P&D Information Systems and Signal Processing

Week 4 - March 7-11, 2021
Frequency-domain GSC and demo

Giuliano Bernardi, Randall Ali, Santiago Ruiz, Paul Didier[1], Marc Moonen

Version 1.0

Part 1: Frequency-domain DOA-informed GSC

The time-domain GSC (Week 3), based on a DAS BF, does not perform very well in scenarios with significant reverberation. To achieve improved performance, the GSC will now be realized in the frequency domain, where better models can be used to replace the DAS BF and blocking matrix. For this, a short-time Fourier transform (STFT) is to be used to first transform successive time-domain signal frames to the frequency domain, then a GSC will be operated in each individual frequency bin, and finally the GSC outputs are to be transformed back to the time domain, to synthesize the overall output signal. Use m-files from Week 3 and Week 4 to create a new m-file `GSC_fd.m`, realizing the GSC in the frequency domain, and containing the following parts:

- The analysis part (time domain to frequency domain) can be adopted from the MUSIC-based DOA estimation procedure (Week 2), and so will serve two purposes now (DOA estimation and GSC). Use a 50% overlap between successive time-domain signal frames, so that the analysis part corresponds to a two-fold oversampled filter bank operation. Optionally, the DFT operations can be supplemented with a windowing operation (e.g. a square-root Hanning window), to achieve a higher spectral resolution.

- For the GSC in an individual frequency bin, the DAS BF will be replaced by a 'filter-and-sum beamformer' (FAS BF), which will be a function of the estimated DOA and read from a pre-filled look-up table (FAS BF as a function of DOA and frequency). For each DOA and frequency, the FAS BF is defined by a vector of multipliers corresponding to a measured or generated steering vector for this DOA and frequency, in the considered scenario (and so including the reverberation). The corresponding blocking matrix can then be computed such that it is orthogonal to this steering vector. Further details are provided in the next paragraph.

- The synthesis part (frequency domain to time domain) will have IDFT and overlap-save operations or alternatively weighted overlap-add operations with a windowing operation matched to the analysis windowing operation (e.g. again a square-root Hanning window)[2].

GSC details:

- Create the steering vector for each DOA and frequency using measured or generated RIRs. For this, take the DFT of the RIRs and normalize them with respect to the first microphone

[1] For questions and remarks: `santiago.ruiz@esat.kuleuven.be` & `paul.didier@esat.kuleuven.be`

[2] See `https://homes.esat.kuleuven.be/~dspuser/DSP-CIS/2020-2021/material/chapter14.pdf` page 17-18.

to obtain $\mathbf{h}(\omega,\theta) = \dfrac{\mathbf{a}(\omega,\theta)}{a_1(\omega,\theta)}$. Create a look-up table with the corresponding steering vectors $\mathbf{h}(\omega,\theta)$.

- Define the FAS BF as $\mathbf{w}_{\mathrm{FAS}}(\omega,\theta) = \dfrac{\mathbf{h}(\omega,\theta)}{\mathbf{h}^H(\omega,\theta)\mathbf{h}(\omega,\theta)}$
- Define the blocking matrix $\mathbf{B}(\omega,\theta)$ such that $\mathbf{B}(\omega,\theta)\mathbf{h}(\omega,\theta) = \mathbf{0}$.
- Check if the steering vectors $\mathbf{h}(\omega,\theta)$ may also be used to replace $\mathbf{g}(\omega,\theta)$ in the MUSIC algorithm (Week 2, formula (5)) to improve the DOA estimation.

Test your `GSC_fd.m` on the scenario defined in Week 3, Part 4.

Hint 1: Check the help file of the command `null`, which may be useful in the above task. Pay attention to the expected dimension of the input arguments.

Part 2: Scenario and experiments

1. Use `mySA_GUI.m` to create a two time-varying acoustic scenarios (one anechoic and the other reverberant) consisting of one target audio source and one interfering source, whose positions change every 10 seconds, as follows:
   - Choose five angle locations for the target audio source, left of the broadside direction, i.e. between azimuthal directions $-90°$ to $0°$. Choose five angle locations for the interfering source, right of the broadside direction, i.e. between azimuthal directions $0°$ to $90°$.
   - Place a 5-microphone linear array with an inter-microphone distance equal to 5cm. Set the room size to 10m and T60 to 0s. Place the target audio source and interfering source on the first chosen angle locations on the left and right, respectively. Store the generated RIRs. Change the T60 to 1s and keep the remaining parameters fixed and generate the new set of RIRs. Make sure to save and name properly each set of RIRs. Repeat this step for each angle location, keeping the microphone array fixed.
   - Five sets of RIRs for a scenario with T60 equal to 0s and five others with T60 equal to 1s should be obtained. Use these to generate microphone signals where the target audio source plays the signal `part1 track1 dry.wav` and the interfering source plays the signal `part1 track2 dry.wav`.
   - Combine the signals accordingly to create the positional changes of the sources every 10 seconds.
2. Create an m-file that estimates the DOAs over time and the desired signal.
3. Quantify and demonstrate the achieved performance by using the signal-to-interference ratio (SIR) in the anechoic and reverberant scenarios (See Appendix). Compare the SIR values obtained in the microphone signals and after the filtering is applied. Tune the parameters of the DOA estimation to achieve the best possible performance.

Part 3: Experiments with measured RIRs

- Create microphone signals for both the head-mounted and linear arrays using the measured RIRs from the measurement session. One source location will be the target audio source, playing the signal `part1 track1 dry.wav` and the other location will be the interfering source, playing the signal `part1 track2 dry.wav`.
- Use these microphone signals in the DOA-informed GSC with VAD implementation of week 3 (Part 4) and compute the SIR. Compare the SIR values obtained in the microphone signals and after the filtering is applied. Tune the parameters of the DOA estimation to achieve the best possible performance.

Part 4: Report

Describe the constructed scenario, the system tuning, as well as the obtained results and achieved performance in a brief report (max 5 pages). Include links to audio files. Send the report and code to `santiago.ruiz@esat.kuleuven.be` & `paul.didier@esat.kuleuven.be`

Deadline: Friday March 11, 2022 (23:59).

Part 5: Demo

Demonstrate the final system to the teaching assistant in the lab session on Friday March 11, 2022.

Appendix

Signal-to-interference ratio (SIR)

The signal-to-interference ratio (SIR) quantifies the contribution of the desired signal with respect to the interference signal(s). In this case we aim to quantify the SIR for the target source signal with respect to the interfering source signal. The SIR is defined as

$$\mathrm{SIR} = 10\log_{10}\left(\frac{\sigma^2_{\hat{s}_d}}{\sigma^2_{\hat{s}_i}}\right) \quad \mathrm{dB} \tag{1}$$

where $\sigma_{\hat{s}_d}$ is the power of the estimated target source signal and $\sigma_{\hat{s}_i}$ the power of the interfering source signal. The file `compute_SIR.m` is provided.

Remark 1: To compute the SIR access to the target audio source and interfering source components in the microphone signals is required. The estimated beamformer is applied to both components separately and then equation (1) is used with the obtained signals.

Remark 2: Given that the scenario is time-varying the beamformer will change over time, hence a single SIR would not suffice. The metric should be computed over small time periods were the beamformer is kept constant.