THE ANALYSIS OF SPEECH CODECS USING PSYCHOACOUSTIC MEASURES

Mohammed Raad, Christian Ritz, Ian Burnett and Alfred Mertins

School of Electrical, Computer and Telecommunications Engineering
University Of Wollongong
Northfields Ave Wollongong NSW 2522 Australia
email: mr10@uow.edu.au

ABSTRACT

This paper analyses two narrowband speech codecs, the 4.8 kbps FS1016 coder and the 8 kbps G729 coder, using objective psychoacoustic measures. Four measures are used; Loudness, Sharpness, Roughness and Tonality. The results show Sharpness and Roughness as the two major contributing factors to the subjective difference between the two coders.

1. INTRODUCTION

In recent years objective measures of speech and audio quality have been considered with increased interest due to the standardization of PEAQ (Perceptual Evaluation of Audio Quality) by the ITU [1]. PEAQ is built around psychoacoustic characteristics of the human auditory system such as those presented in [2]. The objective quality of the audio signal is calculated through the combination of a number of psychoacoustic factors and models into a single value. This has been shown to correlate acceptably well with the subjective measures currently used.

The most widely used subjective measure of speech and audio quality is the Mean Opinion Score (MOS) [3]. It has been shown, during the ITU standardization process, that the proposed measure (PEAQ) can predict the MOS of a speech or audio signal well, but PEAQ has been presented as a technique for measuring perceptual speech and audio quality "on-line" [1]. That is, in situations where it is difficult or impossible to organize a subjective test such as the determination of the speech quality being delivered to a cellular telephone customer.

A possible extension of "on-line" PEAQ would be the introduction of objective quality measures in a feedback system to improve the perceived quality of a coded speech or audio signal. However, this possibility is hindered by the computational complexity of acceptable objective measures. Another approach would be to use some of the individual psychoacoustic factors upon which the objective measures such as PEAQ have been based. These psychoacoustic factors include Loudness, Sharpness, Roughness and Tonality of a sound; these each contribute in a unique manner to the final perceived quality [2].

Most of these psychoacoustic factors have been mathematically modelled in [2]. These models can be utilized to analyze a compressed signal in a psychoacoustic manner. This allows the identification of factors that need to be addressed to improve overall pleasantness and hence improve the perceived sound quality.

This paper presents a psychoacoustic analysis (using the above factors) of two well known speech codecs; The FS1016 coder [4] and the G729 coder [5]. The paper has been divided into three main sections; Section 2 introduces the objective models of the mentioned psychoacoustic measures, Section 3 presents the analysis method used and the results obtained and Section 4 provides a detailed discussion of the significance of the obtained results.

2. THE PSYCHOACOUSTIC MEASURES

2.1. Loudness

Loudness is measured in units of "phon" and is a relative measure indicating Sound Pressure Level (SPL) of a 1 kHz signal that would sound as loud as the given sound. Loudness is a sensation that is developed by the hearing system, that is, a sound incident to the hearing system will not result in instantaneous "loudness" perception. Instead the human auditory system needs time to develop the loudness of the incident signal and if this process is interrupted by another sound before the loudness sensation has been developed the earlier sound may not be heard, depending on its level. In [2], the loudness is modelled by the following equation:

$$N = \int_0^{24} N' dz \tag{1}$$

where N is the Loudness, N' is the loudness in the given critical band (called the "Specific Loudness" and measured in units of sone/bark) and dz is the increment in the critical band scale or Bark scale. The specific loudness is related to the excitation of the hearing system (E) by the sound in the frequency domain through the following equation:

$$N' = 0.08 \left(\frac{E_{TQ}}{E_0}\right) \left[\left(0.5 + 0.5\frac{E}{E_{TQ}}\right)^{0.23} - 1\right] \text{Sone/Bark} \tag{2}$$

Where E_{TQ} is the excitation at the threshold in quite, E_0 is the excitation as related to a reference intensity of $I_0 = 10^{-12}\ W/m^2$ and E is the excitation of the sound of interest.

0-7803-7549-1/02/$17.00 ©2002 IEEE.

2.2. Sharpness

Sharpness may be viewed as a measure of the density of loudness across the spectrum in different critical bands. Sharpness is most heavily influenced by the center frequency of the sound as well as the spectral content [2]. Sharpness (measured in *acums*) increases for sounds with greater loudness spread across more critical bands. Thus, if the spectral envelope (which ultimately determines the loudness) has significant spectral spread across a large number of critical bands then the sound will be sharp. In order to model this effect, Zwicker and Fastel in [2] proposed the following simple equation:

$$S = 0.11 \frac{\int_0^{24} N' g(z) z dz}{\int_0^{24} N' dz} \tag{3}$$

Where S is the sharpness, z is the bark scale value of the band and $g(z)$ is a weighting function.

