An acoustical engineer as a researcher in speech and hearing

Tammo Houtgast

VU University medical center Amsterdam, The Netherlands. Email:T.Houtgast@vumc.nl

Abstract

There is a long tradition of acoustical engineers working in the field of speech and hearing. Although it is well recognized that speech reception, and hearing in general, involves cognitive processes beyond the reach of the engineer, this engineering approach has led to interesting results which will be reviewed briefly.

For the engineer, speech is often described in acoustical terms as a succession of sounds with specific properties in terms of spectra and decibels. It has been possible to identify the acoustical cues in a speech signal which are essential for speech intelligibility. These cues are related to the dynamics, or the pattern of fluctuations, in the speech signal. By applying a specific type of speech analysis, the strength of these cues in an ongoing speech signal can be quantified. It will be shown that the effect of speech transmission and room acoustics on speech intelligibility can be estimated successfully by measuring the degree of preservation of these cues in the speech signal received by the listener.

In the field of hearing, the engineer is concerned mainly with the quality of the peripheral auditory system (i.e., the cochlea) in terms of accurate coding of the incoming sound. Sub-optimal performance (hearing impairment) may be related not only to a raised hearing threshold (hearing loss as determined by the classical tone audiogram), but also to a reduced quality of the coding process in terms of spectral and temporal resolution. The specific part of psycho-acoustical research concerned with this topic will be briefly reviewed: the design of appropriate test signals and measuring procedures to estimate the acuity in the auditory coding of the spectral, temporal and amplitude sound characteristics.

Introduction

This plenary lecture presents an impression of one example of a successful product of the engineering approach in the field of speech and hearing: the development of the Speech Transmission Index STI by Houtgast and Steeneken [1]. This index specifies the effect of the speech transmission from talker to listener on the speech intelligibility, and can either be calculated from design specifications, or actually measured in real conditions with noise, reverberation or distortion. Relevant information may be found in [2, 3, 4].

After introducing the rationale behind the STI development, the further presentation concentrates on two more recent issues which are strongly related to the STI concept. The first issue addresses the problem of a serious limitation of its applicability: the effect of non-linear processing on speech intelligibility, as in noise reduction algorithms, is not correctly accounted for by the STI. The second issue relates to the applicability of the STI for the hearing impaired.

The STI and non-linear signal processing

When speech is mixed with noise, one may try to reduce the noise and increase the speech-to-noise ratio by specific processing algorithms, with the purpose of increasing the intelligibility of the speech. One specific example is the noise reduction algorithm developed by Ephraim and Malah, known as spectral subtraction. It is a common observation that as a result of this type of operations, most listeners report a decrease of the noise, but well-controlled listening experiments do not indicate an improvement of speech intelligibility. However, STI-measurements do predict an improvement of intelligibility. So here is a conflict between STI predictions and actual intelligibility data. Recently, this apparent conflict was reconciled by introducing the concept of the signal-to-noise ratio in the modulation domain (S/N-mod) [5], which will be described briefly.

The key issue underlying the traditional STI-concept is illustrated in Figure 1. This historical picture (adapted from ref [2], 1972) presents the spectrum of the intensity envelope of octave-band filtered speech (the speech-envelope spectrum) along with the envelope spectrum of stationary Gaussian noise, reflecting the statistical characteristics of that noise. The area between the speech and the noise envelope spectra indicates what distinguishes speech from noise, in terms of modulations (the ‘useful modulation area’). This area, properly weighted and summed for the various octave bands, is the basis for the STI. The presence of noise and/or reverberation causes a reduction of the speech modulations, thereby reducing the useful modulation area and reducing the STI. It has been shown [5] that by spectral subtraction the reduced speech modulations are essentially restored, thereby improving the STI but, as indicated above, not improving intelligibility.

