Measurement of speech intelligibility based on the PESQ approach

John G. Beerends[1]*, Erik Larsen*[2] *, Nandini Iyer*[2§]*, Jeroen M. van Vugt*[1]

[1]TNO Telecom, The Netherlands
[2]University of Illinois at Urbana-Champaign, Beckman Inst. for Adv. Sci. and Techn., USA
j.g.beerends@telecom.tno.nl, elarsen@uiuc.edu,
niyer@uiuc.edu, j.m.vanvugt@telecom.tno.nl

Abstract

ITU recommendation P.862, PESQ was developed for measuring speech quality in a wide variety of conditions. This paper shows that PESQ can also be used for measuring speech intelligibility. An experiment was set up to evaluate speech intelligibility in the presence of two interfering talkers at 4 SNR conditions (0, -3, -6, and -9 dB), using 100 sentences in each condition, with five normal hearing subjects. Signals were processed by four different beamforming algorithms, as used in hearing devices. Two different presentations were used, diotic and dichotic. The correlation of P.862 with the intelligibility scores, which ranged from 9 to 98%, was 0.99 for the diotic experiment and about 0.91 for the dichotic experiment.

1. Introduction

In modern networks there is a trend to improve the end-to-end perceived speech quality by using noise suppression, either in the terminal or in the network. In general, these noise suppression schemes improve the quality regarding the background noise but tend to decrease the speech intelligibility.

PESQ ITU Recommendation P.862 [1], [2], was never validated in terms of speech intelligibility and, although there is a relation between speech quality and speech intelligibility, it is not clear that PESQ can be used to predict intelligibility. One should be aware that one can improve quality while decreasing intelligibility.

This paper presents a first attempt to see if PESQ can be used for measuring speech intelligibility. A database is used that was constructed for quantifying the differences in speech intelligibility for hearing aids when they are used in the presence of interfering talkers (two single-talker noise background situations). Experiments were conducted with diotic (identical signals at both ears) and dichotic (different signals at both ears) listening using four different beamforming algorithms that are designed to increase intelligibility by improving the signal-to-noise ratio (SNR) of the target signal.

2. Experiment

Subjects: 5 listeners (4 female, 1 male) with normal hearing participated in the experiment. All listeners had audiometric thresholds below 15 dB HL between 250-8000 Hz. All listeners were paid for their participation and were not familiar with the speech material used in the study.

Signal generation: Binaural room impulse responses (BRIR) were recorded from two hearing aid microphones (Phonak Claro 311 BTE), one mounted on each side of a Knowles Electronic Manikin for Acoustic Research (KEMAR). The BRIRs were recorded for azimuths of 0°, +30° and –90°; elevation was always fixed to 0°. These BRIRs were convolved with speech signals recorded in an anechoic room. The speech signals consisted of HINT [3] sentences spoken by a male talker, which served as target signal and were convolved with the BRIR at 0°. Continuous narrative spoken by 2 male and 2 female talkers were used as interfering signals, and were convolved with BRIRs at +30° and –90°. Thus, a simulation of an auditory scene

[§] Now with Air Force Research Laboratory, Wright-Patterson AFB, Dayton OH.

was created in which target sentences were presented from directly in front of the listener, and two maskers (chosen randomly for each trial and normalized for signal power) were presented at +30° and –90°. This auditory scene served as inputs to four types of beamforming algorithms - three variants of a frequency banded minimum variance beamformer (FMV [4], FMV_{IID}[1] and FMV_{ideal}) as well as a simple summing beamformer. As a control condition, the unprocessed signals (pass-through condition) were also included in the study (details about these beamformers, parameters used, as well as signal generation, are given in [5]). These experimental manipulations resulted in testing of two diotic conditions (FMV and summing beamformer), and three dichotic conditions (pass-through, FMV_{IID} and FMV_{ideal}).

Procedure: Listeners were presented with the sentences from the five described conditions. The order of presentation of the five conditions was varied for each listener. In addition, for each algorithm condition, listeners were presented with four different input SNRs (0, -3, -6, -9 dB). These input SNRs were generated by reducing the power levels of the two interfering signals, and were randomly presented within each listening block. Note that the SNR of the signals the listeners actually heard (output SNR) differs for each beamformers used. In each block, signals were presented to listeners seated in a sound-treated room over headphones (Sennheiser HDA 200). The listeners' task was to repeat the target sentence, while an experimenter scored the responses. Signal presentation as well as response tracking was controlled by a computer.

