Exploiting non-negative matrix factorization for binaural sound source localization in the presence of directional interference

IEEE ICASSP 2021

Poster ID: #3127

Ingvi Örnolfsson, Torsten Dau*, Ning Ma† and Tobias May**

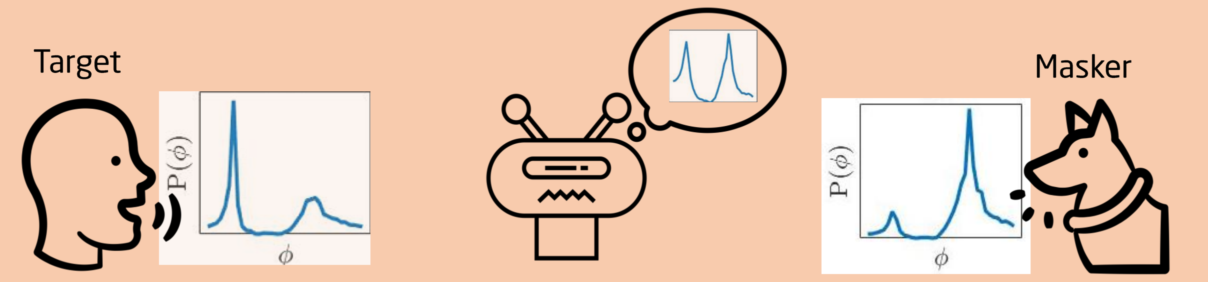
*DTU Health Tech
Hearing Systems Group

†TUOS Department of Computer Science
Speech and Hearing Research Group

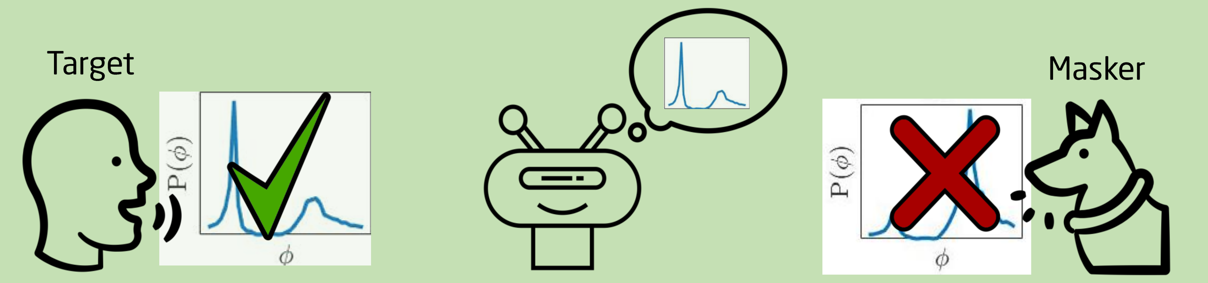
An existing binaural localization system based on a deep neural network (DNN) performs well in reverberant conditions, but performance deteriorates in conditions with directional interference. To overcome this limitation, the system is extended with a separation stage based on nonnegative matrix factorization (NMF). Different approaches to validating and training the system are explored. Extending the system in this way greatly improves performance in conditions with directional interference.

Task to solve:
- Predict azimuth angle ϕ of a speech source
- One non-speech masking source is present

Problem with existing system:
- Masker provides competing ITD and ILD cues



Solution:
- Add separation stage to isolate speech cues



The study of sound localization has been an active research field for over a hundred years. Today, some relevant applications include[1]:

Hearing aids, where beamforming is used for spatial filtering.

Social robots that need to be able to communicate with humans.

Auditory scene analysis, where spatial positioning of sources is important.

Our study extends an existing machine listening system with a separation stage. This extensions allows it to cope with directional interference.

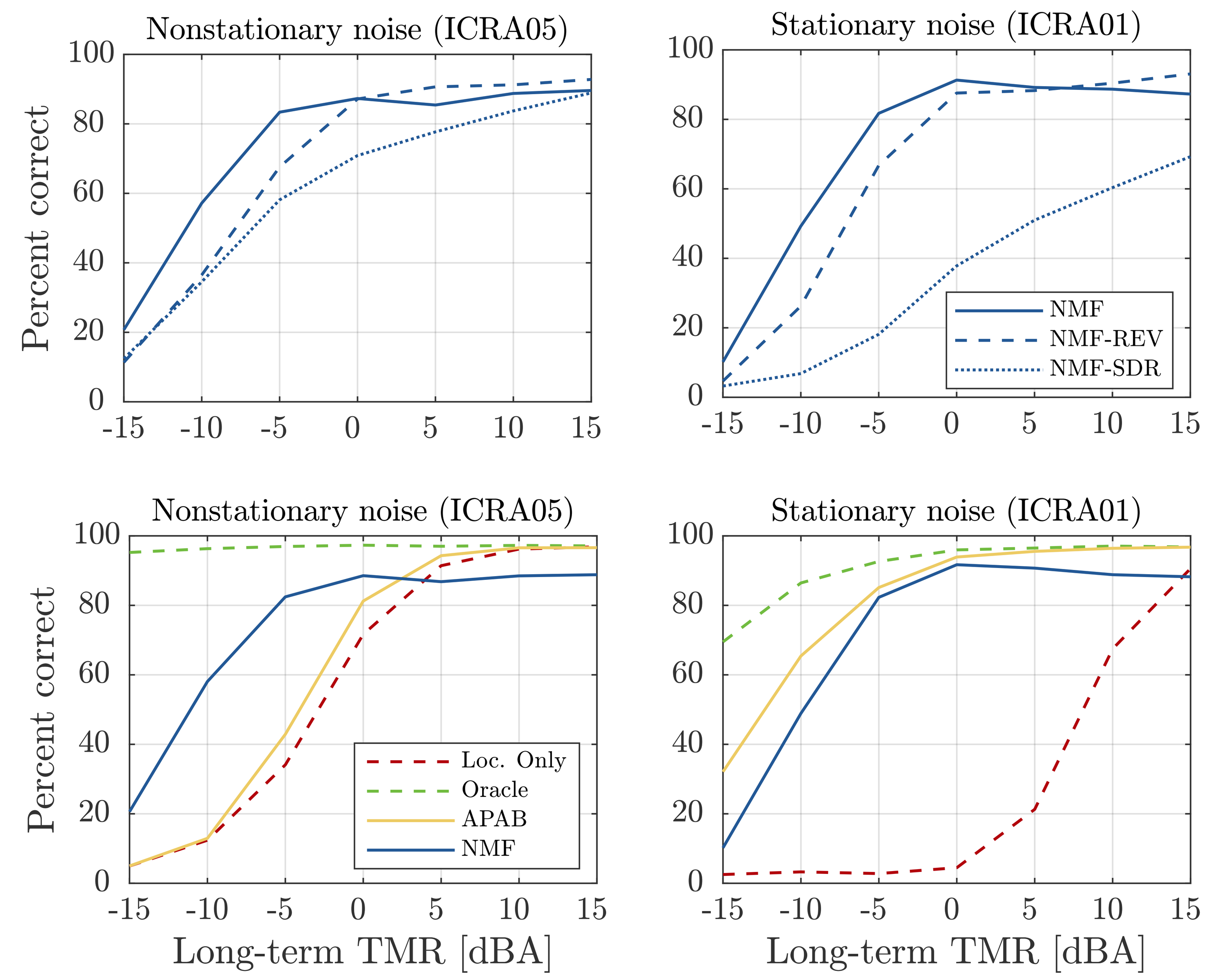
Three main research questions are adressed in our study:

1. Should the NMF be validated on a separation or a localization metric?
2. Should reverberant speech be used as training material for the NMF?
3. Is the proposed separation stage better than a learning-free approach?

To address these questions, three NMF variants are tested. The best of these is then compared to three baseline algorithms.

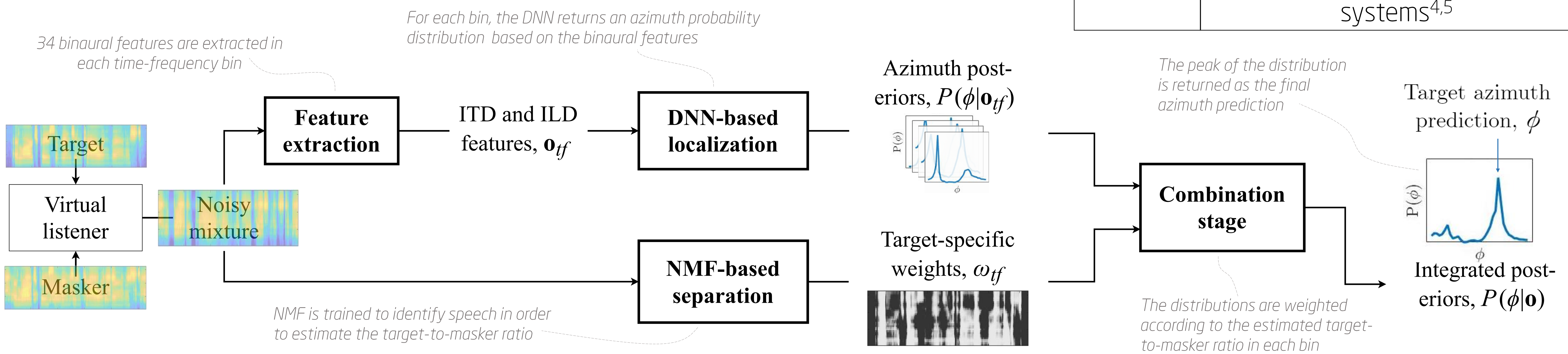
Algorithm	*Validation metric*	*Training*
NMF	Localization (PC)	Anechoic
NMF-SDR	Separation (SDR)[2]	Anechoic
NMF-REV	Localization (PC)	Reverberant
Baselines	*Description*	
Loc. Only	No separation[3]	
Oracle	Ideal separation	
APAB	Adaptive post-filter driven by two single-channel noise reduction systems[4,5]	

The results of our experiments are based on simulations in five different rooms using both stationary an nonstationary noise types.



To conclude, our study showed that using an NMF-based separation stage is a useful way to deal with directional interference in a state-of-the-art localization system. It was also shown that:

1. It is of great importance that the NMF-based separation stage is validated using a localization metric.
2. Training the NMF on reverberant speech only aids localization in high target-to-masker ratio conditions.
3. NMF is better than the learning-free approach in nonstationary noise. NMF also performs well in stationary noise, but not as well as the baseline.



[1] T. Rohdenburg, S. Goetze, V. Hohmann, K. Kammeyer, and B. Kollmeier: "Objective perceptual quality assessment for self-steering binaural hearing aid microphone arrays," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2008, pp. 2449-2452

[2] E. Vincent, R. Gribonval, and C. Févotte: "Performance measurements in blind audio source separation", IEEE Transactions on Audio, Speech and Language Processing, vol. 14, pp- 1462-1469, 2006

[3] N. Ma, J. A. Gonzales, and G. J. Brown: "Robust binaural localization of a target sound source by combining spectral source models and deep neural networks", *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2016

[4] Y. Ephraim, and D. Malah: ""Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* 1984, vol. 32, pp. 1109-1121

[5] C. Ris, and S. Dupont: "Assessing local noise level estimation methods: Application to noise robust ASR," *Speech Communication*, 2001, vol. 34(1-2), pp. 141-158