2.3. Roughness

Roughness describes the inability of the ear to distinguish tonal components in a given signal e.g. a sound that is primarily noise-like would be "rough". The model proposed in [2] for Roughness is based on the assumption that the hearing system is only capable of detecting changes in excitation as given by:

$$R = 0.3 \frac{f_{mod}}{kHz} \int_0^{24} \frac{\Delta L_E(z) dz}{dB/Bark} \tag{4}$$

In Equation (4), ΔL_E is the change in the sensation level in dB; this is different to the change in excitation level but may be calculated from it (see [2]). The term f_{mod} is the modulating frequency of the sound, where it has been assumed that an amplitude modulation model is sufficient to represent the sound.

2.4. Tonality

In [2] it is suggested that tonality must be judged subjectively as no appropriate model exists. It is noted, however, that tonality decreases with increasing critical band rate spread, that is as the sound becomes more noise like it becomes less tonal. It should be noted here that in some published literature, such as [6], the tonality of the sound is approximated by using the Spectral Flatness Measure (SFM)[6]. In [6] it is suggested that as the SFM increases, the tonality decreases which matches what is reported in [2]. Hence, the tonality in this analysis has been obtained by the use of the SFM.

3. ANALYSIS METHOD AND RESULTS

Ten files (five female and five male) of narrow band speech were used to compare the behaviour of the four factor models. These files were extracted from the ANDOSL database [7], resampled to 8 kHz and bandlimited between 300 Hz and 3.4 kHz. The models presented in the previous section were then used to calculate the mean Loudness, Sharpness, Roughness and Tonality of each speech file (the original, FS1016 and G729 coded speech). The results obtained for the compressed versions of each file have been normalized to the results of the original file. As such, a relative measure is obtained showing how the compression techniques tested compare to each other and the original.

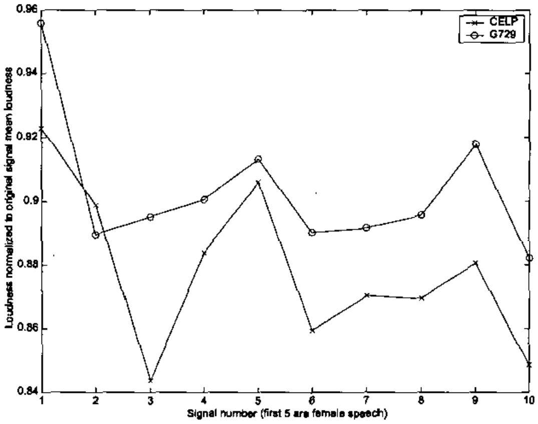

Figure 1: The relative mean loudness of the two coders

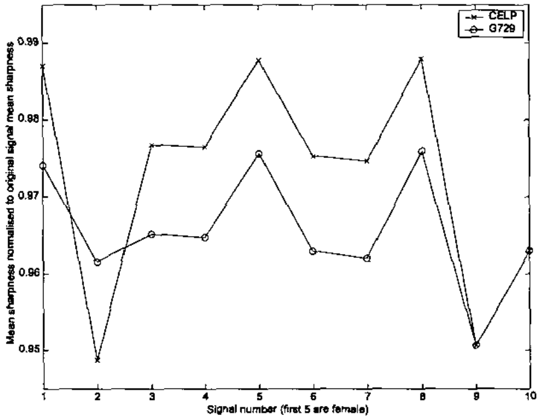

Figure 2: The relative mean sharpness of the two coders

Figures 1 and 2 present the normalized mean loudness and sharpness of the compressed files while Figures 3 and 4 present the roughness and tonality values. It can be seen from Figure 1 that the G729 coder results in a louder synthesized sound than the FS1016 coder while Figure 3 indicates that the FS1016 coder produces the rougher sound, Figure 2 shows that the FS1016 coder also produces a sharper sound than the G729 coder. Finally, Figure 4 shows that the tonality of the synthesized sounds varies but tends to be higher than the tonality of the original sound.

