FILTERING OF SEPARATE AUDIO TRACKS FROM SYNCHRONOUSLY PLAYING MUSICAL INSTRUMENTS USING BEAMFORMING IN THE TIME DOMAIN

Serdar Gareayaghi, Ivo Hagenmaier, Helen Henze and Gert Herold
Technische Universität Berlin, FG Technische Akustik
Einsteinufer 25, 10587 Berlin, Germany

Abstract

This work presents the development and assessment of a time-domain based beamforming algorithm and measurement setup using a linear microphone array for recording and filtering of two synchronously playing musical instruments generating separated audio files. Furthermore, a progressive method of designated octave band specific geometries for the used linear microphone array are proposed and compared to using the complete microphone array geometry for the whole frequency spectrum. To assess the recording results, varying positioning setups are compared to simulation results that were obtained using the Python package *Acoular*. While the initial algorithm and setup deliver satisfactory results for the simulations as well as for real life recordings, it is furthermore found that the progressive method enhances the results to a certain extent. Thus, this work poses a possible base for future research in the field of musical instrument recording and beamforming.

1 INTRODUCTION

The use of frequency-domain beamforming algorithms has become a common method for sound source localization and acoustic power level quantification. For certain applications, time-domain beamforming has shown itself especially suitable, finding use in aeroacoustic assessments [10], moving sound source localisation [8] as well as multi-channel speech processing and enhancement [2, 3, 5]. However, the idea of utilizing time-domain beamforming to generate real-time monitoring audio files and recordings of large musical ensembles like an orchestra has yet to be explored. In this contribution, time-domain based beamforming is employed for producing separated audio files for each instrument from a simultaneous recording of a guitar and a violin using a linear microphone array.

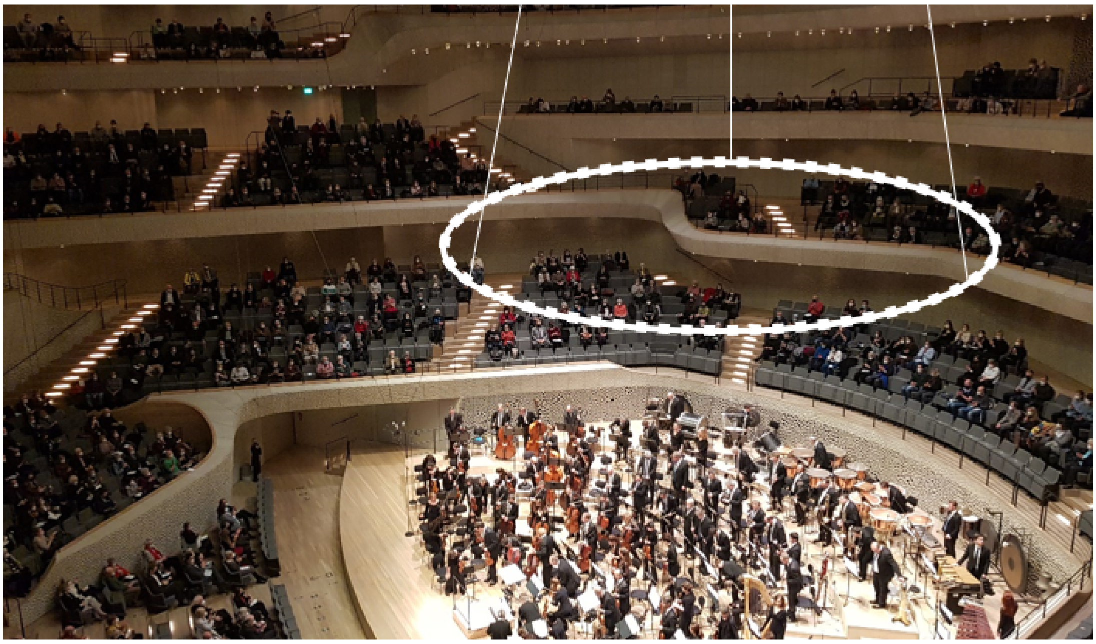


Figure 1: Photo of a concert hall with a schematic sketch of a microphone array above an orchestra.

By means of beamforming in the time domain, the signal captured by a linear microphone array is locally separated using the Python package *Acoular* [7], which was developed at the Chair of Engineering Acoustics at the Technical University of Berlin, and audio files corresponding to the respective instrument are generated. The isolation of the musical instruments is examined for different distances between array and instruments and between the instruments themselves, and the results are then compared. For this purpose, correlation coefficients are calculated and spectrograms are generated. In addition, the resulting audio files are subjectively evaluated.

In the further course of this work, after an explanation of the motivation for the investigations, the methodology and procedure will be explained. For this purpose, the simulations preceding the measurements and the creation of an individual array geometry are discussed. Subsequently, the setup and execution of the measurements in the anechoic room of the Technical University of Berlin will be explained before the results are presented, evaluated and eventually discussed. Finally, an outlook for further research based on this work is proposed.

This work was inspired by the idea of developing an algorithm based setup utilizing microphone arrays and the beamforming technique to generate real-time monitoring audio tracks for orchestras. As we know from live bands and singers, the use of real-time monitoring equipment helps musicians hear themselves or specific musicians. Current research in the field includes Zhong et. al's work on the use of beamforming techniques for surface sound separation, which does not consider generating an audio output from the simulation results [12], as well as Jimenez et. al's paper on stage acoustic auralisation, which uses playback recording for on stage auralisation [4]. Using beamforming algorithms for orchestra recording and monitoring to our understanding has yet to be explored, which is why we have decided to carry out this work.

The problem with monitoring orchestras is that the acoustic instruments used usually do not have a direct line-out, in contrast to, for example, an electric guitar. Therefore, each instrument or instrument group would need its own microphone for monitoring. By using the fixed microphone array in the concert hall, the monitoring would only be influenced by the positions of the instruments. However, the localization of the exact sound source could also be determined with the microphone array and the effort of equipping each concert with microphones for a different orchestra would be eliminated.

Another advantage of recording an entire orchestra with such microphone array could be the many possibilities of mixing the recorded instruments in post processing, as all audio tracks can be edited individually for each instrument. Contrary to proximity recordings using ON-microphones, the reverberation and room acoustic information could be gathered, creating a more profound sound output. A schematic sketch of such a circular array mounted above an orchestra can be seen in Fig. 1.

2 METHODS

The Python based framework *Acoular* is used to develop the time-domain based beamforming algorithm as well as to simulate expected results before the recordings. Furthermore, the framework is utilized to create an individually designed microphone array to meet the needs for recording the wide frequency range covered by the instruments.

The aim of this work is to create isolated audio files for simultaneously playing instruments. Therefore, the use of beamforming in the time domain is suitable. As we will know the sound source locations, we do not need to conduct sound source localization beamforming. Furthermore, the actual spectrum contents of the sounds recorded are not of main interest, which is why beamforming in the frequency domain does not possess any advantages over beamforming in the time domain. The following sections will give further explanation and details on preliminary considerations about the recordings on which the preceding simulations and the results that emerged from them are based. The development of the microphone array geometry is presented as well.