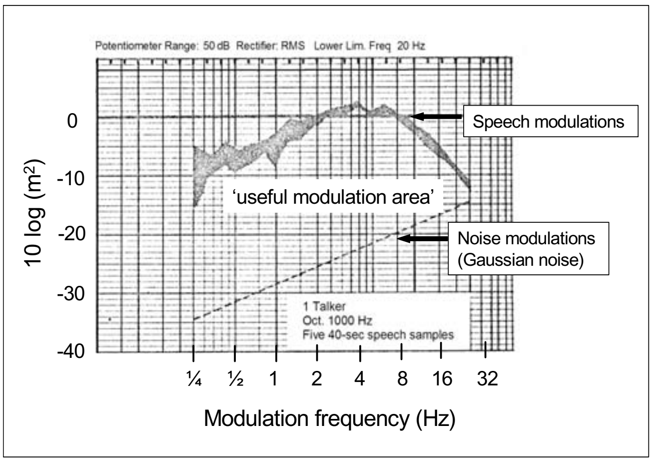

Figure 1: The envelope spectra for speech and for noise. The area enclosed by the two curves is crucial for the STI calculations.

The more recent approach [5] questions the implicit assumption that the lower limit of the useful modulation area, the noise modulations as indicated in Figure 1, is fixed and stable. This lower limit represents a floor of non-relevant or spurious modulations which partly originate from the instantaneous interactions of the speech and noise waveforms. Since that noise floor is (partly) caused by the speech-noise interactions, it cannot be measured simply from the noise-alone signal, but must be derived from the speech-plus-noise signal. For that purpose we developed a speech test signal for which in the (octave-band filtered) envelope all modulations in the 4-Hz region were removed. This provided a 'peephole' to study the underlying floor of spurious modulations in that region.

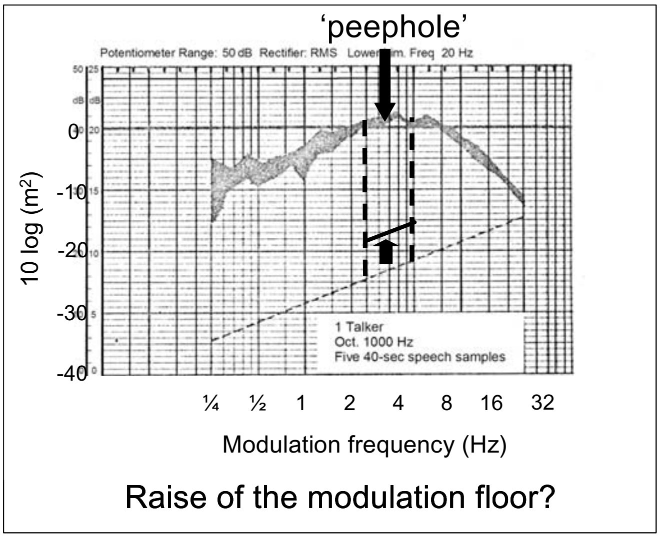

Figure 2: By removing some of the modulations in the speech envelope, that 'peephole' in the test signal can be used to study the underlying modulation floor.

Studies have shown [5] that the floor of spurious modulations is far from stable and invariant. For instance, in the case of spectral subtraction, where the speech modulations are restored, the noise floor moves upward accordingly. It was concluded that the *ratio* between the speech and the noise modulations is the crucial factor for speech intelligibility: the S/N-mod.

Subsequent studies confirmed the importance of S/N-mod in modeling the effect of noise-suppression algorithms on speech intelligibility. Figure 3 presents an example for 30 conditions of processing speech-plus-noise: the effect on the SRT in dB (the Speech Reception Threshold, see below) as predicted from the S/N-mod (abscissa), and the actually measured SRT's with listeners (ordinate). Keeping in mind that negative SRT-changes imply better intelligibility, it can be seen that improvements in SRT up to 2 dB are possible with this type of noise suppression algorithm [6].

