TECHNICAL DOCUMENTATION

cJMAP SINGLE MICROPHONE SPEECH ENHANCEMENT

Chandan K A Reddy, Issa Panahi, SSPRL UT Dallas

We developed a new single microphone SE technique based on *Joint Maximum A Posteriori* (JMAP) for convolutive mixtures, which inherently accounts for the effects of changes in acoustic path and reverberation without the need of pre or post filtering to counteract these effects. The developed method is designed for convolutive mixtures. Hence, we call this method convolutive JMAP (cJMAP). We introduce a parameter in the derived final gain function called the "trade-off" factor, which allows smartphone user to control the amount of noise suppression and speech distortion. This immensely helps hearing impaired in real-world scenarios when the environmental conditions change continuously. The user can personalize the trade-off parameter by varying according to the changing acoustical conditions and their comfort level of hearing.

In this work, we derive a new cJMAP amplitude and phase estimators by considering the new signal model of noisy speech given by,

$$y(n) = s(n) * h_1(n) + v(n) * h_2(n) \tag{1}$$

where $y(n)$ is a discrete time noisy speech composed of clean speech $s(n)$ and additive noise $v(n)$. $h_1(n)$ is the time domain impulse response between the speech source and the observation microphone and $h_2(n)$ is the impulse response between noise source and the observation microphone respectively. (*) is convolution operation. The final gain function is given below in (2). Please refer to our publications section http://www.utdallas.edu/ssprl/publications/ for further details,

$$G_k^{cJMAP} = \frac{\xi_k + \sqrt{\left(\xi_k^2 + 2\left(1 + \frac{\xi_k}{\beta_k}\right)\frac{\xi_k}{\gamma_k}\right)}}{2(\beta_k + \xi_k)} \tag{2}$$

Figure 1 describes the proposed method. The default microphone on the smartphone captures the signal and is transformed to frequency domain using windowing and Short Time Fourier transform (STFT). Once the noise starts, the "Trade-off" factor is adjusted to a desired value to strike a balance between noise suppression and speech distortion. The gain is adjusted based on the selected "trade-off" factor to obtain a value G_k, which is then multiplied to speech spectrum Y_k to obtain the estimate of the clean speech spectrum $\hat{S}_k$. $\hat{S}_k$ is then transformed to time-domain $\hat{s}(n)$ using synthesis process with Inverse Fast Fourier Transform (IFFT).

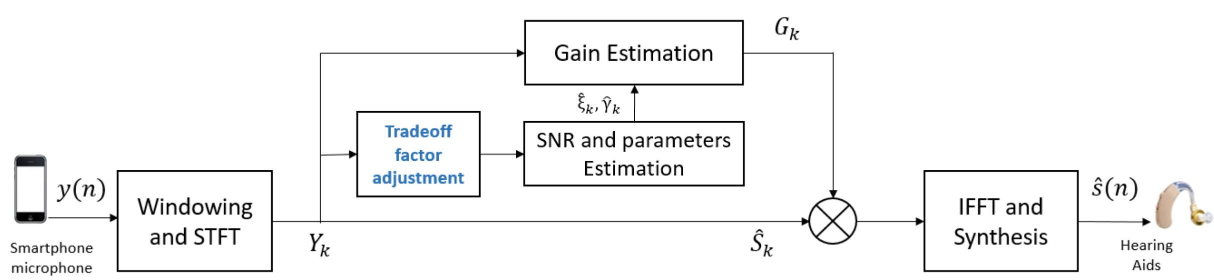
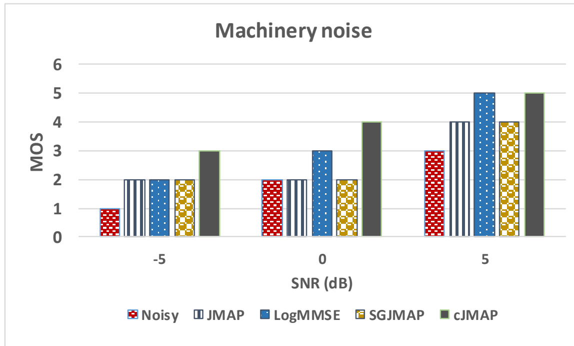
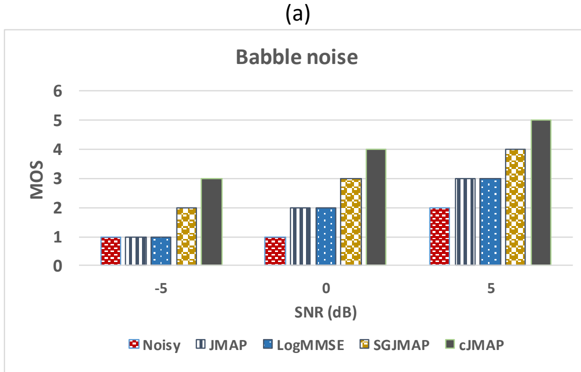


Fig. 1. Block diagram of the proposed SE method

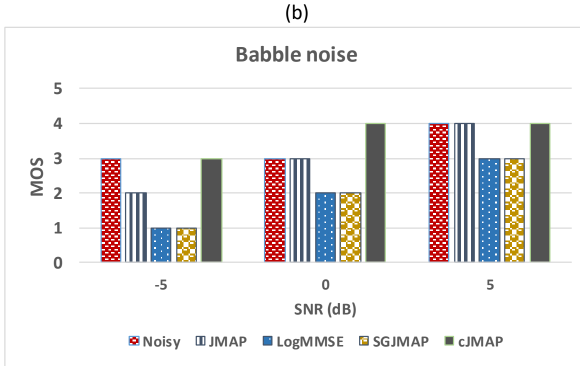
We compared our method with few other benchmark methods that are known to perform well in various noise conditions. The subjective results are shown in Figure 2. The results reflect the usefulness of the developed method in real-world noisy conditions. The clinical testing results can be found in http://www.utdallas.edu/ssprl/hearing-aid-project/clinical-testing/



(a)



(b)



(c)

Fig. 2. Mean Opinion Score (MOS) test scores for 15 normal hearing subjects