A METHOD FOR MEASURING MODULATION TRANSMISSION IN SPEECH TRANSMITTED VIA A NONLINEAR CHANNEL

Juha Backman

Nokia Mobile Phones
PO Box 86, FIN-24101, Salo, Finland

ABSTRACT

The paper presents a method for measuring the transmission of speech transmitted through a channel with linear or nonlinear distortion and arbitrary noise. The method is a generalization of the well-established method of measuring speech intelligibility using modulation transmission function, but instead of measuring only the amount of the modulation in the received signal and comparing it against the amount of modulation in the transmitted signals in given carrier and modulation frequency band, the proposed method cross-correlates the envelopes of the transmitted and received signal.

INTRODUCTION

In measuring the intelligibility of speech in a system with linear distortion (such as coloration and reverberation) and additive noise uncorrelated to speech a generally accepted method is to use a measure called Speech Transmission Index (STI) [1], derived from modulation transfer function [2]. In measuring STI (or its simplified form Rapid STI, or RASTI), the amount of modulation is either directly measured using a suitable test signal or derived from the impulse response and the measured amount of assumedly random noise.

This method, although successful for measuring e.g. loudspeaker installations of speech reinforcement systems in audioria etc., falls short when used for evaluating the intelligibility of speech transmitted through a nonlinear channel. Examples of commonly encountered nonlinear systems are most codec-based speech transmission systems (such as digital telephone systems), systems where the dynamics of the speech is altered (compressors, limiters, noise gates, etc., i.e. typical sound reinforcement systems as whole), and systems that have severe distortion, or noise correlated to the speech signal (such as modulation noise). An impulse-response based STI measure fails since there is no properly defined impulse response for a nonlinear system, and if a test signal with a spectrum and modulation roughly corresponding to that of speech is used, attempting to measure only the amount of modulation in the received signal will lead into a misleading results since the nonlinearities may either increase or decrease the amount of apparent modulation in a manner not corresponding the actual transmission of speech information.

MEASURING MODULATION IN NONLINEAR CHANNELS

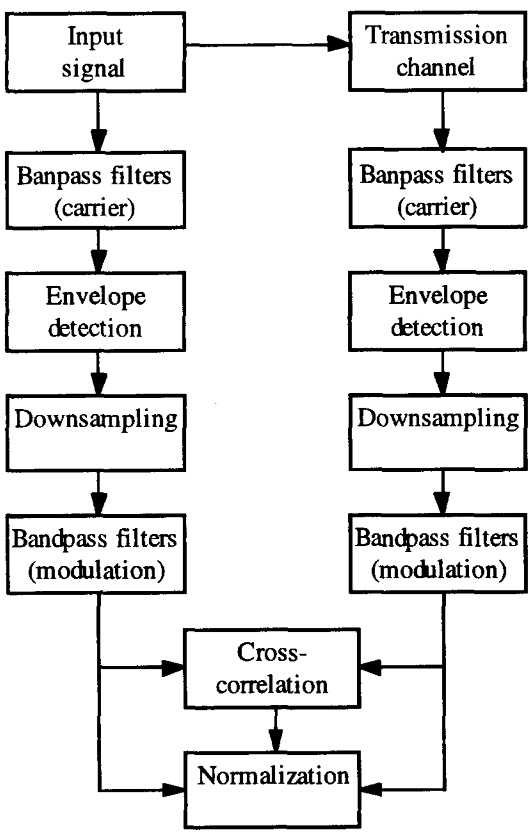
The method proposed in this paper is also a measure of the amount of modulation transmission. However, a measure of a nonlinear channel can only use the actual transmitted signal (such as speech) as its test signal, since otherwise the nonlinearities cannot be properly excited, or random errors, such as noise burst or breaks in transmission, cannot be observed in a proper relationship to the transmitted signal. The method that was found to be best suited for measuring the actual information transmission was normalized cross-correlation (i.e. coherence) of the envelopes of the transmitted and received signals.

Just as in measuring STI, in measuring modulation cross-correlation it is beneficial to divide both the speech signal (carrier) into a small number of frequency bands (usually one octave wide), and then again divide the modulation function into bands (usually one third-octave wide). This enables us to weight the various carrier and modulation bands according to the relative amount of both carrier signal and modulation in each band in average speech. In order to maintain compatibility with the established criteria carrier and modulation bands similar to those used in STI can also be used for measuring modulation cross-correlation.

To keep the criterion simple enough, instead of observing the entire cross-correlation function, an adequate indication of the modulation transmitted is the maximum of the cross-correlation in each carrier and modulation frequency band.

DESCRIPTION OF THE ALGORITHM

The algorithm can be described using the block diagram below.



Although the use of IIR filters in the carrier bandpass filters results in a computationally more efficient realization for given stopband attenuation, it is advisable to implement the filters either as equal-length FIRs or, if IIR implementation is attempted, they should be used as dual-pass linear-phase filters in order to keep the modulation envelopes of all the carrier and modulation bands time-aligned.

The envelope detection can be implemented through Hilbert transformation. As only the modulation below modulation frequencies of about 20 Hz is used in subsequent processing, the amount of data to be processed can be radically cut at this point by the downsampling.

If the system is realized correspond to RASTI measurements, then the signal is filtered with octave bandpass filters having center frequencies 500 Hz and 2 kHz. The envelopes of 500 Hz signal are analyzed in the third-octave bands of 1, 2, 4, and 8 kHz, and the envelopes of 2 kHz signal are analyzed using 0.7, 1.4, 2.8, and 11.2 Hz third-octave bands. The band-pass filtered envelopes of input and transmitted signal are then cross-correlated, and the maximum value of cross-correlation is used as an indicator of modulation transmission for each band. The total speech transmission can be then characterized by a weighted sum of these values.

