On the detection of the intelligibility advantage of clear speech vs. casual speech

M. Koutsogiannaki[1], *C. Mayo*[2], *V. Kandia*[1] *and Y. Stylianou*[1]

[1]Institute of Computer Science, FORTH, and Multimedia Informatics Lab, CSD, UoC, Greece
[2] Centre for Speech Technology Research, the University of Edinburgh, UK
{mkoutsog, vkandia, yannis}@ics.forth.gr, catherin@ling.ed.ac.uk

1. Introduction

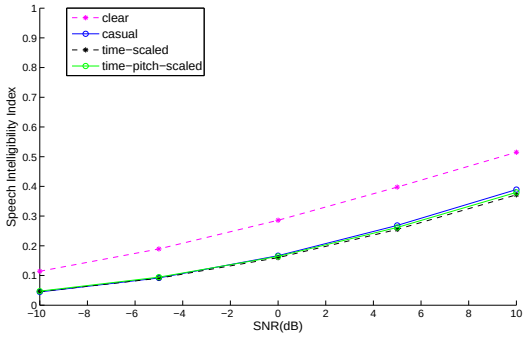
This work focuses on speaking rate and pitch differences between clear and casual signals. Transforming the clear signal to match the casual signal in terms of the prosodic features, the effect of each one factor to the intelligibility advantage of clear speech is examined based on acoustic evaluations and objective measures.

2. Methodology and results

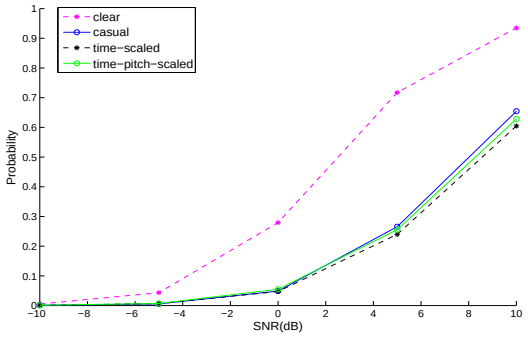
Studies show that even though intelligibility is increased by the decrease of speech rate both in clear and casual speech, clear speech can also be produced without the decrease of rate after training the speakers [1]. This suggests that clear speech has inherent acoustic properties independent of rate, that contribute to improved intelligibility. Many speakers in their effort to elicit clear speech change their pitch both in level and range. However, it is not luminous if pitch modification is a feature that contributes to intelligibility.

In this work, we examine if clear speech signals are still more comprehensive than casual speech signals after equalizing the prosody features on the two signals. To this purpose, a database of clear and casual speech signals is analyzed. Speakers in this database read sentences both in clear and casual way [2]. Clear speech sentences are modified in duration and pitch to match the corresponding attributes of casual speech signals. After the equalization, pilot acoustical test analysis and objective measure tests are performed on the four equal set of signals; on the initial database of clear and casual signals and additionally on the time-scaled and time and pitch-scaled clear signals.

In the acoustical pilot experiments, speech shaped noise is added to the signals to create the test signals, with Signal to Noise Ratio of $0dB$. Results show that on a set of pairs of clear and casual sentences, in 64% of the cases listeners found more intelligible the clear sentences. However, in time-scaled and time-pitch-scaled modified clear sentences intelligibility scores were deteriorated. Objective measure tests were also performed, using a modified version of the extended Speech Intelligibility Index (SII) [3]. SII was evaluated in a separate database giving high correlation scores with perceptual acoustical tests. According to the SSI measure, clear signals have higher intelligibility scores than casual signals (Fig.1(a)) with higher probability (Fig.1(b)) of identifying a sentence for SNR levels above $-5dB$. On the other hand, casual signals, time-scaled and time-pitch-scaled clear signals that have the same duration, give the same score of SII independent of the SNR level (Fig.1(a)). Pilot acoustical experiments and objective measures suggest that duration indeed plays a significant role to intelligibility, whereas pitch modifications do not seem to contribute to intelligibility.



(a)



(b)

Figure 1: Objective Measure Score for the four set of signals for different levels of SNR. a) Speech Intelligibility Index b) Probability of correctly identifying a sentence

3. References

[1] J. Krause and L. Braida, "Acoustic properties of naturally produced clear speech at normal speaking rates," *JASA*, vol. 115, no. 362-378, 2004.

[2] V. Hazan and R. Baker, "Does reading clearly produce the same acoustic-phonetic modifications as spontaneous speech in a clear speaking style?" *DiSS-LPSS*, pp. 7–10, 2010.

[3] K. S. Rherbergen and N. J. Versfeld, "Speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners," *JASA*, pp. 2181–2192, 2005.