BeBeC-2026-D09

INVESTIGATION OF HIGH-RESOLUTION MICROPHONE ARRAY METHODS FOR SEPARATING SIMULTANEOUSLY PLAYING STRING INSTRUMENTS

Ivo Hagenmaier[1], Mikolaj Czuchaj[1], Gert Herold[2] and Ennes Sarradj[1]
[1]Technische Universität Berlin, FG Technische Akustik, Einsteinufer 25, 10587 Berlin, Germany
[2]German Aerospace Center (DLR), Institute of Propulsion Technology,
Engine Acoustics Department, Bismarckstraße 101, 10625 Berlin, Germany

Abstract

Microphone arrays are used to spatially image or record individual sound sources while minimizing contributions from nearby interfering sources. The separation of simultaneously playing instruments poses a particular challenge, but is of practical interest as an alternative to the labor-intensive individual recording of each instrument with clip-on microphones in music production. This research investigates the extent to which a microphone array can be used to record a small ensemble and subsequently separate the individual instruments from one another. Prior to the measurement, an investigation was conducted to determine which array geometry is best suited for this application. The evaluation was based on a comparison of various geometries by analyzing the leakage in the respective other instrument regions. The measurement took place in the anechoic chamber of the Technische Akustik Prüfhalle at TU Berlin. The instruments under investigation were two violins, a cello, and a double bass. Using frequency-domain beamforming and the CLEAN-T algorithm, separate audio tracks for each instrument were extracted from a synthetic ensemble recording, constructed by combining the individual solo recordings. This allowed the separation performance to be evaluated against the known single-instrument reference signals, with the leakage, i.e. contributions from the other instruments, serving as the central evaluation criterion.

1 INTRODUCTION

The recording of acoustic musical instruments in live performance settings traditionally relies on close-microphone techniques, where individual clip-on or spot microphones are placed

directly on or near each instrument. While this approach provides high signal quality and isolation, it introduces practical challenges: the physical attachment of microphones to instruments can interfere with the performer's playing comfort, alter the acoustic radiation of the instrument, and require extensive setup time prior to each performance. Furthermore, managing the signal routing for large ensembles such as orchestras adds significant technical complexity.
A particular challenge in orchestral monitoring is that acoustic instruments — unlike electric guitars or keyboards — do not provide a direct line output. Consequently, each instrument or instrument group requires its own dedicated microphone for real-time monitoring, a setup that musicians in live bands and stage productions rely upon to hear themselves and their fellow performers clearly [2]. A fixed microphone array installed in the concert hall could offer an alternative: by applying beamforming algorithms to exploit the spatial information captured across the microphone channels, individual instruments can be isolated based on their position in the room alone, eliminating the need to equip each concert individually with close microphones.
Beyond monitoring, such an approach offers further advantages for recording and post-production. Unlike proximity microphone recordings, a spatially distributed array captures directional and spatial information of the sound field, potentially yielding a more natural and immersive sound [1]. All instrument contributions could furthermore be edited individually in post-processing, providing considerable flexibility in the mixing stage.
Microphone arrays combined with beamforming algorithms represent a well-established methodology for sound source localization and signal separation [10]. Frequency-domain delay-and-sum beamforming has been widely applied to acoustic source localization and sound power quantification, while time-domain methods directly reconstruct the emitted time signals and have demonstrated suitability for aeroacoustic measurements and moving source localization [4, 5]. The open-source framework Acoular [9], developed at the Technical University of Berlin, implements both approaches and serves as the computational basis for this work. Related work includes beamforming-based audio object separation from mixed recordings [3]; however, the application of array beamforming specifically to the recording and separation of orchestral instruments has not yet been explored systematically. The present study therefore investigates the extent to which a three-dimensional microphone array combined with frequency-domain delay-and-sum beamforming and the CLEAN-T algorithm can be used to separate and individually reconstruct the signals of string instruments from a synthetically combined recording of individual performances.

