The BEST (Beautifully Efficient Speech Test) for evaluating speech intelligibility in noise

Virginia Best, Margot McLelland, Harvey Dillon
National Acoustic Laboratories

Abstract

Efficient speech-in-noise tests are a critical tool in hearing research for the quick and reliable detection of changes in speech intelligibility (e.g. due to hearing aid processing). We describe the development of an Australian English sentence test (the BEST: Beautifully Efficient Speech Test). This test comprises a large set of naturally spoken open-set sentences, in which each morpheme can be scored. Since one proven way to achieve increased efficiency is to reduce variation amongst scored items, we attempted to equalise the intelligibility of each morpheme in the corpus. In an equalisation phase, the intelligibility of each morpheme in the corpus was compared using speech-in-noise data from a group of 40 normally hearing participants. Differences in intelligibility were compensated for on a morpheme-by-morpheme basis by applying continuous gain functions (range ± 6 dB) to each sentence. A validation phase involving 29 new listeners confirmed that the equalisation resulted in more uniform intelligibility thresholds across the corpus and steeper psychometric functions in individual participants. An evaluation phase on a third group of 14 listeners demonstrated that these psychometric changes translated into an adaptive speech test that was 30% more efficient than the original test.

Introduction

Equalisation of the intelligibility of sentences or words in a speech corpus has been used previously as a way of reducing variability and thus increasing the efficiency of a speech test.[1,2] To our knowledge, however, no attempt has previously been made to equalise at the word level within naturally spoken open-set sentences. The goal of this study was to create a highly efficient Australian English speech-in-noise test that uses open-set sentences and enables word-level scoring. This study is based on an earlier pilot study.[3]

Method and Results

- 20 lists (each comprising 16 sentences) in BKB-like[4] format spoken by a single male talker.
- Morphemic scoring - smallest part of a word with semantic meaning (e.g. “stairs” has 2 morphemes “stair” and “s”).
- Automated software package[5] measured performance using fixed SNRs (Phase 1 and 2) or adaptive tracking (Phase 3).

Loudspeakers arranged in a horizontal array

Speech at 0° azimuth

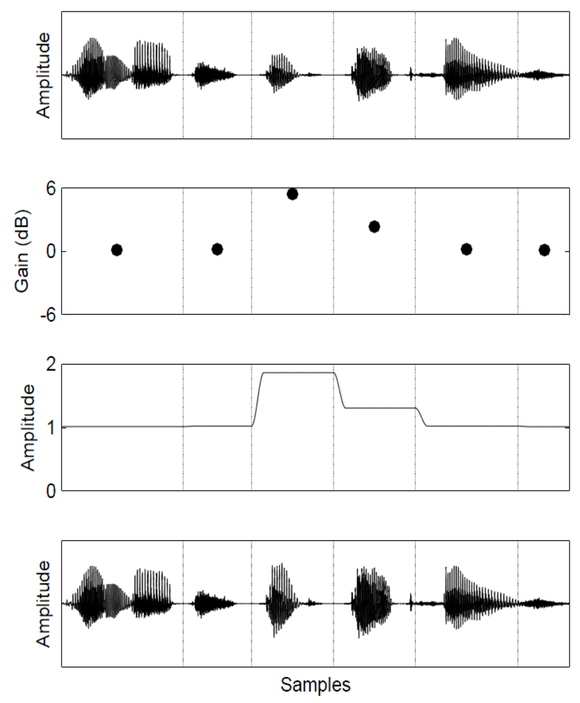
ILTASS shaped multitalker babble at ± 45° and ±135° azimuth

Phase 1 - Equalisation of original sentences (n = 40)

- 5 lists were presented at each of 4 SNRs (-4, -5.5, -7, -8.5 dB).
- Data was pooled across participants to generate psychometric functions for each morpheme.
- Thresholds (SNR for 50% correct) were extracted for every morpheme and compared to the mean across the set (-6.8 dB).