2.1 Array geometry

Regarding the microphone array geometry, the use of a linear array, also known as an acoustic antenna, is sufficient, because no vertical sound source information is to be acquired by the microphones as the musicians will be placed at roughly the same height.

The array consists of 64 microphones arranged in a base 15 logarithmic pattern to reduce aliasing effects. Both instruments should have approximately the same average distance to each microphone of the array as the sound pressure level decreases with distance. Therefore an array roughly symmetrical to the center point was designed. To create this array, every second microphone of the logarithmically spaced setup is mirrored around the y-axis. The hole created by the logarithm in the middle was filled by pushing the mirrored and unmirrored sides of the array together to obtain the smallest spacing in the middle of the array with 0.014 m. The complete array geometry can be seen in Fig. 2 (bottom array labelled as ‘All’).

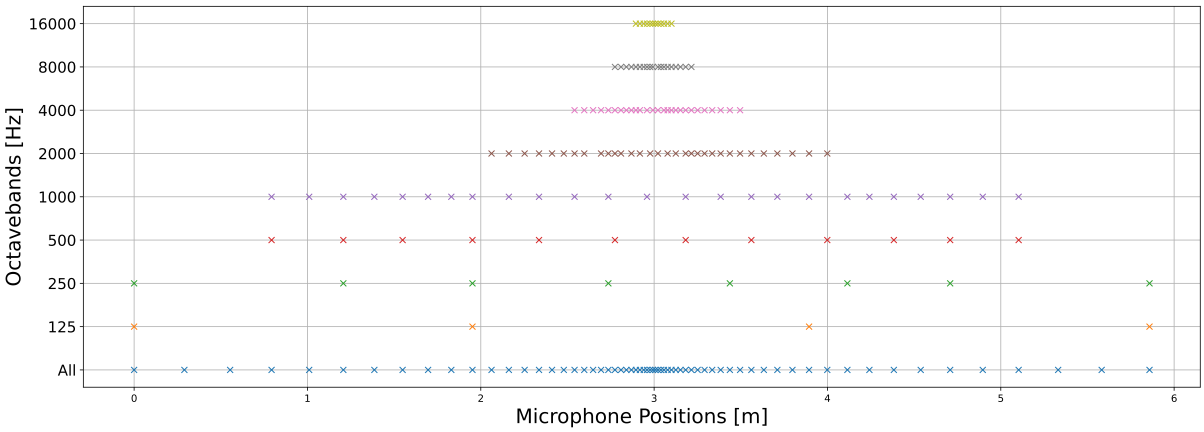
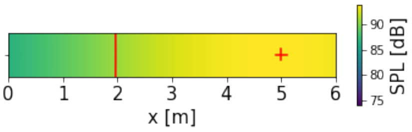


Figure 2: Octave band specific and full array geometries.

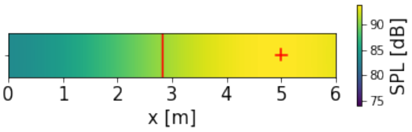
To obtain the relevant frequency information, the aperture of the array is chosen based on the lowest relevant frequency of 80 Hz given by the low E string of the guitar. In general, the aperture for low frequencies should be as large as possible to get a better resolution [6]. Considering the limitation by the width of the anechoic room, an aperture of 5.858 m is chosen. The upper frequency limit is set by the human auditory spectrum at 20 kHz in order to reproduce the individual harmonics of the instruments.

After evaluating the point spread function (PSF) of the entire array, octave band specific array geometries are created by omitting certain microphones in the evaluation. Furthermore, using a logarithmically arranged array with the narrowest microphone distance in the middle and wider distances at the ends makes it possible to choose the amount, position and distance of used microphones per octave band, which shows itself suitable in the simulations. For each of the octave bands from 125 Hz to 16 kHz, an individually simulated geometry is designed by using only selected microphones for beamforming. To do this, the point spread functions are examined along with the previous considerations for aperture, number of microphones, and minimum distances between microphones. For the maximum distance of 6 m with the source located at 5 m, a comparison between the point spread functions of the entire array as well as the octave band specific array geometries investigated for the center frequencies of each octave band can be seen in Fig. 3 and 4.

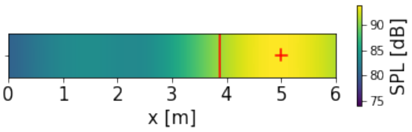
Two main objectives are to enhance the capability of the microphone array to beamform lower frequencies better while still avoiding aliasing for higher frequencies. The tuning of the beamforming for lower frequencies determines how close two sources can be located next to each other. Figure 3a shows that in the whole scope of the microphone array, beamforming for two sources in the lower frequencies would be sub-optimal. However, with the 125 Hz octave band specific array, isolation in certain distances could be achieved as visualized in Fig. 4a. On the other extreme end, the octave band of 16 kHz (Fig. 3i) shows aliasing starting at approximately 1.5 m distance from the source. However, with the 16 kHz octave band specific array seen in Fig. 4i, aliasing was eliminated for most distances from the source and the main lobe is observed to be wider and more defined for the beamforming. The octave band specific array geometries can be seen in Fig. 2.



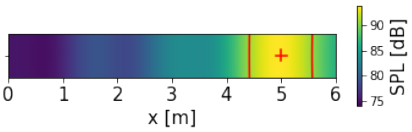
(a) 88 Hz



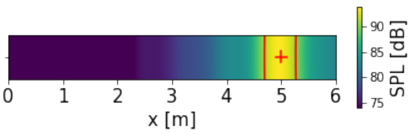
(b) 125 Hz



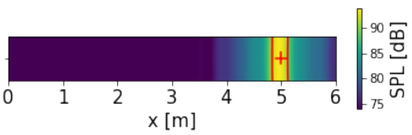
(c) 250 Hz



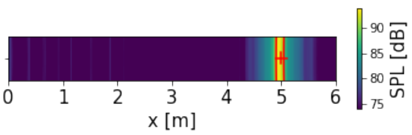
(d) 500 Hz



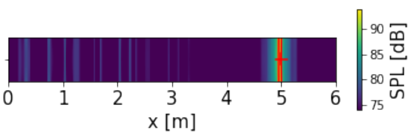
(e) 1 kHz



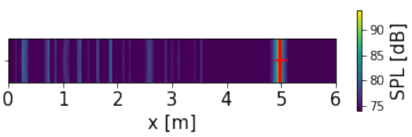
(f) 2 kHz



(g) 4 kHz

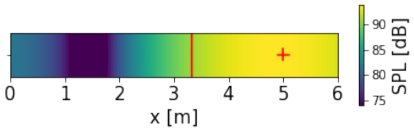


(h) 8 kHz

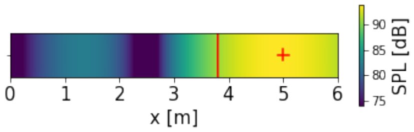


(i) 16 kHz

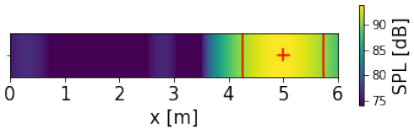
Figure 3: PSF for the center frequency of each octave band using the full array geometry. Red lines mark the beamwidth (-3 dB below max.).