2 METHODS

2.1 Array Geometry

The array geometry used in this experiment was selected by evaluating the frequency-dependent leakage between all instrument positions for several different geometries. For this purpose, a monopole source was placed at the estimated position of one of the two f-holes of each of the five planned instruments (two violins, one viola, one cello, and one double bass). For each candidate geometry, the point spread function (PSF) was computed for every instrument position and subsequently normalized to its own maximum value. For a given target instrument, the normalized PSFs of the four remaining instruments were summed pointwise within a spherical

region of radius $r = 0.25$ m centered on the target instrument position, with a grid spacing of 0.025 m along each spatial axis. The maximum value of this summed interference field within the sphere was then divided by the maximum value of the normalized target PSF within the same region, and the resulting ratio was expressed in decibels as the leakage measure L:

$$L = 10 \cdot \log_{10} \left(\frac{\max(PSF_{\text{others}})}{\max(PSF_{\text{target}})} \right) \tag{1}$$

This procedure was repeated for all five instruments as the target and for all frequencies of interest. The resulting leakage values were summed across all instruments and additionally evaluated in octave bands from 125 Hz to 8000 Hz. Candidate array geometries were then ranked based on these octave-band leakage values, with lower leakage indicating superior spatial separation performance.

All candidate array geometries consisted of 95 microphones and were either two-dimensional or three-dimensional to enable source localization in depth. These included different arrangements of two or three line arrays, similar to the single line-array approach used in [6], as well as Vogel spiral arrangements. The latter were found to be particularly promising. Following the Vogel spiral [11] and a generic approach by Sarradj for synthesizing optimal microphone arrangements [7], the radial distance r and the azimuthal angle ϕ of the m-th microphone are given by:

$$r = R\sqrt{\frac{m}{M}}, \quad m = 1, 2, \ldots, M \tag{2}$$

$$\phi = 2\pi m \frac{(1+\sqrt{V})}{2} \tag{3}$$

where R is the array radius, M the total number of microphones, and V a free parameter controlling the shape of the spiral.

The final array geometry used in the measurement, consisting of contributions from two different Vogel spirals, was constructed as follows. A Vogel spiral of 161 microphones with radius $R = 2$ m and $V = 5$ was used as the basis for the two side walls. A rectangle was iteratively built inside this spiral until 63 microphones fitted within it, resulting in final dimensions of a width of 2.74 m and a height of 1.80 m. The microphones were then moved to the nearest 3D-printed clips mounted on the standing metal mesh walls. The walls were subsequently split in the middle and each half was repositioned: the left wall center point was placed at $x = -1.75$ m, $y = 0.362$ m and $z = 1.195$ m with a 90° counter-clockwise rotation, and the right wall center point at $x = +1.75$ m, $y = 0.405$ m and $z = 1.195$ m with a 90° clockwise rotation, resulting in 32 microphones on the left side and 31 on the right.

For the top section, parts of a second Vogel spiral from [8] were used. This spiral was divided into six parts, of which Q2R and Q3R, consisting of 32 microphones in total, were selected as they yielded the lowest leakage. The final geometry with the approximate position of one f-hole per instrument can be seen in Figure 1.