3. PESQ enhancement and results

As P.862 PESQ was developed for telephone band listening situations an update was made of the input filtering. The P.862 telephone band filter was replaced by a flat frequency response between 125 and 6000 Hz. Both diotic and dichotic databases were processed with this wide-band PESQ algorithm. As PESQ was developed as a monaural model, only the input signals to one ear were used for both databases.

The results are given in Figs. 1 and 2, and show very high correlation for the diotic presentation (r = 0.99), and high correlation for the dichotic presentation (r = 0.91). The correlation between PESQ and left/right ear showed a small difference for the dichotic conditions, 0.86 for left and 0.89 for the right ear.

The correlation of 0.91 for the dichotic condition was achieved by simply implementing the strategy that subjects use the highest SNR speech signal of either ear (the 'better ear'), see Fig. 2. As all conditions were obtained within a single experimental run, the regression lines of both experiments should coincide if there was no binaural gain. But in fact, another experiment with a very similar listening condition (two single talkers at +30° and -90°) showed that in this case binaural advantage (over better-ear only) is very large, about 9 dB [6], measured in terms of speech reception threshold. Figure 3 shows that there is a maximum binaural advantage of about 40% at equal-PESQ values.

The FMV_{IID} beamformer, which has a dichotic output, is designed to pass interferers to one ear only, in situations where there is a large level difference between both ears; otherwise signals are presented diotically. This may be the reason that the four data points for this particular beamformer coincide on the same regression as do the eight data points for the diotic conditions (Fig. 3).

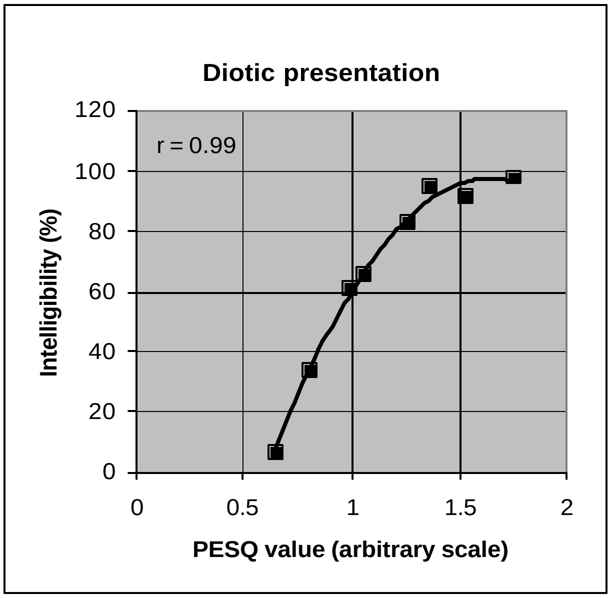

Figure 1. Relation between speech intelligibility and PESQ value in the case of diotic presentation. The scale is not calibrated (note that the original PESQ used band limiting between 300 and 3400 Hz in the training of the MOS scale).

[1] The FMV_{IID} was designed to make use of binaural cues such as interaural intensity differences (IID). In addition to implementing the FMV, the algorithm also computed the interaural intensity differences between the maskers. When the intensity difference between the two maskers exceeded a certain threshold (5 dB), the algorithm presented the masker to the ear

with the higher masker strength. When the difference in intensity was less than the threshold, the maskers were played diotically. Since the threshold criterion was applied to each frequency band, the resulting output was quasi-dichotic (signals were dichotically played in some frequency bands and diotic in others).

