Speech Enhancement: Reduction of Additive Noise in the Digital Processing of Speech

Project Proposal

Avner Halevy

Department of Mathematics

University of Maryland, College Park

ahalevy at math.umd.edu

Professor Radu Balan

Department of Mathematics

Center for Scientific Computation and Mathematical Modeling (CSCAMM)

University of Maryland, College Park

rvbalan at math.umd.edu

Abstract

This project will implement a few standard algorithms for reducing additive white noise in the processing of speech signals. Among these are spectral subtraction and iterative Wiener filtering. The performance of the algorithms will be evaluated on a database of IEEE sentences corrupted by several types of real-world noise. Initially only objective measures will be used to evaluate the quality of processed speech, but if time allows evaluation will be extended to include subjective listening tests as well.

Background

The need to enhance speech signals arises in many situations in which the speech signal originates from a noisy location or is degraded by noise over a communication channel. Speech enhancement algorithms can be used to enhance both quality and intelligibility of speech signals, thus making communication more effective and reducing listener fatigue. The precise goals of speech enhancement algorithms depend on the specific application, and the specific type of noise involved, as well as its statistical relation to the clean signal. The main challenge in designing effective speech enhancement algorithms is reducing noise without introducing perceptible distortion to the speech signal.

This project will focus on the reduction of additive white Gaussian noise which is statistically uncorrelated with the clean speech signal. We are assuming that y(n), the noisy signal, is composed of the clean speech signal x(n), and the additive noise signal, d(n), i.e. y(n) = x(n) + d(n).

If time allows, we will also explore the possibility of dealing with colored noise

Approach

Since speech signals are highly non stationary, short time Fourier analysis will be used for the analysis and syntheses of the signal, with frames typically consisting of 15-20 msec of speech, during which the properties of the signal are assumed not to change significantly. The discrete Fourier transform (DFT), computed via the fast Fourier transform (FFT), will be used for this purpose. The overlap and add method (OLA) will be used for reconstructing the enhanced signal.

Initially, two main algorithms will be implemented:

<u>Spectral subtraction</u>

This algorithm estimates the noise spectrum when speech is absent from the signal and subtracts it from the spectrum of the noisy signal to recover (an estimate of) the clean signal. The magnitude of the noise spectrum can be estimated by computing its average value during speech pauses, and the phase of the noise can be replaced by the phase of the noisy signal, which has been shown to be good enough for practical purposes. Precise upper bounds for the error thus introduced can be obtained. It has been shown that as long as the spectral SNR is larger than about 8 dB, the error will not be perceptible by the auditory system. Performing the simplest spectral subtraction may result in negative spectral magnitude components, but various methods exist for rectifying this situation.

The main challenge in designing these methods is to avoid introducing “musical noise”, which can be especially prominent in unvoiced segments, where SNR values are low.

As always, the simple subtraction comes at a price. If too much is subtracted, too much distortion will be introduced, whereas if too little is subtracted, too much noise will remain. This tradeoff will be explored and dealt with in several ways.

Iterative Wiener filtering

This algorithm constructs a linear time invariant (LTI) finite impulse response (FIR) filter which is optimal in the sense that it minimizes the mean square of the estimation error. The filter is constructed iteratively. At each iteration, given a previous estimate of the clean signal, linear prediction is used to estimate the parameters of a speech production model assumed to hold for the clean speech signal, and these are in turn used to update the filter and obtain a new estimate of the clean signal. The optimal number of iterations may vary with the characteristics of the clean signal.

Various parameters in these two algorithms may be adjusted in order to optimize performance for specific types of signals. Different values will be experimented with and the effects will be reported.

Additional algorithms:

If time allows, two other algorithms will be considered. The first is the MMSE estimator of the spectral magnitude proposed by Ephraim and Malah. The second is an SVD-based algorithm.

Implementation

The algorithms will be implemented on a standard PC using MATLAB. Computational complexity is not expected to require special resources or a need for parallelization.

Validation and Testing

A basic validation of the spectral subtraction algorithm implementation will be done by setting the estimate of the noise spectrum magnitude equal to zero, in which case the output signal should be the input signal. Validation of the linear prediction module of the Wiener filtering algorithm will be done by artificially constructing a speech production model with known parameters, in which case the module should predict very nearly the same parameters.

A noisy speech corpus (NOIZEUS) containing 30 sentences from an IEEE database corrupted by eight different real-world noises at different SNRs, will be used for testing the algorithms. This corpus is available to researchers free of charge, to facilitate comparison of speech enhancement algorithms developed by different research groups.

Performance of speech enhancement algorithms is usually judged in terms of quality and intelligibility of the enhanced speech. Quality measures may assess several different dimensions of the enhanced speech, and are concerned with *how* words were said. Intelligibility measures assess *what* words were said. Quality can be measured both subjectively and objectively. Subjective quality tests, as well as intelligibility tests, are usually highly time-consuming, and may possibly also require access to trained listeners. For this reason, in this project, evaluation of the algorithms will initially be confined to the use of objective measures.

The measures that will be used are called SNR, where we consider the energy of the clean signal as compared with the energy of the error in estimation. Two versions will be used: The first, performed in the time domain, is called *segmental SNR*, and is defined as

$$SNR_{seg} = \frac{10}{M}\sum_{m=0}^{M-1} \log_{10} \frac{\sum_{n=Nm}^{Nm+N-1} x^2(n)}{\sum_{n=Nm}^{Nm+N-1} (x(n) - \hat{x}(n))^2}$$

where *x(n)* is the clean signal, $\hat{x}(n)$ is the enhanced signal, *N* is the frame length, and *M* is the number of frames. The second, performed in the frequency domain, is called *frequency weighted segmental SNR*, and is defined as

$$fwSNR_{seg} = \frac{10}{M}\sum_{m=0}^{M-1} \frac{\sum_{j=1}^{K} B_j \log_{10}[\frac{E_c(m,j)}{E_e(m,j)}]}{\sum_{j=1}^{K} B_j}$$

where *K* is the number of frequency bands, B_j is the weight placed on the *j*th frequency band, $E_c(m,j)$ is the short term clean signal energy contained in the *j*th frequency band in the *m*th frame, and $E_e(m,j)$ is the analogous quantity for the energy of the error. $fwSNR_{seg}$ has been shown to perform moderately well in predicting subjective overall quality.

If time allows, we will also conduct a Mean Opinion Score test of quality, which is a widely used subjective quality measure. In this test listeners rate the signal on a five-point scale (1 corresponding to unsatisfactory, 5 to excellent) where the numerical value reflects the listener’s subjective impression of overall quality.

Schedule, Milestones and Deliverables

September:

Preliminary background reading and formulation of project scope

October:

9 – Research Proposal and Presentation

31 – Finish background reading

November:

Implementation of the spectral subtraction algorithm, including several variations

December

End of semester progress report and presentation

January

Spring break

February

Implementation of iterative Wiener filtering, including several variations

March

Testing, modification, and finalization of code

April

Prepare final report and presentation

May

Deliver final report and presentation

Bibliography

[1] Deller, J., Hansen, J., and Proakis, J. (2000) *Discrete Time Processing of Speech Signals*, New York, NY: Institute of Electrical and Electronics Engineers

[2] Quatieri, T. (2002) *Discrete Time Speech Signal Processing*, Upper Saddle River, NJ: Prentice Hall

[3] Loizou, P. (2007) *Speech Enhancement: Theory and Practice*, Boca Raton, FL: Taylor & Francis Group