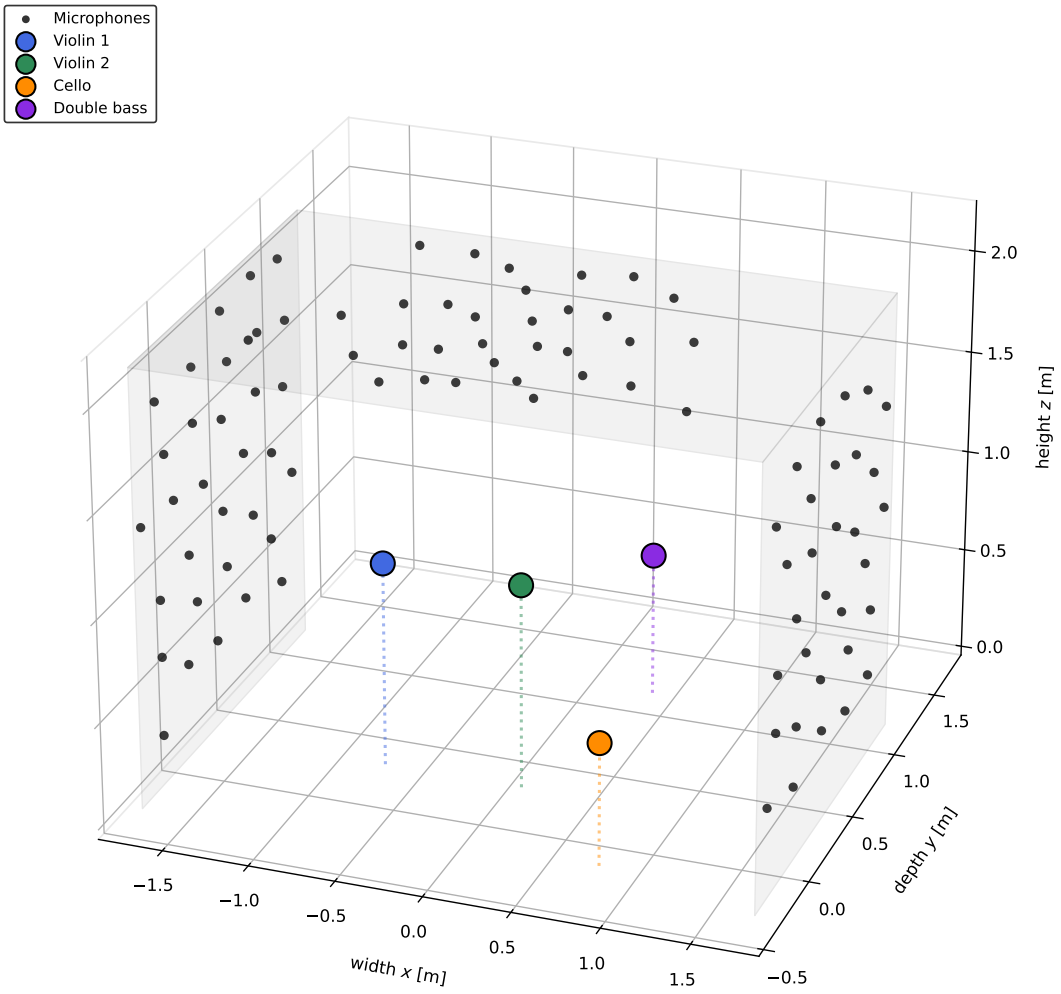

Figure 1: 3D plot of the array geometry consisting of 95 microphones and the four instruments.

2.2 Frequency-domain delay-and-sum beamforming

For the localization of each instrument, frequency-domain delay-and-sum beamforming was applied using the `BeamformerBase` implementation of the Acoular framework [9]. The beamformer steers the array response to a set of focus points on a predefined spatial grid and evaluates the frequency-dependent sound pressure level at each grid point, allowing the identification of source positions as local maxima in the beamforming map.

2.3 CLEAN-T Algorithm

The CLEAN-T method proposed by Cousson et al. [4] is an iterative deconvolution algorithm that aims to remove the influence of the point spread function from the delay-and-sum beamforming result, yielding a cleaner spatial representation of the source field. While originally

proposed for moving sources, it is applied here to the stationary case, as the instruments remain fixed during the measurement.
In a first step, the beamforming map $\Phi^{(0)}$, the clean representation $\Gamma^{(0)}$, and the residual microphone signals $p_m^{res(0)}$ are initialized:

$$\Phi^{(0)}(t,x_g) = b\left(t,x_g,\{p_m\}\right) \tag{4}$$

$$\Gamma^{(0)}(t,x_g) = 0 \tag{5}$$

$$p_m^{res(0)}(t) = p_m(t) \tag{6}$$

In each subsequent iteration i, the grid point with the highest energy is identified as the dominant source location $\hat{x}_g$:

$$\hat{x}_g = \underset{x_g}{\operatorname{argmax}} \left(\int_0^T \left| \Phi^{(i-1)}(t,x_g) \right|^2 dt \right) \tag{7}$$