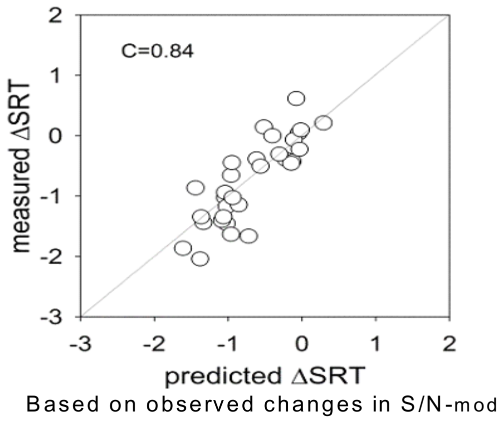

Figure 3: The relation between the predicted effect on the SRT, based on changes in S/N-mod, and actually measured changes in SRT's, for various conditions of signal processing applied to the speech-plus-noise signal.

The STI and the hearing impaired

In clinical practice speech intelligibility by an individual is often measured by the SRT in noise (the Speech Reception Threshold), see [7, 8]. It uses simple short sentences, and the SRT is the speech-to-noise ratio at which 50% of these sentences are understood correctly. The typical value for normal hearing listeners (using stationary noise with a spectrum shaped to the average speech spectrum) is an SRT of -4 to -5 dB.

With regard to the STI and the hearing impaired, there is good news as well as bad news. The good news is that on average, for groups of hearing impaired persons, the STI concept, in representing the combined effect of noise and reverberation, remains valid. This is illustrated in Figure 4, adapted from [3]. This figure displays iso-STI contours within the plane defined by the S/N ratio and the reverberation time T. It also displays data on the SRT measured for various T-values, for different groups of listeners, from [9].

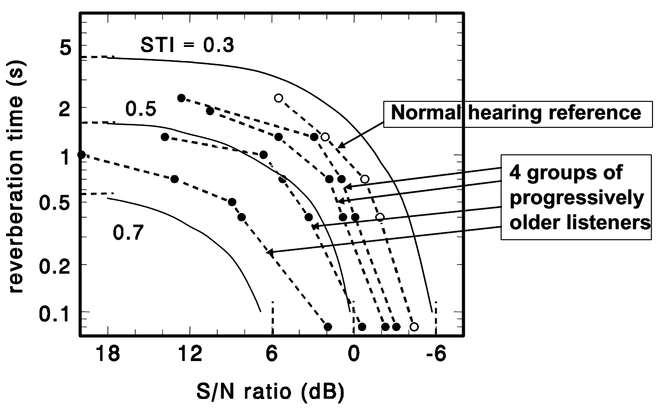

Figure 4: The data for normal-hearing as well as for groups of older, hearing impaired persons indicate the relevance of the iso-STI contours for all groups.

The normal-hearing SRT-data in Figure 4 follow the iso-STI contour of about STI=0.35 (for noise-only, this corresponds to the SRT of -4 to -5 dB mentioned before). The crucial point of Figure 4 is that also the data for groups of older (hearing-impaired) listeners essentially follow iso-STI contours, be it corresponding to higher STI-values. This means that hearing impaired need a higher STI to reach 50% intelligibility of simple sentences, and that the way in which the STI concept combines the effect of noise and reverberation remains valid (see [10] for recent additional information).

Beyond the engineering approach: Cognitive factors in top-down processes

It is well known that, on an individual level, there remain unexplained differences between the measured SRT and that predicted by the STI, or other related models, even when taking the hearing loss (pure-tone audiogram) into account. There is a long history of research to identify psycho-acoustic tests on possible deficits in auditory processing (e.g., tests on spectral or temporal processing) which may be relevant to explain the remaining differences. Still, the general opinion is that these strictly auditory-based bottom-up type of models fall short in fully explaining all inter-individual differences in speech reception in noise. There is growing awareness of the need to include the role of cognitive factors when modeling speech reception in noise for the older and/or hearing impaired listeners. References [11, 12] are among the first experimental studies related to this issue. For more recent studies and reviews in this field, see [13, 14, 15]. This topic was also addressed in a recent European-wide integrated project [16].

