Target Speaker Selection for Neural Network Beamforming in Multi-Speaker Scenarios

Luan Vinícius Fiorio
Department of Electrical Engineering
Eindhoven University of Technology
Eindhoven, The Netherlands
l.v.fiorio@tue.nl

Bruno Defraene
NXP Semiconductors
Leuven, Belgium
bruno.defraene@nxp.com

Johan David
NXP Semiconductors
Leuven, Belgium
j.david@nxp.com

Alex Young
NXP Semiconductors
Eindhoven, The Netherlands
alex.young@nxp.com

Frans Widdershoven
NXP Semiconductors
Eindhoven, The Netherlands
frans.widdershoven@nxp.com

Wim van Houtum
NXP Semiconductors
Eindhoven, The Netherlands
wim.van.houtum@nxp.com

Ronald M. Aarts
Department of Electrical Engineering
Eindhoven University of Technology
Eindhoven, The Netherlands
R.M.Aarts@tue.nl

***Abstract*—We propose a speaker selection mechanism (SSM) to enhance the training of a beamforming neural network. Our approach is motivated by the observation that listeners typically orient themselves toward the target speaker at a slight undershot angle. The mechanism enables the neural network to learn which speaker to focus on in multi-speaker scenarios, based on the relative positions of the listener and speakers. Importantly, only audio input is required during inference. We conduct acoustic simulations to evaluate the effectiveness of the SSM, demonstrating its impact on performance. Results show significant increase in speech intelligibility, quality, and distortion metrics, outperforming both the ideal minimum variance distortionless filter and the same neural network model trained without SSM.**

***Index Terms*—Speaker selection mechanism, neural network, audio beamforming, cocktail party problem**

I. Introduction

"How do we recognize what one person is saying when others are speaking at the same time?" [1, p. 117]. This simple question formulates the *cocktail party problem*, which refers to the ability of the human hearing to separate voices that are mixed, in frequency and time. While such an ability is present in normal hearing, hearing impaired listeners might face difficulty in segregating auditory streams [2].

Hearing impaired listeners frequently rely on hearing aids, sound-amplifying devices which employ beamforming strategies. Such devices usually beamform in front of the listener, while recent findings show that the listener's head has a tendency to undershot the target speaker's position [3]. Beamforming algorithms help improving speech intelligibility and sound quality [4], however, in reverberant multi-speaker scenarios, the performance of algorithms such as the minimum variance distortionless response (MVDR) filter is reduced [5]. More recently, audio beamforming was developed using neural networks (NN), end-to-end [6] or estimating signals fed into a beamforming filter [7]. Such approaches usually do not take multi-speaker scenarios into account, limiting its application, or employ additional sensors (e.g., cameras) for guiding the beam, which can be prohibitive for most hearing aid devices.

Inspired from the findings of [3] regarding the presence of an undershot angle between listener's head and speaker direction, we propose a speaker selection mechanism for the training of beamforming neural networks. The mechanism teaches the model to focus on the target speaker based on the smallest undershot angle, requiring only audio information during inference. To the best of our knowledge, this is the first study to propose a solution for this task that neither requires the listener to face the speaker at any time nor relies on additional sensors. Through acoustic simulations, we show that a neural network trained with the mechanism is able to outperform the baseline model, trained without it, and the MVDR filter [8]. We also show that the proposed algorithm is robust to changes in number and position of speakers, a significant progress toward solving the cocktail party problem.

II. Preliminaries

The problem we tackle consists of multi-microphone audio beamforming in a multi-speaker scenario, where the microphones are positioned as of simulating hearing aid devices wore by a listener. N speakers and a listener are randomly positioned in a reverberant room. The listener can look toward one of the speakers directly, or with an *undershot* azimuth angle. For generality, we also consider the overshot (though we prioritize the term undershot for readability) when the listener looks further than the desired speaker angle. Our objective is to extract the clean reverberant speech of the desired speaker only using audio information.