The beamformed signal at this location is then added to the clean representation, scaled by the loop gain factor γ:

$$\Gamma^{(i)}\left(t,\hat{x}_g\right) = \Gamma^{(i-1)}\left(t,\hat{x}_g\right) + \gamma\, \Phi^{(i-1)}\left(t,\hat{x}_g\right) \tag{8}$$

The microphone signals corresponding to this source contribution are then modelled and subtracted from the residual microphone signals:

$$p_m^{res(i)}\left(t+\frac{r_{x_g m}}{c}\right) = p_m^{res(i-1)}\left(t+\frac{r_{x_g m}}{c}\right) - \gamma \frac{\Phi^{(i-1)}\left(t,\hat{x}_g\right)}{r_{\hat{x}_g m}(t)\left(1-M\cos\theta_{\hat{x}_g m}(t)\right)^2} \tag{9}$$

Since the sources are stationary, $r_{\hat{x}_g m}$ is time-invariant and $M = 0$, such that the term $(1 - M\cos\theta)^2$ reduces to unity.
Finally, beamforming is performed on the updated residual microphone signals to obtain the next iteration map:

$$\Phi^{(i)}\left(t,x_g\right) = b\left(t,x_g,p_m^{res(i)}\right) \tag{10}$$

This process is repeated until the termination condition is reached, which can be defined either as a fixed number of iterations or as an increase in energy within the residual beamforming map. All computations were performed using the *Acoular* framework [9].

3 MEASUREMENT

A photograph of the experiment in the anechoic room at the Technische Universität Berlin is shown in Figure 2, the measurement setup is also shown schematically in Figure 1. The array consists of 95 G.R.A.S. 40PK CCP free-field microphones attached to 3D-printed microphone mounts clipped to the metal mesh walls, which are mounted on the metal grid floor of the anechoic room. The top microphone mounts are attached to aluminum profile bars fixed on two round metal rods, which are in turn attached to the metal mesh walls. The positions of the

microphones are considered accurate to within one centimeter. The recordings were made at a sampling rate of 51200 Hz.
Three of the musicians (violins and cello) were seated on chairs placed on top of carpets, while the double bass player was standing with the double bass endpin resting on a small 0.15×0.15 m rubber mat. The four instruments are roughly symmetrical about the center line ($x = 0$ m), with the two violins and the cello positioned in the front line ($y = 0$ m) and the double bass offset in the back ($y = 1$ m). This arrangement results from the fact that a viola was originally planned for the back line next to the double bass, but the viola player was not available for the measurement.

Figure 2: Photograph of the measurement setup with the four instrumentalists inside the anechoic room at the Technische Universität Berlin.

Each of the four instruments — two violins, a cello, and a double bass — bowed a note of the G major chord in succession. First, violin V1 played a *G4*, followed by violin V2 with a *D4*, then the cello C with a *B3*, and finally the double bass DB with a *G2*.

4 RESULTS

4.1 Spectrogram of the measurement

Figure 3 shows the spectrogram of the measurement with each instrument playing in succession, averaged over all 95 microphones. The frequency resolution is $\Delta f = 51200$ Hz $/$ $4096 = 12.5$ Hz. The four colored areas, each with a duration of 2.75 s, mark the time intervals during which each instrument plays solo and are used for all subsequent analyses.

