Effects of Speech Coders and Speech Disorders on Speech Quality and Intelligibility*

M. Price[1], D.G. Jamieson[1], V. Parsa[1] and L. Deng[2]
1. Hearing Health Care Research Unit, University of Western Ontario, London, ON, Canada N6G 1H1
2. Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada N2L 3G1

1. INTRODUCTION

Telecommunication providers employ speech coders to compress signals for more economical spoken language telecommunication. Such coder processing always introduces some distortion. To increase efficiency, speech coders make certain assumptions about speech input. These assumptions are violated for talkers who have certain voice and speech disorders. As a result, modern speech coding systems may interact with voice and/or speech disorders to degrade communication differentially for particular talkers. This study investigated the effect of five common speech coders on both the intelligibility and the voice quality of various disordered speech samples.

2. METHOD

Our database of disordered speech samples [1] was obtained from 26 voices. Samples consisted of 22 isolated consonants in the UWODFD[2] format (the /aCi/ environment), 15 isolated vowels in the /hVd/ environment, and two continuous speech samples consisting of the first two sentences of the Rainbow Passage [3]. Speech disorders sampled included dysarthria and multiple sclerosis, while voice disorders included hypernasality, breathiness, harshness, pitch problems, vocal fold granules and vocal fold paralysis. Most of the disordered voices exhibited a combination of these elements. Whenever possible, the full range of severity from mild to severe was represented for both voice and speech disorders. Twenty disordered and six normal talkers were sampled.

Five speech coders were investigated. GSM 6.10 (a standard European coder), CELP and LPC-10e (commonly used in North America), and two sub-band coders, a 16 kbps (SBC16) and a 32 kbps (SBC32) version. Each speech sample was processed through each of the five speech coders resulting in six conditions for each sampled item including the original. The isolated consonants and vowels were used to evaluate intelligibility while the Rainbow Passage was used to measure speech quality.

To evaluate the effect of the coders on speech intelligibility, 10 young adults with normal hearing were asked to identify the processed and unprocessed isolated consonant and vowel samples. To evaluate the effects of the coders on speech quality, 10 young adults with normal hearing were asked to rate the quality of each voice in each of the six conditions. Ratings were obtained using a set of 28 visual analog scale items including voice pitch, speed of speech, breathiness, harshness, tinniness, presence of extraneous sounds, acceptability, naturalness and overall disorderedness.

3. RESULTS AND DISCUSSION

Intelligibility was highest for the original samples, and declined in a consistent manner for the coders examined [cf figure 1], with GSM speech often as intelligible as the original, with CELP next, followed by LPC, SBC32 and SBC16. Coding scheme and talker often interacted, so that the magnitude of the differences observed between processing techniques was different for different talkers.

Voice quality was rated highest for the original unprocessed samples, and declined in a consistent manner for the coders examined, with GSM speech being sometimes rated as highly as the original, with CELP next, followed by LPC, SBC32 and SBC16. Coding scheme and talker interacted so that the magnitude of the differences observed between processing techniques was different for different talkers. These interactions were sometimes quite pronounced [cf figure 2].

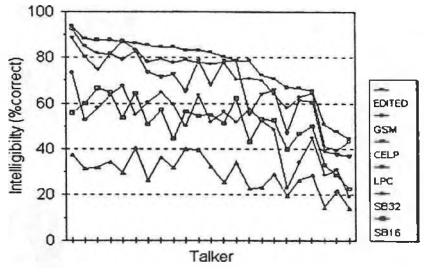

Figure 1. Intelligibility (in percent correct) on UWODFD.

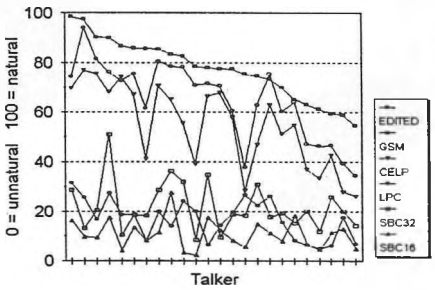

Figure 2. Voice quality ratings for the scale item "Naturalness".

These results have important implications for telecommunications. GSM and CELP appear to use coding schemes that degrade the input significantly less than LPC, or either of the subband coders. LPC, SBC32 and SBC16 substantially degrade the quality of the speech and therefore are not desirable for use in telecommunications. However, CELP and to a less extent GSM, *interact* with different voices so that degradation is differential.

4. CONCLUSIONS

Coded speech is less intelligible and perceived to be of lower quality than the natural speech of the same talker. Talker and coder interact to affect the intelligibility and perceived quality of speech at the output of the coder, so coder effects are much larger for some talkers than for others.

REFERENCES

[1] Jamieson, D.G., Deng, L., Price, M., Parsa, V. & Till, J. (1996). Interaction of speech disorders with speech coders. International Conference on Spoken Language Processing (ICSLP96) Philadelphia. 737-740.

[2] Cheesman, M. F. & Jamieson, D. G. (1996). Development, evaluation and scoring of a nonsense word test suitable for use with speakers of Canadian English. Canadian Acoustic, 24(1), 3-11.

[3] Fairbanks, G. (1960). Voice and articulation drillbook, 2nd ed. New York: Harper and Bros. 127.

*Works supported by a grant from NSERC to DJG & LD and another from the Barbershoppers to MP. Correspondence should be addressed to jamieson@audio.hhcru.uwo.ca.