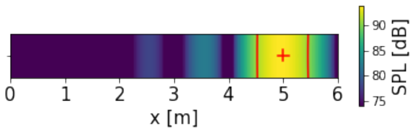
(a) 88 Hz lower cutoff frequency



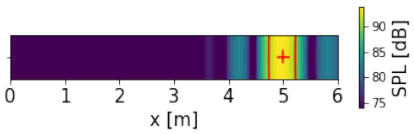
(b) 125 Hz



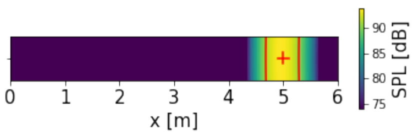
(c) 250 Hz



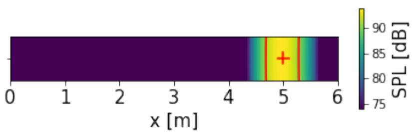
(d) 500 Hz



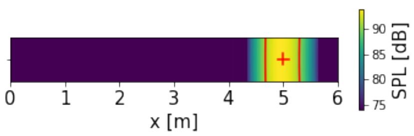
(e) 1 kHz



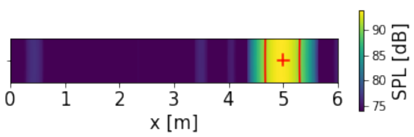
(f) 2 kHz



(g) 4 kHz



(h) 8 kHz



(i) 16 kHz

Figure 4: PSF for the center frequency of each octave band using octave band specific array geometries. Red lines mark the beamwidth (-3 dB below max.).

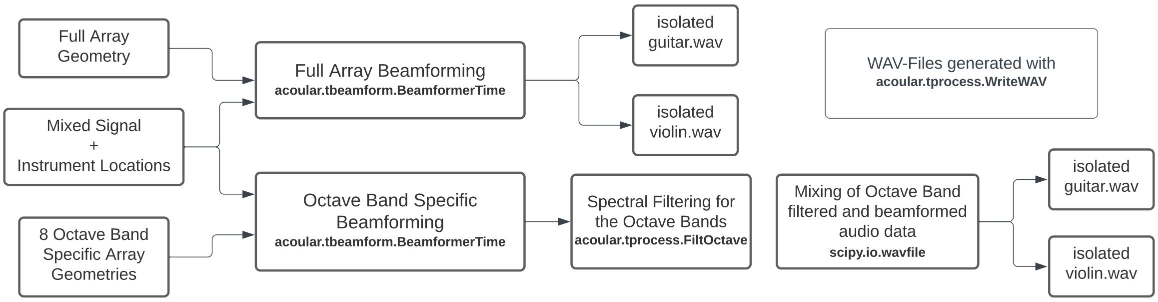


Figure 5: Diagram of the two beamforming methods for two sources.

2.2 Beamforming

For the isolation of audio tracks, a basic time domain beamformer with time signal output for a given spatially fixed grid is used. The focused signal $p(\vec{x_f}, t_f)$ as a function of the focus point $\vec{x_f}$ and the emission time t_f for an acoustic antenna can be determined as [6]:

$$p(\vec{x_f}, t_f) = \frac{1}{M} \sum_{m=1}^{M} p_m(t_f + t_{fm}) w_m r_{fm} / r_{ref}. \tag{1}$$

Equation 1 is the sum of all time-delayed microphone signals ('delay-and-sum') with t_{fm} being the retarded time at the focus point, r_{fm} the distance between the focus position $\vec{x_f}$ and the microphone at the position $\vec{x_m}$ at the time t_f, $p_m(t_f + t_{fm})$ the delayed signal of the microphone m, r_{ref} the reference removal and w_m an optional weight sequence of the microphone signals [6]. Considering potential movement of instrumentalists and the spatial acoustic characteristics of instruments, two 0.3 m wide grids are positioned on each of the central locations of the instruments. The process follows two different methods regarding the utilization of the developed microphone array geometry. The first method, following a basic model, applies beamforming to the whole of the recording with the complete linear microphone array. The second method aims to enhance the isolation and the audio quality of the results by utilizing the individually designed microphone geometries for each of the octave bands.

For the second method, a recording is beamformed for each octave band proportion with the designated microphone array geometry, followed by filtering respective to the octave bands, thus creating eight octave band filtered and separated outputs for each instrument. The multiple outputs for each instrument are then mixed with equal weights using *Scipy* [11] to generate the final separated audio track, spanning the frequency range between 88 Hz and 20 kHz. A diagram for both of the methods can be seen in Fig. 5, in which the fundamental functions are noted under the corresponding segments.

2.3 Simulation

To test and develop the design of the microphone geometries as well as the beamforming methods, a simulation was set up. For the beamforming methods to be applied, recordings of the two instruments were separately simulated as fixed point sources in arbitrary positions. For the

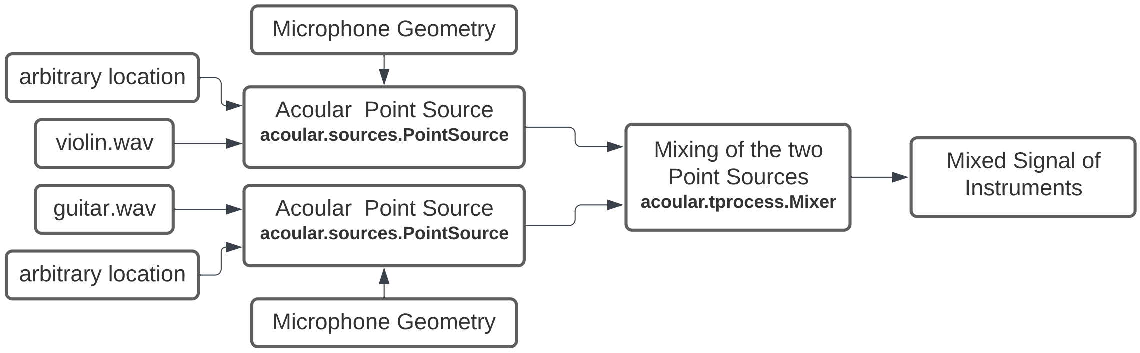


Figure 6: Diagram of the simulated recording with two sources.

input signals of fixed point sources, a WAV audio track was generated in the DAW *Cubase* [9] using the instrument simulator plugin 'Halion Sonic SE 3' with the 'Campfire Guitar' preset for guitar and the presets '[GM 041] Violin' and 'Pizzicato Strings' for the violin following the composed melody for each instrument, which can be seen in section 2.4. Following the designed microphone array geometry, each of the point sources had 64 channel recordings, which were then mixed to create the final recording simulation of two instruments in two different arbitrary positions. A diagram for the simulation of the recordings, along the utilized fundamental functions can be seen in Fig. 6. The simulated positions of the instruments, which followed the positioning of the instrumentalists during the measurements are visualized in Fig. 9. Furthermore, the simulation results obtained by the utilized array geometry and beamforming methods can be seen in Tab. 1, Tab. 2 and Fig. 12a.

