Ming Hsieh Department of Electrical Engineering

Multistream Diarization Fusion Using the Minimum Variance Bayesian Information Criterion

Tae Jin Park, Panayiotis Georgiou
SIPI / SCUBA

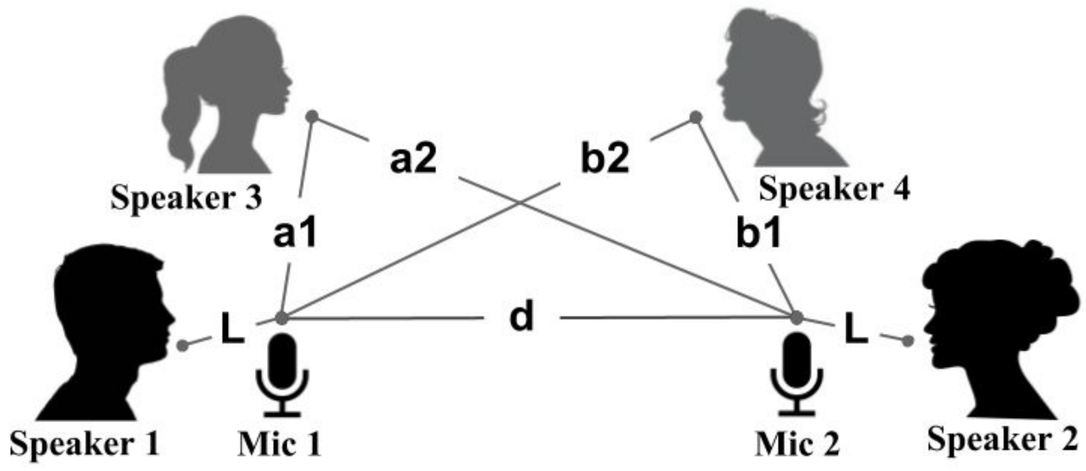
USC Viterbi School of Engineering

1. Introduction

- We present a method of improved fusion over multiple diarization streams.
- We introduce a new database that realistically simulates a range of extremely challenging acoustic conditions.
- We also propose a Minimum Variance of BIC (MVBIC) method to combine information from the various diarization streams.
- Our proposed method exploits the complementarity of the individual diarization streams and outperforms static fusion mixing weights.
- In a real scenario, where people are going about their daily lives, interacting with third parties, and under variable acoustic conditions and relative locations, the diarization task becomes more difficult.

2. The creation of a dataset

- USCDiarLibri2,4 Database:



- USCDiarLibri2,4 assumes 2 speakers of interest among 4 active speakers.
- It models reverberation, overlap, and interfering sources.

3. MVBIC

- We propose the Minimum Variance of BIC (MVBIC) technique that efficiently weights BIC distances according to their reliability towards improved clustering accuracy.
- We assume that there is an underlying correct BIC stream that we are observing through a noisy channel. The hidden, correct BIC stream will be represented by b and its two observed, noisy versions by b̃i.

$$\tilde{b}_i = b + n_i \tag{1}$$

- Obtain the optimal fusion weights that will lead to accurate estimation of the true b value:

$$\widehat{b} = \sum_{i=1}^{M} \omega_i \, b_i = \mathbf{w}^T \mathbf{b}$$
$$Var\left[\widehat{b}\right] = \mathbf{w}^T \Sigma_b \mathbf{w} \tag{2}$$

- An assumption that the noise random variable is mean zero and the two noise streams are uncorrelated:

$$\sigma_{b,i}^2 = \sigma^2 + \sigma_{n,i}^2$$
$$\sigma_{b,ij} = \sigma_{b,ji} = \sigma^2 \tag{3}$$
$$\text{where } i \neq j \text{ and } i, j \in [1, M]$$

- Minimization problem:

$$\begin{aligned} \textbf{Minimize:} \quad & Var\left[\widehat{b}\right] = \mathbf{w}^T \Sigma_b \mathbf{w} \\ \textbf{Subject to:} \quad & \mathbf{w}^T \mathbf{1} = 1 \end{aligned} \tag{4}$$

- The solution to the equation would be given as below:

$$\widehat{\mathbf{w}} = \frac{\Sigma_b^{-1} \mathbf{1}}{\mathbf{1}^T \Sigma_b^{-1} \mathbf{1}} \tag{5}$$

4. Experimental Results

- MVBIC keeps the DER lower than the single feature diarization methods regardless of the location of the interfering speakers.
- We observe that the MVBIC method approaches the optimize-on-test-set performance of the static weight.