An example can be seen in Fig. 1, where in a reverberant room, a listener L looks toward a speaker S_2 with an undershot angle θ_u. In this case, S_2 is the desired speaker while S_1 is undesired. The undershot angle can be described in terms

This work was supported by the Robust AI for SafE (radar) signal processing (RAISE) collaboration framework between Eindhoven University of Technology and NXP Semiconductors, including a Privaat-Publieke Samenwerkingen-toeslag (PPS) supplement from the Dutch Ministry of Economic Affairs and Climate Policy.

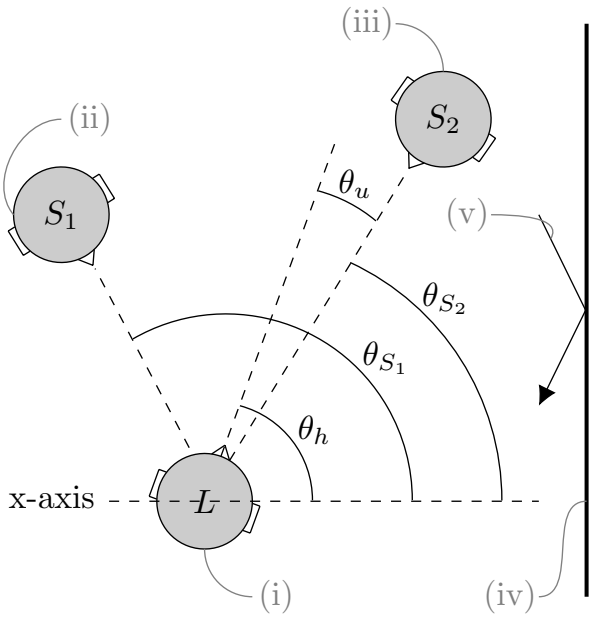

Fig. 1: Example scenario of the considered problem. The indication arrows point out to: (i) listener L; (ii) speaker S_1; (iii) speaker S_2; (iv) wall; and (v) reverberation.

of the listener's head center axis angle θ_h and the angle of the desired speaker θ_{S_2} (θ_{S_1} for the undersired speaker), in relation to the listener's x-axis, as $|\theta_u| = |\theta_h - \theta_{S_2}|$. In this example, the objective would be to extract the reverberant speech of speaker S_2 as received by a reference microphone.

The output $y_m(t)$ of each microphone m is defined by the speech fragment $s_n(t)$ of speaker n, convolutioned ($*$) with a room impulse response (RIR) $g_{nm}(t)$ from speaker n to microphone m, summed for all speakers. This is described as

$$y_m(t) = \sum_n s_n(t) * g_{n,m}(t). \tag{1}$$

Our objective is to extract the desired (subscript d) reverberant speech at the reference microphone $s_{n=d}(t) * g_{n=d,m=ref.}(t)$, solely given microphone outputs $y_m(t),\ \forall\ m \in [1,\ ...,\ M]$, while speech coming from other speakers is treated as interference. Additional noise is not considered in order to facilitate the demonstration of the proposed method.

The system operates in the time-frequency domain, where $Y_m(t,f)$ and $S_n(t,f)$ are, respectively, the short-term Fourier transform (STFT) of the microphone outputs and the reverberant speech signal $s_n(t)$ captured by a reference microphone.

III. Speaker selection mechanism

We propose a speaker selection mechanism (SSM) for allowing a neural network to learn which speaker is desired, and beamform toward it. This approach can be applied to a NN that is: estimating the steering vector of a classical beamforming algorithm, like the MVDR filter [7]; estimating the position of the target speaker [9]; estimating a time-frequency mask, which can be applied to the microphones' outputs via filter-and-sum operation [6]; among other uses. In this work, we choose to validate the proposed mechanism with an end-to-end neural network that estimates a multi-channel time-frequency mask for beamforming. Nevertheless, we assume that we have access to all microphones' outputs in the array, and that the position of listener and speakers is known during training.

