found with key words that contained /t/ and /p/ sounds using the 350 Hz bandwidth, although using 200 Hz bandwidth were better even than the original signal.



(a)

(b)

Figure 2: Speech spectrogram and time signal **(a)** original signal and **(b)** sharpened signal of the male's voice that says *"You will say does*".

4. DISCUSSION AND CONCLUSION

The processing algorithm uses the sinusoidal model to enhance the spectral peaks of the spectrum of the speech. This enhancement is adaptive to the phoneme processed, not only for choosing the main-frequencies supposed to be the formant frequencies but also changing the number of these main-frequencies. So, the algorithm chooses more peaks to synthesize a flat spectrum than for a compact spectrum.

Although peak detection is poorer for medium-high frequency than for a medium-low frequency for hearing-impaired listeners (e.g., Nelson *et. al.* [11]), introducing more peaks for a flat spectrum does not reduce the intelligibility. The reason is that the chosen peaks are more separated than for a compact spectrum.

The signal processed here lends itself well to real-time processing because the number of peaks used to synthesize the output signal is less than for the unsharpened signal. Future plans call for more extensive testing under a variety of listening conditions and with more hearing-impaired listeners.

5. REFERENCES

[1] Baer, T., and Moore, B. C. J., *Spectral enhancement to compensate for reduced frequency selectivity,* J. Acoust. Soc. Am., 95, 2992. 1994.

[2] Boers, P. M., *Formant enhancement of speech for listeners with sensorineural hearing loss* in IPO Annual Progress Report, No. 15 (Institut voor Perceptie Onderzoek, The Netherlands), pp. 21-28. 1980.

[3] Bunnell, H. T., *On enhancement of spectral contrast in speech for hearing-impaired listeners*, J. Acoust. Soc. Am., 88, 2546-2556. 1990.

[4] Bustamante, D. K., and Braida, L. D., *Wideband compression and spectral sharpening for hearing-impaired listeners*, J. Acoust. Soc. Am., Suppl. 1, 80, S12-S13. 1986.

[5] Bustamante, D. K., and Braida, L. D., *Principal-component amplitude compression for the hearing impaired*, J. Acoust. Soc. Am., 82, 1227-1242. 1987.

[6] Ching, T. *et. al., Prediction of speech recognition from audibility and psychoacoustic abilities of hearing-impaired listeners*, in Modeling Sensorineural Hearing Loss, Jesteadt, W., Ed., Lawrence Erlbaum Associates, Publisher, Mahwah, NJ, 433-445. 1997.

[7] Dreschler, W.A. *et. al., Relations between psychophysical data and speech perception for hearing-impaired subjects* II, J. Acoust. Soc. Am., 78, 1261–1270. 1985.

[8] Kates, J. M., *Speech enhancement based on sinusoidal model*, J. Speech Hearing Research, 37 (2), 449-464. 1994.

[9] McAulay, R. J., and Quateri, T. F., *Speech analysis/synthesis based on a sinusoidal Representation*, IEEE Trans. Acoust. Speech and Sig., ASSP-34, N 4, pag. 744-754. 1986.

[10] Nelson, P., and Revoile, S., *Factors affecting identification of stop and glide consonants by listeners with moderate and severe hearing loss*, J. Speech Lang. Hear. Res. 1998. (Submitted).

[11] Nelson, P., and Revoile, S., *Detection of spectral peaks in noise: Effects of hearing loss and frequency regions*, J. Acoust. Soc. Am., 1998. (Submitted).

[12] Revoile, S., *et. al., Some auditory indices related to consonant and vowel transition use by hard of hearing listeners*, J. Acoust. Soc. Am., 1998. (Submitted).

[13] Summerfield, Q., *et. al., Influences of formant bandwith and auditory frequency selectivity on identification of place of articulation of place of articulation in stop consonant,* Speech Communication, 4, 213-229. 1985.