2.4 Played melody

The melodies played by the guitar and the violin are shown in Fig. 7. Each bar includes picking and strumming of the guitar, the first bar includes picking from the violin and the second bar stroking with the bow. This way all combinations between picking, strumming and stroking of both instruments are varied to gather comprehensive information.

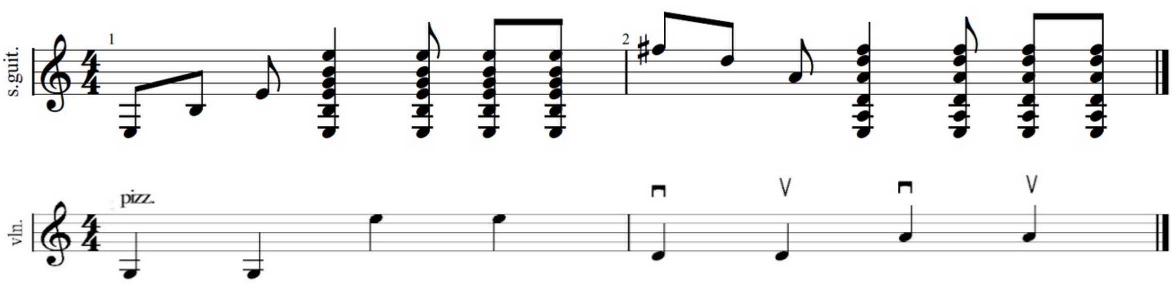


Figure 7: Simulated and played melody of the guitar (top) and the violin (bottom).

3 MEASUREMENT SETUP

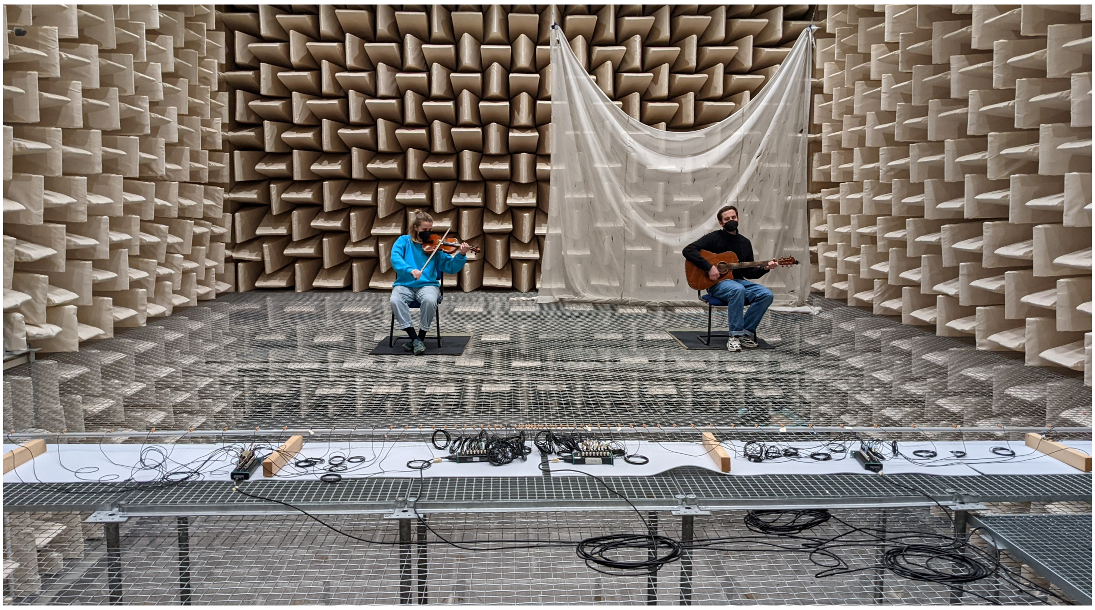


Figure 8: Photograph of the measurement setup.

A photograph of the experiment in the anechoic room at Technical University of Berlin is shown in Fig. 8, the measurement setup is also shown in Fig. 9 schematically. The distance between instruments and array and the distance between instruments themselves was varied between 2 m and 6 m.

To create the array 64 G.R.A.S. 40PK CCP freefield microphones were attached to a 5.585 m metal beam by using 3D-printed mounts. The distances between microphones can be considered accurate to half a millimeter. The microphone array was attached above the metal grid inside the anechoic room and padded with insulation foam underneath and in front of the array to reduce reflections. The musicians were positioned on chairs on top of a carpets, that were moved within the room for each positioning setup.

4 RESULTS

To assess the results of the measurements in the anechoic room three methods will be considered. As the aim of this work is to produce separated audio files that are perceived as sufficient separation of instruments by the human hearing, evaluating the results needs to take human perception into consideration and thus cannot be based solely on functional descriptions such as correlation coefficients. However, there is no previously set numerical point at which the separation of audio tracks is considered ‘good’ or ‘sufficient’. Therefore a combination of three methods - subjective assessment of audio files, correlation coefficients and spectrograms - will be used to discuss the results. The four major findings to be considered in the following are: (1) The best results have been found at the positioning of instruments closest to the array and widest apart between the instruments, (2) similar results for simulations and measurements have

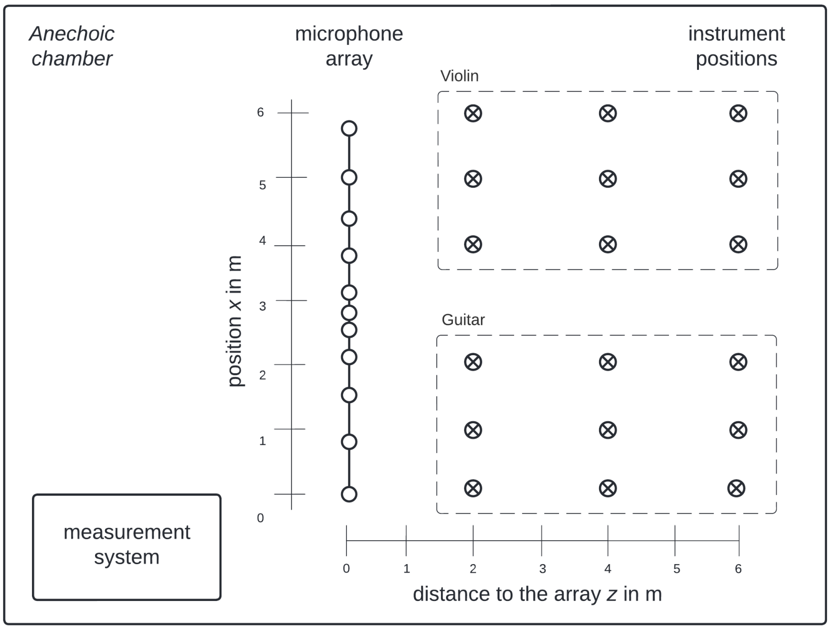


Figure 9: Measurement setup for the recordings in the anechoic room. The instrument positions were varied along both axes on the shown position spots. The recordings that were analyzed followed a symmetric setup with both instruments having the same distance to the array and to the center of the array.

been found with slightly deviating results, (3) the use of octave band specific array geometries enhances the results for the position and distance they were tuned for and (4) the stroking of the violin is very prominent across all recordings and poses thus the most difficulty for audio track isolation.