Algorithm 1 Speaker selection mechanism for two speakers

procedure SSM(positions, $|\theta_u^{\max}|$)
  Input: 2-dimensional position of listener $[a_L, b_L]$ and speakers $\Big[[a_{S_1}, b_{S_1}],\ [a_{S_2}, b_{S_2}]\Big]$ of an audio utterance
  Parameter: Maximum undershot angle $|\theta_u^{\max}|$
  Output: Index of desired speaker
  Calculate the speakers' angles relative to the listener's x-axis:

$$\theta_{S_1} = \text{atan2}(b_{S_1} - b_L,\ a_{S_1} - a_L)$$
$$\theta_{S_2} = \text{atan2}(b_{S_2} - b_L,\ a_{S_2} - a_L)$$

  Determine the admissible range for θ_h:
Ensure that the listener's head angle is always closer than $|\theta_u^{\max}|$ from both speakers:

$$\theta_h^{\min} = \min\{\theta_{S_1}, \theta_{S_2}\} - |\theta_u^{\max}|$$
$$\theta_h^{\max} = \max\{\theta_{S_1}, \theta_{S_2}\} + |\theta_u^{\max}|$$

  Sample the listener's head angle:

$$\theta_h \sim \text{Uniform}(\theta_h^{\min},\ \theta_h^{\max})$$

  Calculate the undershot angles for each speaker:

$$|\theta_{u1}| = |\theta_h - \theta_{S_1}|$$
$$|\theta_{u2}| = |\theta_h - \theta_{S_2}|$$

  Select the speaker:
  if $|\theta_{u1}| < |\theta_{u2}|$ **then**
    Return 1 ▷ Speaker 1 is desired
  else
    Return 2 ▷ Speaker 2 is desired
  end if
end procedure

The SSM works as follows. For each training utterance, we calculate the absolute value of the undershot angles for all speakers. We then identify the speaker that results in the smallest absolute undershot angle and set it as desired. Moreover, the desired speaker is used as a target for calculating the loss, during training, for that specific utterance. The target speaker in the loss function changes dynamically according to the smallest undershot angle. We consider the criteria of smallest undershot angle for changing desired speaker, but the movement of the head could be more explored, being out of scope for this paper. Alg. 1 details the speaker selection mechanism for an example situation of two speakers. Notice that, in inference mode, there is no need to provide any information regarding position. The NN-based system trained with the proposed mechanism is able to beamform toward the desired speaker solely with audio information.

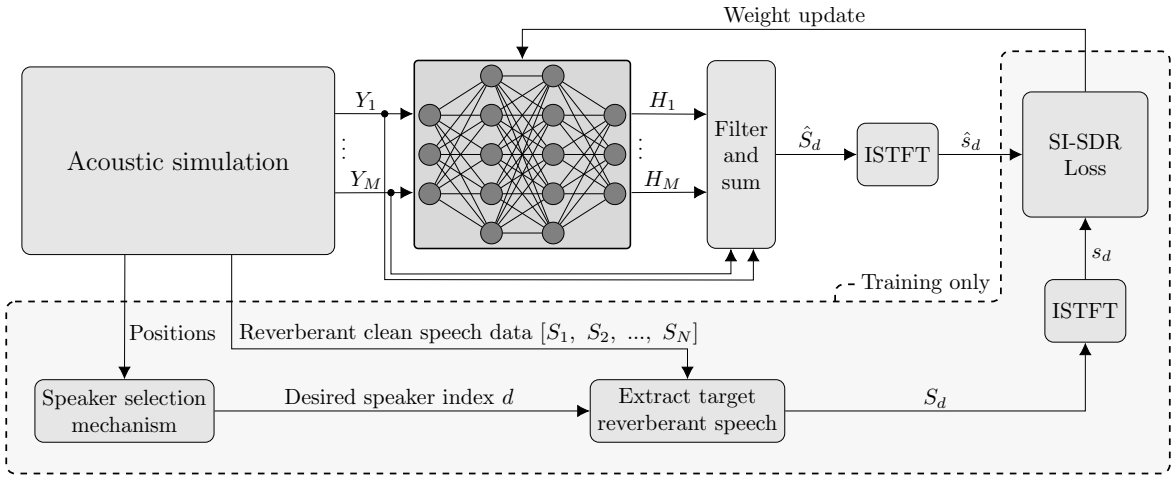

Fig. 2: End-to-end beamforming neural network training system employing speaker selection mechanism.

Differently from [10], our approach does not require the listener to face the target at any moment, and only one neural network is used. We also don't take visual cues into account, e.g., as considered in [7], since we assume that the neural network can obtain spatial information from multi-microphone audio features. Next, we describe the model and simulation framework for evaluating the proposed SSM.