EXAMPLE OF RESULTS

With linear systems (corresponding to reverberation), with or without uncorrelated additive wide-band noise the results correspond rather well (within ±10%) to the STI values computed using conventional methods. The differences in results can be ascribed to the variations of natural speech ignored by the conventional STI computation methods.

As a simple example of the results for nonlinear systems, values for speech intelligibility of half-wave and full-wave rectified speech are presented. Experimental results for such distortions were obtained already in late 1940s by Licklider [3]. His results indicated that half-wave rectification reduces intelligibility only by a small amount (per cent word articulation about 93%), but full-wave rectified speech is almost unintelligible (per cent word articulation about 8%). The results obtained using modulation cross-correlation and carrier and modulation bands described above correspond to STI values of 0.92 for half-wave rectified speech and 0.34 for full-wave rectified speech.

ADVANTAGES OF THE PROPOSED METHOD

- With linear distortion and uncorrelated noise results are equal to those obtained using the well-established STI measurements.
- Method is suitable to any speech transmission or coding system and for evaluating the degradation of intelligibility caused by any type of noise and distortion.
- The effects of different type of transmission disturbances can be objectively compared.

LIMITATIONS OF THE METHOD

- The method requires synchronized recording of the signal at both ends of the transmission chain. This limitation can be partially circumvented using a known recording with precise playback rate, enabling the measurement of physically long transmission chains, e.g. a complete telecommunications network.
- The method does not give in the form described above any indication of the coloration, perceived amount of noise, or the subjective pleasantness of the transmitted speech; however, extending the method to measure e.g. coloration is rather straightforward.
- Using current desktop computers, the method is relatively slow. The testing of the method was performed on a 33-MHz 68030-processor based computer, with code written using Matlab. In such an implementation the computation took about a hundred times real time, with speech signal sampled at 44.1 kHz, but using a signal processors a real-time implementation is feasible, especially if a lower sampling rate is used.

A common problem of all attempts of objective speech transmission measurement, shared also by this method, is that the effects of linguistic redundancy on speech perception are difficult to predict. Such effects arise for instance when speech is interrupted either by short pauses or bursts of noise [4] [5].

POSSIBLE FURTHER DEVELOPMENTS

The aim of the work was to develop a simple, single-number criterion for measuring speech transmission; however, it is instructive to take a look at the information possibly discarded when forming such a criterion.

In the discussion above, the magnitude of the correlation peak has been identified as a measure of the information transmission. The location of the correlation peak can be interpreted as a measure of time distortion of the signal envelope, which corresponds to a group-delay like measure for nonlinear systems.

The conventional STI uses a fixed set of weights, which are a function of carrier and modulation frequency. This leads into a slight language dependency, especially in non-European languages [6]. As the method proposed here measures also the modulation of the actual transmitted speech signal, a possible improvement over conventional measurement methods might be to use a set of weight adapted to the relative amounts of modulation actually present in the signal.

Apparently, the use of the proposed method is not limited only to evaluating speech transmission. The method can be used for evaluating the transmission of information also for other audio signals, such as music, and also for non-audio signals, such as video signals. For these purposes, the proposed method may prove to be useful for objective evaluation of otherwise elusive criteria, such as “clarity” or “definition”. However, great care must be taken if the use of the described measure is to be extended to these applications, since there are no established criteria for non-speech signals, and the exact formulation of the measurement must be entirely reconsidered. As the temporal and spectral structures of music vary widely (from, say, organ music recorded in a reverberant church, to drum solos), a measurement method that uses the actual input signal as a test signal and also possibly adjusts the weights according to the measured modulation contents of the input signal could enable the analysis of relevant parameters. Also, for these applications, the measurement of modulation transmission must be complemented by other, more traditional performance measurements, such as frequency response and (perceived) signal-to-noise ratio.

An interesting feature of the method is that it can be extended to analyze the temporal variation of modulation transmission and frequency response. As systems with time-variant characteristics are usually perceived as subjectively less pleasant than those with stationary characteristics, this result might provide clues to objectively measuring overall perceived quality, in addition to the information transmission capability.

REFERENCES

[1]H. J. M. Steeneken, T. Houtgaast: "A Physical Method for Measuring Speech-Transmission Quality", J. Acoust. Soc. Am., vol. 67, pp. 318 — 326, 1980.
[2] M. R. Schroeder, "Modulation Transfer Functions: Definition and Measurement", Acustica, vol. 49, pp. 180 — 182, 1981.
[3] J. C. R. Licklider, "Effects of Amplitude Distortion upon the Intelligibility of Speech", J. Acoust. Soc. Am., vol. 29, pp. 429 — 434, October 1946.
[4] T. Houtgaast, H. J. M. Steeneken, A. W. Bronkhorst, "Speech Communication in Noise with Strong Variations in the Spectral or the Temporal Domain", Proceedings of 14th International Congress on Acoustics, Beijing, China, vol. 3, paper H2-6, 1992.
[5] P. A. Howard-Jones, "Fluctuation of Noise Background: Measurement and Significance in Relation to Speech Masking", Proceedings of Eurospeech 91, Genova, Italy, vol. 3, pp. 1121 — 1124, 1991.
[6] Hao Shen, "Modulation transfer function of Chinese and its application", Proceedings of 14th International Congress on Acoustics, Beijing, China, vol. 3, paper G3-6, 1992.