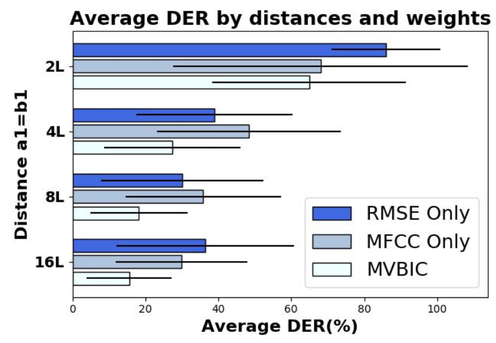


Fig. 1. Average DER by distances of interfering speakers

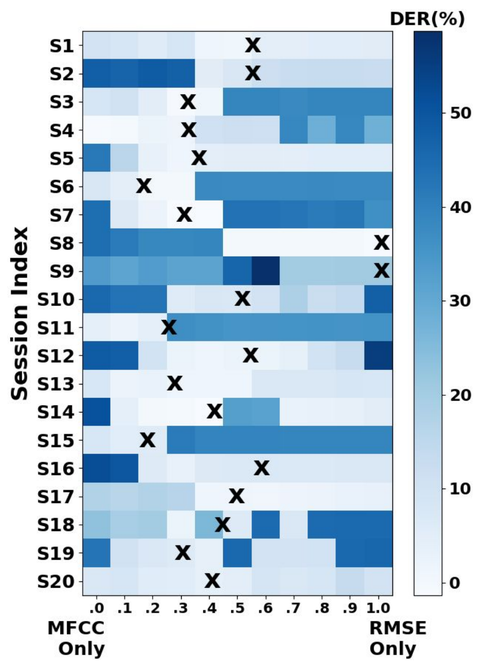


Fig. 2. Estimated weights (X) and DER for USCDiarLibri2,4

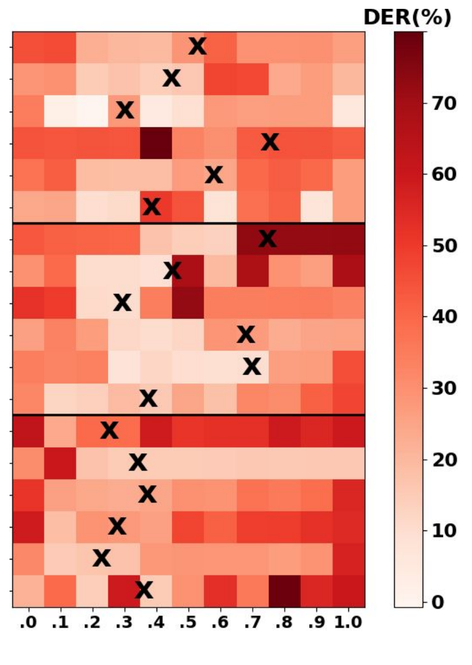


Fig. 3. Estimated weights (X) and DER for RT06S.

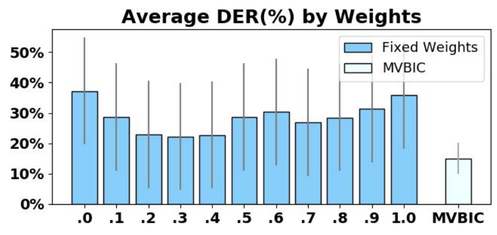


Fig. 4. Average DER by fixed weights and MVBIC for USCDiarLibri2,4.

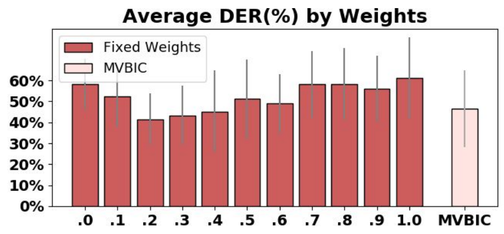


Fig. 5. Average DER by fixed weights and MVBIC for subset of RT06S dataset.

Tae Jin Park: taejinpa@usc.edu / Panayiotis Georgiou: georgiou@sipi.usc.edu

Ming Hsieh Institute
Ming Hsieh Department of Electrical Engineering