IV. Model and simulation framework

We evaluate the SSM with an end-to-end neural network beamforming system, as per Fig. 2. A simulation environment outputs multi-microphone recordings, which are preprocessed and fed into the NN model in the time-frequency domain. The output of the neural network consists of a complex multi-channel mask $H_m(t,f),\ \forall m \in [1,\ ...,\ M]$, applied to the microphone recordings with a filter-and-sum operation, as

$$\hat{S}_d(t,f) = \sum_m Y_m(t,f) \cdot H_m(t,f). \tag{2}$$

The model description is given in the following.

A. Audio beamforming model

We consider a NN-based beamforming approach with filter-and-sum, similar to [6], but we simplify the model by using only real-valued operations, with a real-imaginary split at the input, concatenating both in the frequency axis. Consequentially, the output is recombined as a complex mask. Further on reducing the model's complexity, the convolutions are defined only in the frequency axis, as we did not observe significant performance difference against kernels in both frequency and time axis. The NN model is depicted in Fig. 3.

The model is trained to maximize the scale-invariant signal-to-distortion ratio (SI-SDR) of filtered microphone outputs in relation to the desired speaker's reverberant speech at the reference microphone. Differently from scale-invariant signal-to-noise ratio used in [6], we consider the SI-SDR since it is a lower bound to both SDR and SNR [11].

For comparison, we train the same model twice. First, trained with the SSM for speaker-aware beamforming. Second, without using the proposed mechanism, by always setting a random speaker as the desired target, creating a NN baseline for the task that we are aiming to solve – beamforming on multi-speaker scenarios with undershot angles. We also compare it to an ideal MVDR filter, obtained as [8]

$$\boldsymbol{w}_{MVDR}(f) = \frac{\boldsymbol{\Phi}_{uu}^{-1}(f)\boldsymbol{\Phi}_{ss}(f)}{\mathrm{Trace}(\boldsymbol{\Phi}_{uu}^{-1}(f)\boldsymbol{\Phi}_{ss}(f))}\boldsymbol{r}, \tag{3}$$

where $\boldsymbol{\Phi}_{uu}(f)$ and $\boldsymbol{\Phi}_{ss}(f)$ are the power spectral density matrices of undesired speech and desired speech, respectively, and $\boldsymbol{r}$ is a one-hot vector representing the reference channel. For the considered ideal case, $\boldsymbol{\Phi}_{uu}(f)$ and $\boldsymbol{\Phi}_{ss}(f)$ are known. The weights are applied to the multi-microphone outputs as $\hat{\boldsymbol{S}}_d(f) = \boldsymbol{w}_{MVDR}(f)^H \boldsymbol{Y}(f)$. Additionally, the ideal MVDR is also equivalent to the optimal case for when MVDR parameters are calculated with NNs, e.g., [7].

B. Acoustic simulation setup

We simulate a reverberant room with four microphones and two speakers. First, a rectangle-shaped room of size 5.15 x 3.75 x 2.65 m is defined. Although the room size is fixed, the time it takes for sound pressure to reduce by 60 dB (T60) is defined over a variable range, assuring generality. The room impulse response (RIR) for each speaker in relation to each microphone is generated using gpuRIR [12]. We set up the simulation as described in the following.

Four omnidirectional microphones are positioned similarly as in a hearing aid device wore by a person. First, we (randomly) define the position of the listener, and we assume a radius equal to 0.15 m, which is similar to the average head breadth of an adult person. Two groups of two microphones are positioned in the east-most point and equivalently at the west-most point. The microphones within each group are split from each other by 0.50 cm, and positioned at the same height, with the front left microphone taken as reference.