4.1 Comparing the Simulated and Measured Results with the Full Array Method

The most straightforward way to decide whether the separation of audio tracks by beamforming with the geometry of the full microphone array was successful or not is to listen to the generated audio files and compare the results for the corresponding guitar file and violin file as well as comparing simulation and measurement results for each instrument. The comparison shows that there is a separation of instruments, which can also vary depending on the combination of picking, stroking and strumming. For instance the stroking of the violin is much louder compared to the guitar, which is why it is the most audible part of the violins melody in the separated audio file of the guitar. Regarding the positioning of the sources, the separation with full array beamforming sounds clearer when the instruments are farther apart, and yet closer to the microphone array. However, by listening to the results alone, objective assessments can not be made. Hence, the cross-correlation between the two separated audio tracks are calculated using *Scipy*.

Table 1 shows the correlation coefficients for nine different measurement setups as well as the correlation coefficients yielded from the simulation results beforehand for the full array beamforming method. The cross-correlation between corresponding guitar and violin files shows

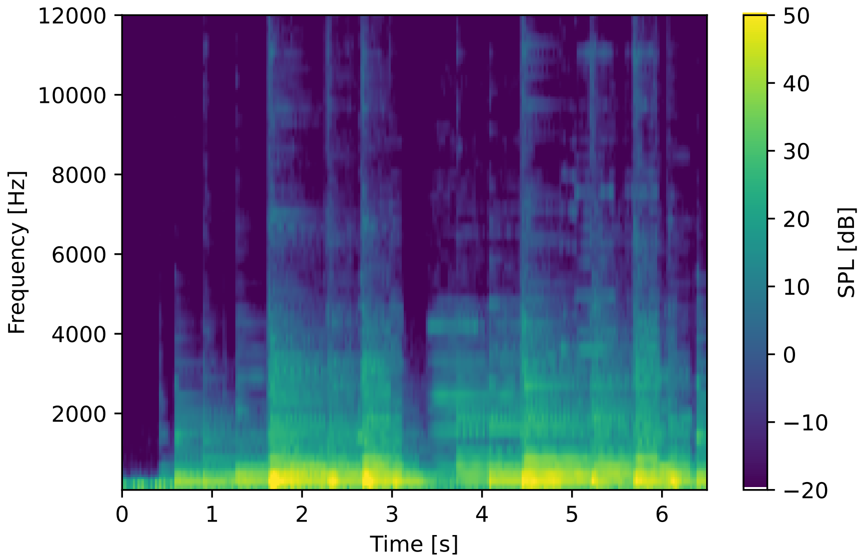
x [m] \ z [m]	2	4	6
guitar 2; violin 4	0.38 (0.44)	0.60 (0.55)	0.65 (0.66)
guitar 1; violin 5	0.24 (0.23)	0.32 (0.36)	0.43 (0.47)
guitar 0; violin 6	0.16 (0.18)	0.23 (0.26)	0.38 (0.33)

Table 1: Full Array Beamforming correlation coefficients for different values of x (distance between the instruments) and z (distance from the instruments to the microphone array). Values in brackets indicate simulation results.

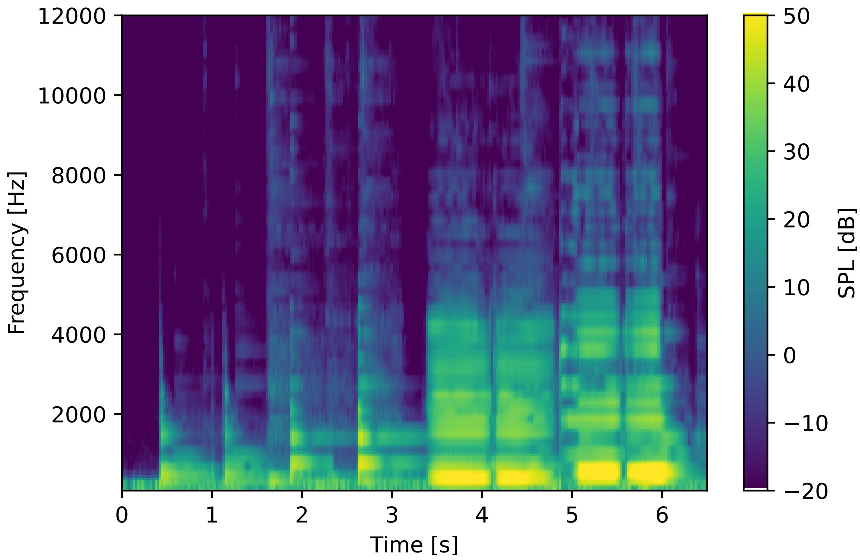
how identical the separated audio files are in average for the total duration of signals. A low correlation coefficient means a smaller portion of the opposing instrument is found in the separated audio file of the other instrument, which hints to better separation of instruments. A strong tendency for decreasing correlation coefficients can be found towards greater distance between the instruments and smaller distance to the microphone array, which confirms the subjective impressions from listening to the audio files. Following this tendency, the lowest correlation in the recordings was obtained for instruments with 6 m distance between each other and 2 m distance to the array. Comparing the recording results with the simulations, the same tendency can be found, which confirms the reliability of the results, however, the measurement correlations are overall lower than the simulation results by 1-9 %.

Additionally, to observe and compare the separation of instruments over time by their spectral attributes, spectrograms are generated from the separated audio tracks as well as the unseparated tracks. Figure 11 shows the unisolated signal between 80 Hz and 12 kHz recorded by one of the microphones at the center of the array. Both the guitar and violin signals are included in this spectrogram and a lot of noise is visible. Figure 10 shows spectrograms for both instruments between 80 Hz and 12 kHz after isolation. The signal-to-noise ratio is increased and it is clearly deductible from the spectrograms when which instrument was played: the picking and stroking is visible for each instrument. For example, we can see that between seconds 1 and 2, the guitar played a faster picking pattern before entering a slower strumming pattern between seconds 2 and 4. This pattern is repeated from second four. The violin played a slow picking pattern between seconds 1 and 4 and then proceeded with a stroking pattern in the same rhythm. Clear differences between each instrument's pattern are noticeable from which it can be visibly concluded as well, that the audio tracks have been separated.

However, a closer inspection of the spectrograms reveals that fragments of each instrument can still be found in the opposing spectrogram. For example, just after second 2 the violin's spectrogram shows a touch which cannot be linked to the melody the violin played. This is clearly a fragment from the guitar's playing. As mentioned previously, especially the stroking of the violin is very noticeable across all recordings. This is mirrored in the spectrograms as well, as the guitar spectrogram turns slightly more blurry after second four. The stroking of the violin distorts the image of the guitar and strumms/picks become less clear.



(a) Guitar



(b) Violin

Figure 10: Exemplary time-beamformed spectrograms for (a) Guitar and (b) Violin between 80 Hz and 12 kHz using the whole microphone array geometry. The distance between microphone array and instruments was $z = 2$ m, the instruments were positioned at $x_g = 0$ m and $x_v = 6$ m. The musicians started playing at second 1; the guitar started with three picking notes (between seconds 1 and 2) followed by strumming (between seconds 2 and 4) and then starting over again; the violin was played picking four notes (between seconds 1 and 4) followed by four notes of stroking (from second 4).