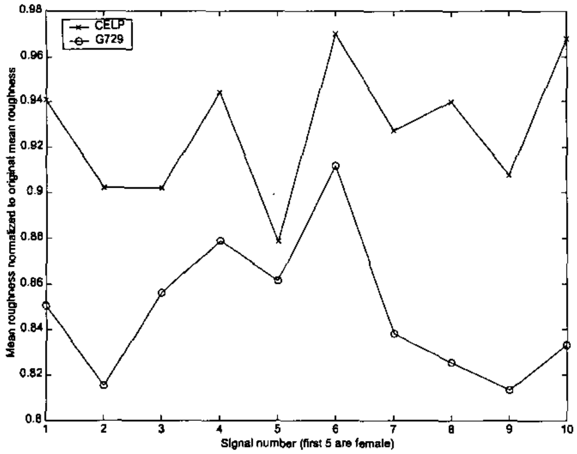

Figure 3: The relative mean roughness of the two coders

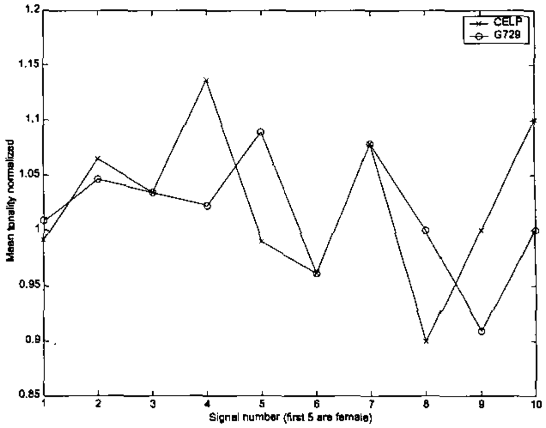

Figure 4: The relative mean tonality of the two coders

4. DISCUSSION AND CONCLUSION

The results of the previous section show that the FS1016 coder is rougher and sharper than the G729 coder. According to [2], as the roughness of a sound increases, the pleasantness decreases and similarly as the sharpness increases pleasantness again decreases. The results are in line with subjective results that suggest that the G729 coder produces synthesized speech of higher perceptual quality than the FS1016 coder [8]. On the other hand, the loudness of the G729 coder is higher than that of the FS1016 coder which suggests that the G729 coder is less pleasant than the FS1016 coder. It should be noted that the loudness results presented in [2] show a considerable amount of scatter and the model presented is less accurate than the models presented for sharpness and roughness.

The inconclusiveness of the tonality result can be simply explained as a direct result of the fact that both coders utilize linear prediction and post filtering as the basis of speech coding. Linear prediction, by the use of an all-pole model generally increases the tonality of a signal and this is deliberately enhanced further by post filtering, hence the tonality of the synthesized speech appears to be higher than the original speech. Significantly, in a consistent manner between the speech files similar curve shapes result.

In summary, this paper has analysed the performance of two widely used LP speech coders using psychoacoustically based models. The constituent measures of roughness and sharpness were found to be reliable and potentially useful indicators of narrowband coder performance. We further propose that these two measures could be usefully employed to improve coded speech quality in e.g. an Analysis-by-Synthesis coding scheme.

5. ACKNOWLEDGEMENT

Mohammed Raad is in receipt of an Australian Postgraduate Award (industry) and a Motorola (Australia) Partnerships in Research Grant. Christian Ritz is in receipt of an Australian Postgraduate Award.

6. REFERENCES

[1] T. Thiede *et al*, "PEAQ- the ITU standard for objective measurement of perceived audio quality," *Journal of the Audio Engineering Society*, vol. 48, no. 1/2, pp. 3–29, January/February 2000.

[2] E. Zwicker and H. Fastel, *Psychoacoustics*, Springer-Verlag, Berlin, second edition, 1999.

[3] T. Ryden, "Using listening tests to assess audio codecs," in *Collected Papers on Digital Audio Bit Rate Reduction*, Neil Gilchrist and Christer Grewin, Eds., USA, 1996, pp. 115–125, Audio Engineering Society, Inc.

[4] T. E. Tremain J. P. Campbell, V. C. Welch, "An expandable error protected 4800bps CELP coder (us federal standard 4800 bps voice coder)," in *Proc. ICASSP'89*, USA, 1989, vol. 2, pp. 735–738.

[5] R. Salami *et al*, "Design and description of CS-ACELP: a toll quality 8 kb/s speech coder," *IEEE transactions on Speech and Audio Processing*, vol. 6, no. 2, pp. 116–130, march 1998.

[6] James D. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE Journal On Selected Areas In Communications*, vol. 6, no. 2, pp. 314-323, Feb. 1988.

[7] "Australian national database of spoken langauge (ANDOSL)," CD-ROM.

[8] R. V. Cox, "Speech coding standards," in *Speech coding and Synthesis*, W.B. Kleijn and K. K. Paliwal, Eds., Netherlands, 1995, pp. 49–78, Elsevier Science.