Moreover, the speakers are randomly positioned following a few constraints. The first constraint is that the speakers cannot be closer than 1.00 m from the listener, and they cannot be closer than 1.00 m from each other. At the moment of

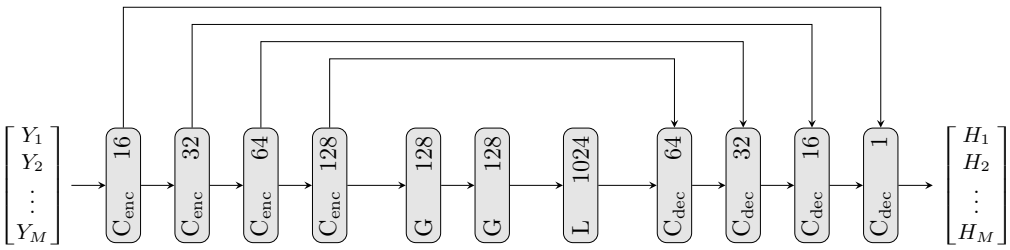

Fig. 3: Schematic of the considered neural nerwork model. The number in each layer indicates output channels. C_{enc} consist of Conv2D encoder layers with BatchNorm2D and ReLU functions in all layers. A Tanh is applied to the encoder output to bound values, ensuring stable inputs for the recurrent layers. G layers are gate recurrent units (GRUs), and L is a linear layer. C_{dec} are Conv2D.T decoder layers, with BatchNorm2D and ReLU activation in all layers but the last, without normalization or activation. All $\mathrm{C}_{enc/dec}$ kernels are (8,1) with stride (2,1) and padding (3,0). Upper lines represent skip connections.

TABLE I: Variable parameters and ranges for the acoustic simulation.

Parameter	Min. value	Max. value
T60 (s)	0.20	1.00
SNR (dB)	-10.00	20.00
Listener/speaker height (m)	1.50	1.95
Undershot angle (°)	-30	+30

positioning, the absolute angle difference of both speakers in relation to the listener must be of at least 45 degrees, avoiding that a speaker would be too close or behind the other speaker. Both listener and speaker positions are limited to be distant from any wall at least twice the head breadth value. The listener and speaker points are positioned with a height ranging from 1.50 and 1.95 m, similar to most adult humans' height. When all speakers and listener are positioned, the head angle of the listener is defined by randomly rotating the center of the two groups of microphones in the azimuth direction, but not exceeding a maximum undershot of 30 degrees. The maximum undershot constraint provides a better sense of reality, as a listener would not look too far from the desired speaker. Additionally, the signal-to-noise ratio (SNR), calculated with the mixed utterance (representing noisy signal) against the desired-speaker-only utterance (representing signal) is randomly varied from -10 to 20 dB. Note that speech traces are combined such that there is minimum silence period, but still sounding natural. Table I summarizes the variable parameters in the simulation.

C. Data

We use the LibriTTS dataset [13] for the acoustic simulation. For each speaker, traces of speech are randomly selected and resampled to 16 kHz, combined until a duration of 10 seconds is reached, with a random fade-in and fade-out of 0.05 to 0.20 seconds. Each speech trace is multiplied by a gain, randomly defined from -3 to 3 dB. Both speech utterances are adjusted to avoid clipping when combined. Each utterance is then convolved with the RIR referent to that speaker and microphones, which are obtained as described in Section IV-B, according to (1), resulting in the microphone outputs.

Moreover, the STFT operation is applied to the microphone outputs for 256 samples, with a Hann window of size 256 and a hop of 128 samples. The STFTs used as input to the neural network are normalized by their mean and standard deviation. Real and imaginary parts are then concatenated in the frequency axis, forming the input to the neural network. For training, the 'train-clean-360' subset of LibriTTS is used, with 360 hours of raw audio. The evaluation is performed on the 'test-clean' set, with approximately 8.6 hours of data.

V. Results and discussion

We train the neural network model described in Section IV-A with and without the SSM proposed in Section III, for $N = 2$ speakers, according to the acoustic parameters defined in Section IV-B, with the data mentioned in Section IV-C. We also consider the (ideal) MVDR filter as a baseline, calculated as in (3) with access to all separate (reverberant) signals, i.e., always beamforming in the target speaker direction. In Table II, we show the average values over the 'test-clean' set of LibriTTS of short-time objective intelligibility (STOI) [14], perceptual evaluation of speech quality (PESQ) [15], and SI-SDR, for the mixed audio at the reference microphone and the filtered signals.

