SUBBAND WEIGHTING FOR BINAURAL SPEECH SOURCE LOCALIZATION

Girija Ramesan Karthik, Parth Suresh and Prasanta Kumar Ghosh
SPIRE Lab, Electrical Engineering, IISc, Bengaluru.

Objective

- Machine localization of sound sources is necessary for applications such as human-robot interaction, surveillance and hearing aids.
- Adding more microphones can help increase the localization performance. However, humans have an incredible ability to localize sounds with just two ears using two major cues: i) interaural time difference (ITD) and ii) interaural level difference (ILD).
- Our objective is to localize a speech source from a binaural recording using ITDs and propose a new method using subband reliabilities.

Localization setup

- This difference in the distance travelled causes interaural time difference (ITD) and level difference (ILD) between the two microphone signals.
- Given omnidirectional microphones, an impulsive source will have frequency independent ITD & ILD.
- However, in binaural recordings, reflections & diffractions caused by the head makes ITD & ILD frequency dependent. This dependency is captured by the Head Related Transfer Function (HRTF).

Frequency dependent ITD extraction

- The ITD distributions of different subbands are different and hence they have different reliabilities.
- The frame level localization accuracy of each subband is obtained using clean speech. The reliability (R) of the i-th subband is calculated by subtracting the localization error (S) from a chosen threshold (T).
- To account for a larger set of possible weights, nonlinear warping is performed on the obtained reliabilities.

Subband reliability and nonlinear warping

- The rate of change of weights is not linear with α. The cumulative rate of change of weights w.r.t. α is shown below. It can be seen that most of the rate of change is concentrated towards α=1. Hence, the need for more α samples in close to 1.
- A set of 100 α values in the range of -0.9999 to 0.9999 and threshold (T) from 14.42 to 40 in steps of 0.5 have been considered.
- The (α,T) pair with the lowest localization error at each SNR is calculated. The corresponding weights are referred to as Best SNR Specific Weights (BSSW).
- The best (α,T) pair corresponding to the Best SNR Independent weights (BSIW) is calculated as below.
  \[
  (\alpha_{\text{BSIW}}, T_{\text{BSIW}}) = \arg \min_{\alpha, T} \sum_{k=1}^{N_{\text{SNR}}} (E_{\theta,k}(T) - E_{\min,k}(T))
  \]
- On similar lines, the worst SNR independent weights (WSIW) have also been calculated.
- BSW have α close to 1. This is equivalent to choosing only the most reliable subband. WSIW have α close to -1 which is equivalent to ML.

Localization

- GMMS are trained on the ITDs of each direction for each subband.
- Obtain likelihood of ITDs w.r.t. each direction (k) in each frame (j) in each subband (i).
- Calculate subband reliability using frame level localization error.
- Proposed Weighted Maximum Likelihood (WML) based localization.

Results

- We propose a weighted Maximum Likelihood (WML) method for binaural speech source localization.
- This reliability of each subband is used as the weights with the inclusion of non-linear warping to account for changes in reliabilities with SNR and also to span a larger space of possible weights. Experimental results with the best set of weights show that WML performs better than ML.

Conclusion

- Authors thank the Pratiksha Trust for their support.

References


https://spire.ee.iisc.ac.in/spire

karthikgr, prasantg@iisc.ac.in