Robust speech recognition using temporal masking and thresholding algorithm

Chanwoo Kim[1], *Kean K. Chin*[1], *Michiel Bacchiani*[1], *Richard M. Stern*[2]

Google, Mountain View CA 94043 USA [1]
Carnegie Mellon University, Pittsburgh PA 15213 USA [2]
{chanwcom, kkchin, michiel}@google.com, rms@cs.cmu.edu

Abstract

In this paper, we present a new dereverberation algorithm called Temporal Masking and Thresholding (TMT) to enhance the temporal spectra of spectral features for robust speech recognition in reverberant environments. This algorithm is motivated by the precedence effect and temporal masking of human auditory perception. This work is an improvement of our previous dereverberation work called Suppression of Slowly-varying components and the falling edge of the power envelope (SSF). The TMT algorithm uses a different mathematical model to characterize temporal masking and thresholding compared to the model that had been used to characterize the SSF algorithm. Specifically, the nonlinear highpass filtering used in the SSF algorithm has been replaced by a masking mechanism based on a combination of peak detection and dynamic thresholding. Speech recognition results show that the TMT algorithm provides superior recognition accuracy compared to other algorithms such as LTLSS, VTS, or SSF in reverberant environments.

Index Terms: Robust speech recognition, speech enhancement, reverberation, temporal masking, precedence effect

1. Introduction

$$\begin{aligned} y[n] &= x[n] * h[n] \\ \log(Y[m, e^{j\omega_k})) &= \log(X[m, e^{j\omega_k})) \\ &\quad + \log(H[m, e^{j\omega_k})) \end{aligned}$$

$$\begin{aligned} \log\left(\left|Z[m, e^{j\omega_k})\right|\right) &= \log\left(\left|Y[m, e^{j\omega_k})\right|\right) - \frac{1}{M}\log\left(\sum_{m=0}^{M-1}\left|Y[m, e^{j\omega_k})\right|\right) \\ &= \log\left(\left|X[m, e^{j\omega_k})\right|\right) - \frac{1}{M}\log\left(\sum_{m=0}^{M-1}\left|X[m, e^{j\omega_k})\right|\right) \end{aligned}$$

2. Conclusion

In this paper, we describe a new dereverberation algorithm, TMT, that is based on temporal enhancement by estimating the peak sound level and applying the temporal masking. In the experimental results, we have observed that TMT algorithm is simple, but has shown better speech recognition accuracies than existing algorithms like LTLSS or VTS. The matlab code for this algorithm may be found at `http://www.cs.cmu.edu/~robust/archive/algorithms/tmt`.

3. Acknowledgements

This research was supported by Google.