We can see from Table II, for $N = 2$ speakers, that the use of the SSM in training can significantly increase the performance of the neural network-based beamforming model, for all considered SNRs. As expected, the proposed mechanism is able to teach the network which speaker to target at each utterance. When the model is trained without such information, a lower signal-to-noise ratio condition causes the performance to be drastically affected since the NN model "confuses" the choice of speaker, to the point of achieving lower metrics than the mixed signal. We can see that, as the SNR of the speech combination increases, the NN without SSM becomes able to separate desired from undesired speaker, indicating that it is focusing on the higher-amplitude signal, a major feature in the audio combination. However, even for higher SNR levels, the performance of the baseline NN is insufficient, as the model trained with SSM almost always forms an upper bound for the NN's performance.

Moreover, the MVDR filter is outperformed by the NN with SSM training for almost all cases. The baseline NN provides a similar or better performance than the MVDR filter at higher SNRs. That is due to the MVDR formulation, which assumes an acoustic scene with anechoic conditions, while the NNs can learn to suppress the effects of reverberation. For higher

TABLE II: Average STOI, PESQ, and SI-SDR over the 'test-clean' LibriTTS set for the mixed audio and the NN-filtered speech, trained with and without SSM for two speakers and evaluated for $N = 2$ and $N = 3$ speakers.

		-10 dB SNR			0 dB SNR			10 dB SNR			20 dB SNR		
N	**Method**	**STOI**	**PESQ**	**SI-SDR**	**STOI**	**PESQ**	**SI-SDR**	**STOI**	**PESQ**	**SI-SDR**	**STOI**	**PESQ**	**SI-SDR**
2	None (mixed)	0.384	1.237	-9.970	0.634	1.535	0.032	0.849	2.334	10.031	0.958	3.431	20.030
	MVDR filter	0.447	1.322	-6.445	0.682	1.740	2.043	0.838	2.430	5.109	0.868	2.707	1.866
	NN	0.346	1.219	-11.172	0.629	1.532	-0.007	0.861	2.430	10.913	**0.963**	3.617	**20.790**
	NN + SSM training	**0.526**	**1.366**	**-1.608**	**0.736**	**1.851**	**5.012**	**0.891**	**2.790**	**12.541**	**0.963**	**3.746**	20.227
3	None (mixed)	0.313	1.237	-9.990	0.580	1.465	0.016	0.828	2.211	10.017	0.954	3.368	20.018
	MVDR filter	0.371	**1.285**	-7.314	0.634	1.626	1.786	0.827	2.327	5.690	0.872	2.714	2.238
	NN	0.299	1.225	-10.582	0.583	1.468	0.156	0.845	2.323	11.025	0.960	3.569	**20.740**
	NN + SSM training	**0.400**	1.253	**-6.311**	**0.669**	**1.652**	**3.361**	**0.874**	**2.621**	**12.222**	**0.961**	**3.703**	20.273

SNR, the reverberation of the desired speaker has more energy, contaminating the direct path and deteriorating the MVDR performance, which can be noticed in terms of SI-SDR.

We also check the robustness of the proposed mechanism against changes in the environment by re-evaluating all methods for a different acoustic scenario. Now, we consider a more challenging case of $N = 3$ speakers, with minimum distance between listener to speakers, and speakers to speakers, of 0.5 m, and minimum absolute angle difference of speakers in relation to the listener center axis of at least 20 degrees. All other simulation parameters are kept as before. The training of the neural networks is not re-executed and their parameters are kept exactly the same as for $N = 2$ speakers.

As shown in Table II, with $N = 3$ speakers, the proposed SSM is robust to changes in the number of speakers and positioning, even at a very low SNR (-10 dB), outperforming the baselines for almost all cases. As the target speaker is successfully extracted, all other speech traces are filtered out, independently of number of undesired speakers. This gives a strong indication that the NN trained with SSM treats all undesired sound homogeneously, as if it was noise. When the SNR is low, the NN without SSM again fails to extract the desired speech, however, it manages to beamform toward the desired speaker under higher SNRs conditions, given the easier settings. The MVDR filter is again affected by the presence of reverberation, which becomes clear with the obtained SI-SDR at the highest considered SNR, as previously explained.