Sentence modification

- Individual gains were applied to each morpheme (Figure 1) based on how far its threshold deviated from the mean (e.g. if a morpheme was 2 dB easier than average its level was reduced by 2 dB).
- Gain values were capped at ± 6 dB to avoid large level variations impacting the natural sound of the sentence.



Original waveform, with dashed lines indicating the transition point between morphemes.

Each morpheme is associated with a gain value.

The gain value is used to create a continuous gain function.

After applying the gain function a new waveform is created.

Figure 1: An example of the sentence modification procedure

Phase 2 – Validation of modified sentences (n = 29)

- Identical procedure to Phase 1, but using modified sentences.
- Tighter clustering of thresholds in Phase 2 confirmed that the equalisation procedure reduced variation across *morphemes* (Figure 2).
- Reduced variance after equalisation was also seen across whole *sentences* and whole *lists.*
- Slope of the group mean psychometric function (Figure 3) increased from 12% (Phase 1) to 16% (Phase2).

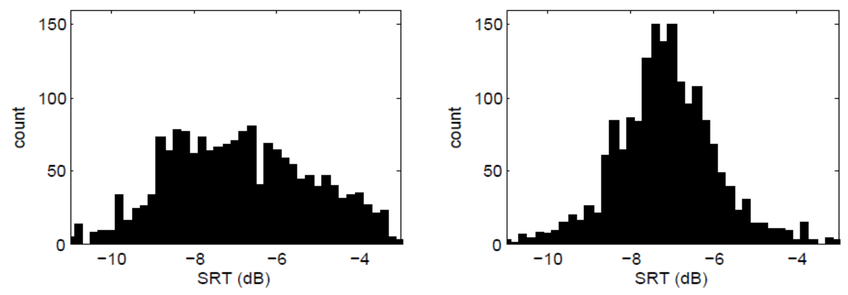


Figure 2: Histograms of speech reception thresholds measured in Phase 1 (left) and Phase 2 (right) for each morpheme.

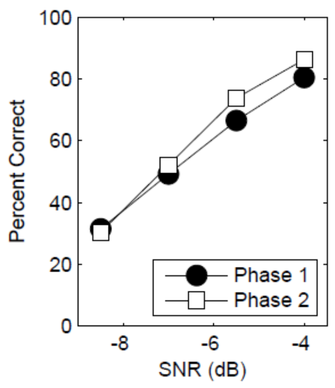


Figure 3: Psychometric functions for the pooled data in Phase 1 and Phase 2.

Phase 3 – Evaluation: Original vs. BEST (n = 14)

- Adaptive tracks using the original and the BEST sentences were compared (5 tracks per condition per subject).
- BEST test reached convergence in fewer trials (N) and produced a smaller standard error (SE) across reversals.
- A new efficiency measure that combines these two measures was defined: Efficiency = $1/(SE^2 \times N)$.
- Mean efficiency increased from 0.10 (original) to 0.14 (BEST), suggesting a gain in efficiency of around 30%.

Conclusions and Recommendations

This study describes a method for increasing the efficiency of a speech test by equalising intelligibility at the morpheme level and presents a new, highly efficient, open-set Australian English speech-in-noise test . The BEST may be a useful clinical tool for the rapid assessment of speech intelligibility in noise and for detecting subtle changes in intelligibility under different conditions.

References

1.Cameron S & Dillon H (2007). Development of the LiSN-S. *Ear and Hearing*, 28,196-211.
2.Wagener K et al (2003). Design optimization and evaluation of a Danish sentence test in noise. *Int. J Audiol*, 42, 10-17.
3.Schmitt N. (2004). A new speech test (BEST Test). Practical Training Report, NAL, Sydney.
4.Bench J & Bamford J.M. (eds) 1979. Speech-hearing tests and the spoken language of hearing-impaired children. London & New York: Academic Press
5.Keidser G et al (2013). An algorithm that administers adaptive speech-in-noise testing to a specified reliability at selectable points on the psychometric function. *Int. J Audiol*, 52, 795-800.