It is clear that the modeling of all factors relevant for speech reception in noise is far from complete. It also is clear that such modeling will certainly include aspects which are not the exclusive domain of the engineer, but requires the input from other disciplines as well. It is interesting to note that this growing awareness of the importance of top-down processes applies to sound perception in general, not only to speech reception. Reference [17] provides some interesting and stimulating reading on this issue.

References

[1] Houtgast, T. & Steeneken, H.J.M. A review of the MTF-concept in room acoustics and its use for estimating speech intelligibility in auditoria. J. Acoust. Soc. Am. **77** (1985), 1069-1077.

[2] Houtgast, T. & Steeneken, H.J.M. Envelope spectrum and intelligibility of speech in enclosures. Proceedings IEEE Speech Conference (1972), Newton, MA, 392-395.

[3] Houtgast, T., Steeneken, H.J.M. & Plomp, R. Predicting speech intelligibility in rooms from the Modulation Transfer Function. I. General Room Acoustics. Acustica **46** (1980), 60-72.

[4] Houtgast, T. & Steeneken, H.J.M. The roots of the STI approach. In: *Past, present and future of the Speech Transmission Index*, published by TNO, PO Box 23, 3769 ZG, Soesterberg, The Netherlands (2002), ISBN 90-76702-02-0.

[5] Dubbelboer, F. & Houtgast, T. The concept of the signal-to-noise ratio in the modulation domain and speech intelligibility. J. Acoust. Soc. Am. **124** (2008), 3937-3946.

[6] Dubbelboer, F. & Houtgast, T. The effect of varying the signal-to-noise ratio in the modulation domain on speech intelligibility in noise. Submitted to J. Acoust. Soc. Am. (2009).

[7] Plomp, R., & Mimpen, A.M. Improving the reliability of testing the speech reception threshold for sentences. Audiology **18** (1979), 43-52.

[8] Plomp, R., & Mimpen, A.M. Speech-reception threshold for sentences as a function of age and noise level. J. Acoust. Soc. Am. **66** (1979), 1333–1342.

[9] Duquesnoy, A.J. & Plomp, R. Effect of reverberation and noise on the intelligibility of sentences in case of presbyacusis. J. Acoust. Soc. Am. **68** (1980), 537-544.

[10] George, E.L.J., Goverts, S.T., Festen, J.M. & Houtgast, T. Measuring the effects of reverberation and noise on sentence intelligibility for hearing impaired listeners. Submitted to J. Acoust. Soc. Am. (2009).

[11] Van Rooij, J.C.G.M. & Plomp, R. Auditive and cognitive factors in speech perception by elderly listeners. III. Additional data and final discussion. J. Acoust. Soc. Am. **91** (1992), 1028-1033.

[12] Smoorenburg, G.F. Speech perception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram. J. Acoust. Soc. Am. **91** (1992), 421-437.

[13] Houtgast, T & Festen, J.M. On the auditory and cognitive functions that may explain an individual's elevation of the speech reception threshold in noise. International Journal of Audiology **47** (2008), 287-295.

[14] Kramer, S.E., Zekveld, A.A. & Houtgast, T. Measuring cognitive factors in speech comprehension: The value of using the Text Reception Threshold test as a visual equivalent of the SRT test. Submitted to the Scandinavian Journal of Psychology (2009).

[15] Zekveld, A.A., Kramer, S.E. & Houtgast, T. Cognitive factors in speech perception. These Proceedings (2009).

[16] EU 6th Framework, IST Integrated Project HearCom: Hearing in the Communication Society, 2004-2009.

[17] Plomp, R. The Intelligent Ear. On the Nature of Sound Perception. Lawrence Erlbaum Associates, Mahwah, NJ, 2002.