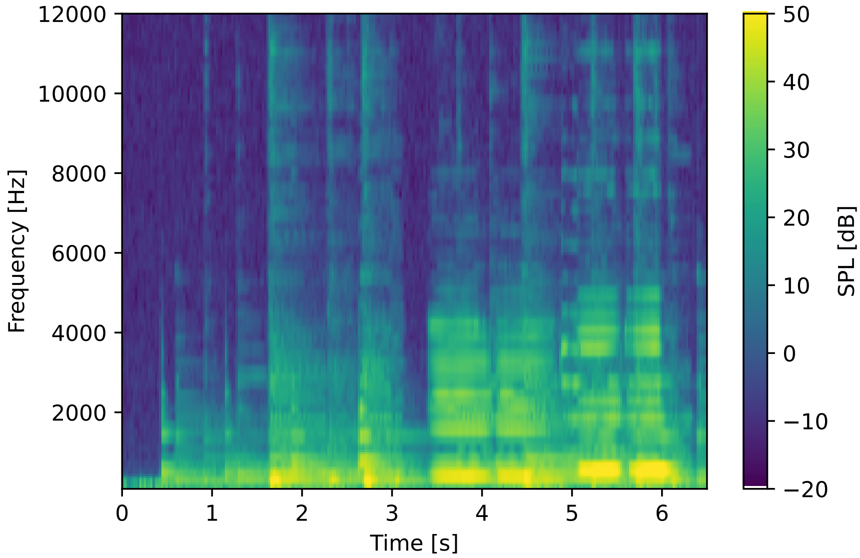


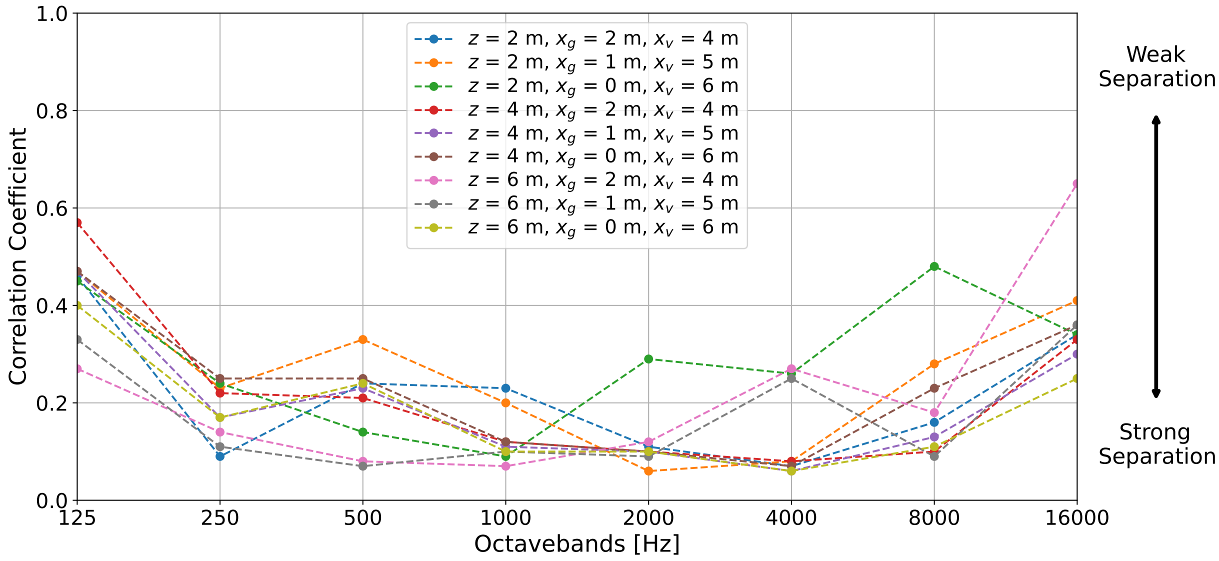
Figure 11: Exemplary time-beamformed spectrogram between 80 Hz and 12 kHz for the unisolated recording of one microphone $x = 3.007$ m in the center of the array at a distance of $z = 2$ m between instruments and array and the instrument positions at $x_g = 0$ m and $x_v = 6$ m.

4.2 Comparing the Simulated and Measured Results with the Octave Band Specific Array Method

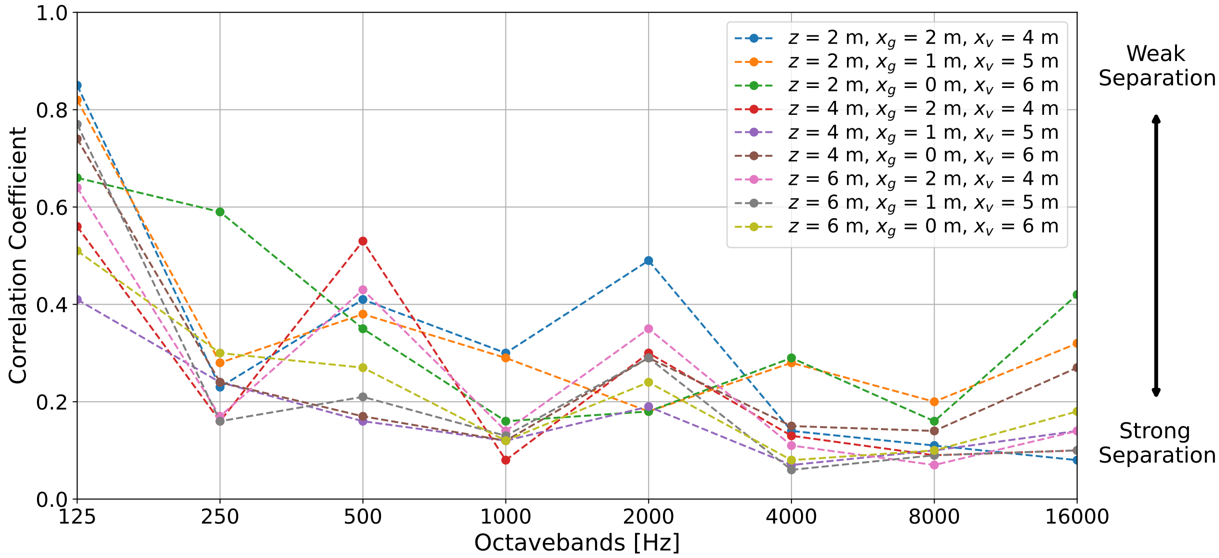
Listening to the mix of octave band specific beamforming results, it can be heard that the mixed signal has a higher timbre than the full array beamforming. This might be due to using octave band filtering as the lowest frequency in the mix is the lower frequency limit for the 125 Hz octave band (88 Hz). Regarding the separation for the mixed signals of each instrument, a better separation for the instruments is not realized for the distance of 2 m (where the full array beamforming yielded its best results). However, with increasing distance - when the instruments are 4 m and 6 m away from the microphone array - the separation sounds clearer for beamforming with octave band specific arrays. Table 2 shows the cross-correlation coefficients for the simulation and measurement results that were obtained by the use of octave band specific

x [m] \ z [m]	2	4	6
guitar 2; violin 4	0.38 (0.16)	0.28 (0.15)	0.30 (0.21)
guitar 1; violin 5	0.31 (0.25)	0.18 (0.17)	0.21 (0.10)
guitar 0; violin 6	0.31 (0.23)	0.18 (0.20)	0.19 (0.20)

Table 2: Octave band beamforming correlation coefficients for different values of x (distance between the instruments) and z (distance from the instruments to the microphone array). Values in brackets indicate simulation results.