VI. Conclusion

We proposed a speaker selection mechanism for the training of a neural network model on the task of audio beamforming. The SSM dynamically changes the target speaker in the loss function, at every utterance, focusing on the closest speaker to the listener's head center axis. Through acoustic simulations, the neural network model trained with SSM was able to outperform the baseline NN model (trained without it) and the (ideal) MVDR filter, achieving significantly higher performance metrics. Additionally, we showed that the SSM is robust to changes in the acoustic scene – number of speakers and positioning. The proposed speaker selection mechanism represents a leap toward the solution of the cocktail party problem. For future work, we suggest the employment of the SSM in different NN-based beamforming systems, like the estimation of classical beamforming filters, and to the solution of multi-speaker source localization problems.

References

[1] A. W. Bronkhorst, "The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions," *Acustica*, vol. 86, pp. 117–128, 2000.

[2] M. A. Bee and C. Micheyl, "The cocktail party problem: What is it? How can it be solved? And why should animal behaviorists study it?" *Journal of Comparative Psychology*, vol. 122, no. 3, pp. 235–251, Aug 2008.

[3] H. Lu, M. F. McKinney, T. Zhang, and A. J. Oxenham, "Investigating age, hearing loss, and background noise effects on speaker-targeted head and eye movements in three-way conversations," *J Acoust Soc Am*, vol. 149, no. 3, p. 1889, Mar 2021.

[4] J. Gerald Kidd, C. R. Mason, V. Best, and J. Swaminathan, "Benefits of Acoustic Beamforming for Solving the Cocktail Party Problem," *Trends in Hearing*, vol. 19, p. 2331216515593385, 2015, pMID: 26126896.

[5] B. Cauchi, I. Kodrasi, R. Rehr *et al.*, "Combination of MVDR beamforming and single-channel spectral processing for enhancing noisy and reverberant speech," *EURASIP Journal on Advances in Signal Processing*, vol. 2015, no. 61, 2015.

[6] Y. Chen, Y. Hsu, and M. R. Bai, "Multi-channel end-to-end neural network for speech enhancement, source localization, and voice activity detection," 2022. [Online]. Available: https://arxiv.org/abs/2206.09728

[7] Z. Zhang, Y. Xu, M. Yu, S.-X. Zhang, L. Chen, and D. Yu, "ADL-MVDR: All Deep Learning MVDR Beamformer for Target Speech Separation," in *2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 6089–6093.

[8] M. Souden, J. Benesty, and S. Affes, "On Optimal Frequency-Domain Multichannel Linear Filtering for Noise Reduction," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 2, pp. 260–276, 2010.

[9] E. Grinstein, C. M. Hicks, T. van Waterschoot, M. Brookes, and P. A. Naylor, "The Neural-SRP Method for Universal Robust Multi-Source Tracking," *IEEE Open Journal of Signal Processing*, vol. 5, pp. 19–28, 2024.

[10] B. Veluri, M. Itani, T. Chen, T. Yoshioka, and S. Gollakota, "Look Once to Hear: Target Speech Hearing with Noisy Examples," in *2024 CHI Conference on Human Factors in Computing Systems*, ser. CHI '24. New York, NY, USA: Association for Computing Machinery, 2024.

[11] J. L. Roux, S. Wisdom, H. Erdogan, and J. R. Hershey, "SDR - half-baked or well done?" *arXiv preprint 1811.02508*, 2018.

[12] D. Diaz-Guerra, A. Miguel, and J. R. Beltran, "gpurir: A python library for room impulse response simulation with gpu acceleration," *Multimedia Tools and Applications*, vol. 80, no. 7, pp. 5653–5671, 2021.

[13] H. Zen, V. Dang, R. Clark, Y. Zhang, R. J. Weiss, Y. Jia, Z. Chen, and Y. Wu, "LibriTTS: A corpus derived from librispeech for text-to-speech," *arXiv preprint 1904.02882*, 2019.

[14] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, "A short-time objective intelligibility measure for time-frequency weighted noisy speech," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2010, pp. 4214–4217.

[15] A. Rix, J. Beerends, M. Hollier, and A. Hekstra, "Perceptual evaluation of speech quality (pesq)-a new method for speech quality assessment of telephone networks and codecs," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*, vol. 2, 2001, pp. 749–752 vol.2.