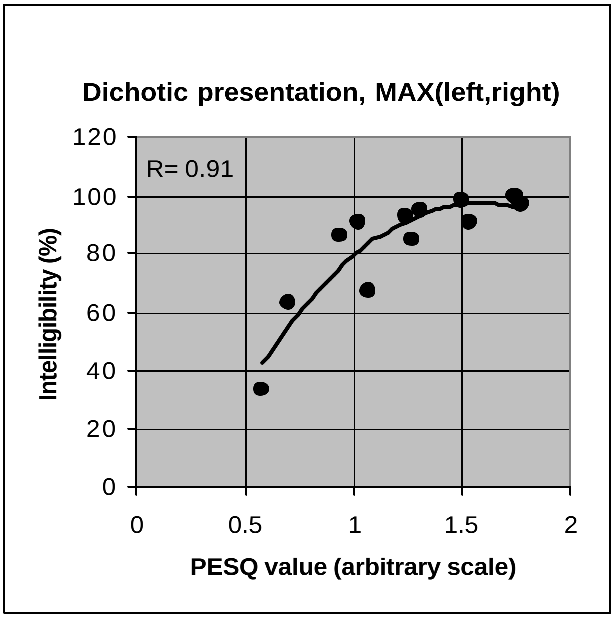

Figure 2. Relation between speech intelligibility and PESQ value in the case of dichotic presentation.

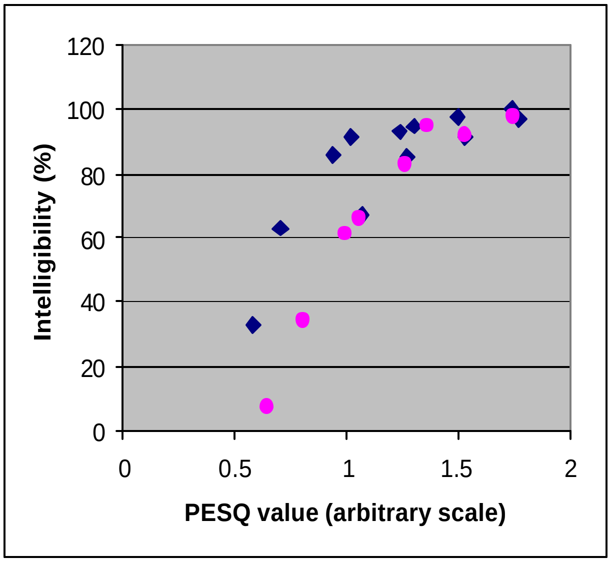

Figure 3. Relation between speech intelligibility and PESQ value in the case of dichotic ♦ and diotic ● presentation.

4. Conclusions and further research

The results show that PESQ can be used for assessing speech intelligibility in the case of two-talker interference. As the intelligibility in this experiment was measured for each condition as an average over 100 sentences, the exact details of the speech intelligibility need not to be modeled. Further research is needed to see whether the results also hold for the intelligibility of single sentences. Furthermore, there is a need to investigate if these results hold over a wide range of distortion conditions, e.g. as occur in mobile telecommunication networks. Assessment of intelligibility is getting increasingly important as noise suppression systems that are being implemented in handsets tend to decrease noise at the expense of intelligibility.

5. References

[1] A. W. Rix, M. P. Hollier, A. P. Hekstra and J. G. Beerends, “PESQ, the New ITU Standard for Objective Measurement of Perceived Speech Quality, Part 1 - Time Alignment,” J. Audio Eng. Soc., vol. 50, pp. 755-764, 2002.

[2] J. G. Beerends, A. P. Hekstra, A. W. Rix and M. P. Hollier, “PESQ, the New ITU Standard for Objective Measurement of Perceived Speech Quality, Part II - Perceptual model,” J. Audio Eng. Soc., vol. 50, pp. 765-778, 2002.

[3] M.J. Nilsson, J.D. Soli, and J. Sullivan, “Development of the Hearing in Noise Test for the Measurement of Speech Reception Thresholds”, J. Acoust. Soc. Am., vol. 95, pp. 1085-1099, 1994.

[4] M.E. Lockwood, D.L. Jones, R.C. Bilger, C.R. Lansing, W.D. O’Brien Jr., B.C. Wheeler, and A.S. Feng, “Performance of Time- and Frequency-Domain Binaural Beamformers Based on Recorded Signals from Real Rooms”, J. Acoust. Soc. Am., vol. 115, pp. 379-391, 2004.

[5] C.D. Schmitz and N. Iyer, “On the Reduction of Masking Effects while Preserving Competing Binaural Audio Streams”, Proc. 37th Asilomar Conf. Signals, Systems and Computers (Pacific Grove, CA), Nov. 2003.

[6] M.L. Hawley, R.Y. Litowsky, and J.F. Culling, “The Benefit of Binaural Hearing in a Cocktail Party: Effect of Location and Type of Interferer”, J. Acoust. Soc. Am., vol. 115, pp. 833-843, 2004.