(a) Simulation



(b) Recordings

Figure 12: Correlation coefficients for octave band beamforming for the simulation (top) as well as recordings (bottom) at different positions.

beamforming. At a distance of 2 m, the correlations are higher than the full array beamforming, however, at 4 m and 6 m distances, the results are indeed enhanced as the correlations are lower than the full array beamforming. As the octave band specific arrays have been tuned specifically for the maximum distance of 6 m and 4 m between the instruments, highest separation improvement is observed in this distance. In the simulation results, the lowest correlation is found for this position as well, which shows that tuning of the octave band specific geometries is considerably location-oriented. Most of the simulation correlation results are lower than the measurements, and they deviate most for positions where the instruments are closest to each other. This deviation is heavily related to the difference between virtual point sources and real instruments, which will be further discussed in section 4.4.

To be able to compare the performance of octave band specific beamforming for each individual octave band, the correlation coefficients per octave band are visualized in Fig. 12. A general tendency towards lower correlation is found for the simulation results compared to recording results. For the recordings, the highest correlation is consistently calculated for the lowest octave band of 125 Hz, whereas the lowest correlation can be found towards the higher frequencies. For the measurement results the correlation coefficients vary strongly depending on the position of the instruments, and we find the lowest correlations overall at the octave bands 1 kHz and 4 kHz. Although there is no position with the lowest correlation in every octave band, lower correlations for octave bands are observed consistently for 4 m and 6 m distance, which goes in line with the results observed in Table 2.

4.3 Comparing Full Array Beamforming and Octave Band Specific Beamforming

As mentioned before, from listening to generated audio files, a better separation was obtained by using the full array geometry in the distance of 2 m instead of octave band specific array geometries. However, an enhanced separation was realized with octave band specific beamforming in greater distances, with the location of tuning having the best simulated result and enhancement. For a better comparison of the performance between the two methods, the correlation coefficients for beamforming using the full array geometry were calculated for individual octave bands as well. The correlation coefficients for both simulation and measurements in the

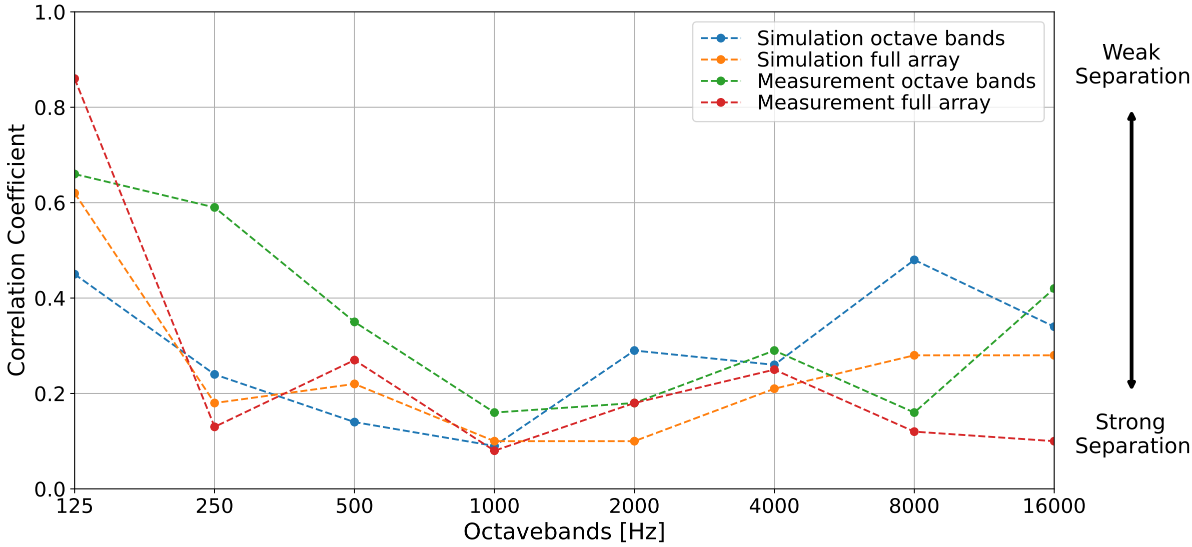


Figure 13: Comparison of the two beamforming methods with an octave band breakdown for the measurement and simulation of $z = 2$ m, $x_g = 0$ m and $x_v = 6$ m.

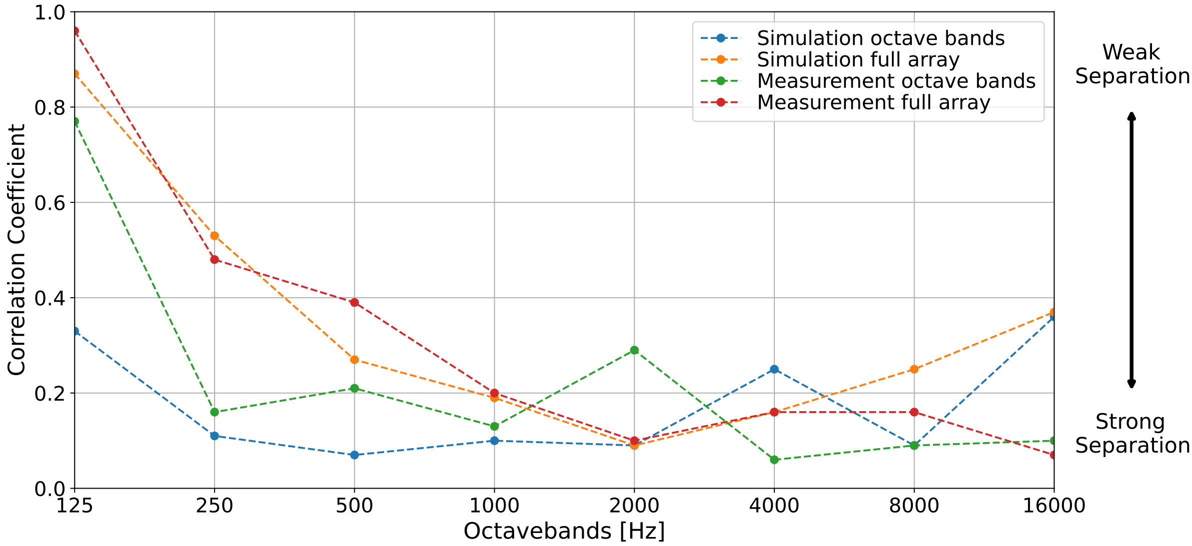


Figure 14: Comparison of the two beamforming methods with an octave band breakdown for the measurement and simulation of z = 6 m, x_g = 1 m and x_v = 5 m.

'best' position for full array beamforming at z = 2 m, x_g = 0 m and x_v = 6 m can be seen in Fig. 13, as well as the correlation coefficients for the 'best' position for octave band specific beamforming at z = 6 m, x_g = 1 m and x_v = 5 m in Fig. 14.

Taking a look at Fig. 13, the correlation coefficients for the octave bands from 250 Hz to 1 kHz have lower values for simulated octave band specific beamforming results, yet the measurement deviates strongly and delivers a weak separation in these octave bands. It is worth noting, that the correlation for the lowest octave band of 125 Hz is enhanced with the use of octave band specific arrays. For the position, where the octave band specific arrays where tuned for (see Fig. 14), it is clear that the octave band specific beamforming delivers superior results both in simulation and measurement for the octave bands except for some deviations in the octave bands of 2 kHz and 4 kHz. Best enhancement is observed for the first four octave bands from 125 Hz to 1 kHz.

Considering these results, octave band specific beamforming shows promise and further tuning the octave band specific arrays will enhance the separation in each octave band improving the overall separation for each instrument's mixed signal. However, octave band specific array beamforming only shows improvement compared to the full array beamforming in a certain perimeter from the position of tuning. In our results, when this perimeter exceeds a distance of 4 m, so when the instruments are 2 m away from the microphone array, full array beamforming shows better results. Thus, if results in a distance of 2 m are to be improved, a second set of octave band specific microphone geometries which are tuned better for this distance should be considered.

4.4 Error Minimisation

For both beamforming methods and all of the octave bands, minor to significant deviation is observed between the simulation and measurement results. These deviations can be traced back to the limitations of the recording setup as well as the difference between virtual and real sound sources. Real instruments have a volume and unlike the sources simulated beforehand, they cannot be considered as a source without expansion. On top of that, the sound emission of an instrument varies across the whole area of the instrument with certain areas emitting a large part of the sound but not all of it, as sound is also emitted from the rest of the instrument's corpusses. Within the developed algorithm spatially fixed grids were manually placed on the estimated position of the sound hole of the guitar and the f-holes of the violin, as these parts of the instruments emit a large part of the sound. Taking the imprecision of the instruments' positions into account an offset along the x-axis of the fixed grid during and between the recordings is inevitable. The directionality of the instruments also plays a part, as the virtual point sources used for beamforming are simulated as omnidirectional, whereas the actual instruments are directional sources. Another possible cause for errors are reflections within the anechoic room caused by human bodies and the instruments.

Several of the above mentioned errors are caused by factors which can not be avoided or can only be minimised to some extent. Especially if the setup is to be used for an increasing number of instruments, some of these factors will not only remain but possibly increase. Regarding the algorithm, one option to improve both methods used in this research could be the use of an iterative beamforming sequence, where in short intervals beamforming in the frequency domain is conducted to find current sound source maxima, followed by a second run of beamforming in the time domain to isolate the audio tracks using the sound source locations according to the just ascertained sound source maxima. The CLEANT method could act as a useful approach here as with the known movement of the sound source a better separation of the time signals could be achieved [1].

5 CONCLUSION

In summary, it was shown that the separation of the guitar and violin worked successfully by beamforming the recordings with the full linear array geometry. The array geometry, limited by the measurement room, the microphone holder and the number of microphones, has shown that a line array is sufficient to separate the two instruments. It is worth mentioning in the determination of the quality that the separated audio files were evaluated with the human hearing perception, the cross-correlation of the separated audio files and the use of the spectrograms. Although the quality of separation can be quantified with cross-correlation coefficients, the quality and sufficiency of the separated signals could not be evaluated by scientific means. A suggestion for evaluating the results further would be a listening test with a group of test subjects. The best performance with the full array beamforming method is achieved if the instruments are closest to the array but furthest away from each other. The attempt to improve the instrument separation results with octave band specific array geometries proved to be a success, while showing that its performance is location oriented. It was observed that the octave band specific arrays work best in enhancing the results when closest to the position they are tuned for. Taking this into account, multiple sets of octave band specific microphone array geometries tuned for different positions can be considered as an outlook. While the current octave band specific array geometries are open for optimisation, use of a more advanced beamforming method such as CLEANT could be an additional improvement to minimise the errors that are caused by instrument positioning and acoustic sound source properties.

References

[1] R. Cousson, Q. Leclère, M.-A. Pallas, and M. Bérengier. "A time domain clean approach for the identification of acoustic moving sources." *Journal of Sound and Vibration*, 443, 47–62(12), 2019. URL `https://doi.org/10.1016/j.jsv.2018.11.026`.

[2] J. Jensen, S. Karimian-Azari, M. Christensen, and J. Benesty. "Harmonic beamformers for speech enhancement and dereverberation in the time domain." *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, 18, 602–613, 2018. URL `https://doi.org/10.1016/j.specom.2019.11.003`.

[3] T. Jiang, L. Hongqing, L. Gan, and Y. Zhou. "A complex neural network adaptive beamforming for multi-channel speech enhancement in time domain." 2021. URL `https://doi.org/10.13140/RG.2.2.31690.08649`.

[4] D. Jimenez, N. Vardenberg, L. Miranda, M. Yadav, and D. Cabrera. "Auralisation of stage acoustics for large ensembles." In *Proceedings of Acoustics 2013 - Victor Harbour*. 2013.

[5] Y. Luo. "A time-domain generalized wiener filter for multi-channel speech separation." 2021. URL `https://doi.org/10.48550/arXiv.2112.03533`.

[6] M. Möser and U. Michel. "Akustische Antennen." In *Messtechnik der Akustik*. Springer-Verlag, 2010.

[7] E. Sarradj and G. Herold. "A python framework for microphone array data processing." *Applied Acoustics*, 116, 50–58, 2017. URL `https://doi.org/10.1016/j.apacoust.2016.09.015`.

[8] P. Sijtsma, S. Oerlemans, and H. Holthusen. "Location of rotating sources by phased array measurements." *7th AIAA/CEAS Aeroacoustics Conference*, 2001. URL `https://doi.org/10.2514/6.2001-2167`.

[9] Steinberg Media Technologies GmbH. "CUBASE PRO 11 (version: 11.0.0).", 2020. URL `https://www.steinberg.net/de/`.

[10] T. Takaishi, K. Yamamoto, T. Kobayashi, and T. Yokota. "Localization of noise sources around aircraft in flight based on time-domain beamforming technique." *Speech Communication*, 116, 1–11, 2020.

[11] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, et al. "Scipy 1.0: fundamental algorithms for scientific computing in python." *Nature methods*, 17(3), 261–272, 2020.

[12] Z. Zhong, M. Shakeel, K. Itoyama, K. Nishida, and K. Nakadai. "Assessment of a beamforming implementation developed for surface sound source separation." In *2021 IEEE/SICE International Symposium on System Integration (SII)*, pages 369–374. 2021. URL `https://doi.org/10.1109/IEEECONF49454.2021.9382648`.