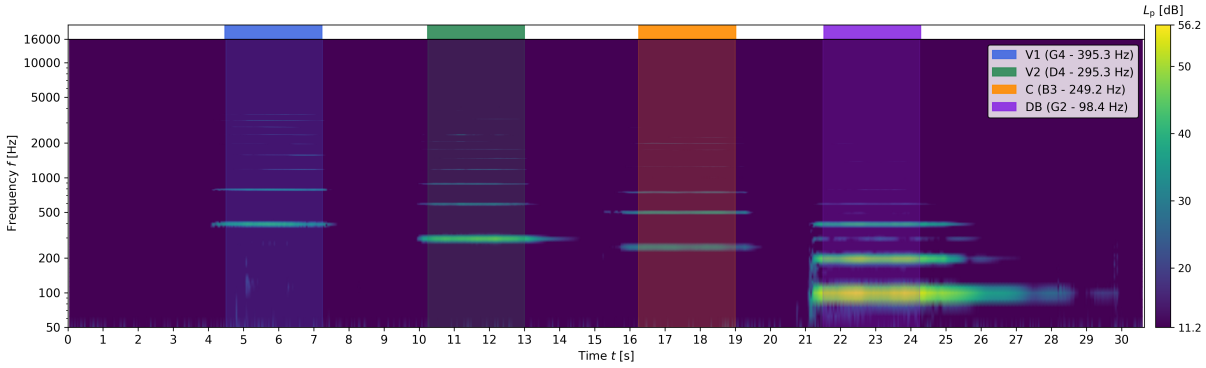

Figure 3: Spectrogram of the measurement averaged over all 95 microphones.

The results are summarized in Table 1. The average sound pressure levels (SPL) during the individual playing sections were highest for the double bass, despite it playing the lowest note of all four instruments, more than two octaves below violin V1.

Instrument	Color	Played note	Fundamental frequency	Average SPL L_p	Time interval
Violin V1	blue	G4	395.3 Hz	67.8 dB	4.50 s – 7.25 s
Violin V2	green	D4	295.3 Hz	69.2 dB	10.25 s – 13.00 s
Cello C	orange	B3	249.2 Hz	67.6 dB	16.25 s – 19.00 s
Double bass DB	purple	G2	98.4 Hz	72.2 dB	21.50 s – 24.25 s

Table 1: Summary of the measured instrument parameters.

For further analysis, a synthetic chord was constructed by combining the four colored time intervals of the individual solo recordings. For each of the 95 microphone channels, these time intervals were added samplewise without any level adjustment, resulting in a synthetic multichannel signal representing all four instruments playing simultaneously. This synthetic signal was used as input for the CLEAN-T algorithm, which was configured with a loop gain of $\gamma = 0.6$ and a maximum of 100 iterations as the termination criterion.

4.2 Localization of frequency-dependent maxima

For localization, the peaks of the frequency spectra of each solo measurement, averaged over all 95 microphones and the 2.75 s interval seen in Figure 3, were analyzed. The peaks were identified with a minimum level of 20 dB for the violins and the cello, and 30 dB for the double bass due to its higher sound pressure level. For all instruments a prominence of 20 dB relative to neighboring frequency points was required. With these conditions, the violins each had 16 peaks ranging from 395.3 Hz to 6323.4 Hz (V1) and 295.3 Hz to 5013.3 Hz (V2), the cello had 19 peaks from 249.2 Hz to 4740.6 Hz, and the double bass 23 peaks from 98.4 Hz to 2268 Hz. First, the beamforming maxima for the identified peak frequencies were searched within an initial rectangular grid of ± 0.30 m on each axis around the roughly measured f-hole position of each instrument. The rounded mean position of the initially found maxima was then used

Instrument	dx [m]	dy [m]	dz [m]	Grid spacing [m]
Violin V1	±0.35	±0.35	±0.35	0.025
Violin V2	±0.35	±0.35	±0.35	0.025
Cello C	±0.35	±0.45	±0.50	0.025
Double bass DB	±0.60	±0.60	±0.60	0.025

Table 2: Search grid dimensions for each instrument.

to define instrument-specific search grids, whose dimensions are listed in Table 2. Frequency-domain beamforming was then repeated within these refined grids to determine the final maxima and the instrument-specific mean position. The maxima, mean positions, and search grids are shown for three projection planes in Figure 4, where the gray area marks the microphone array boundary.

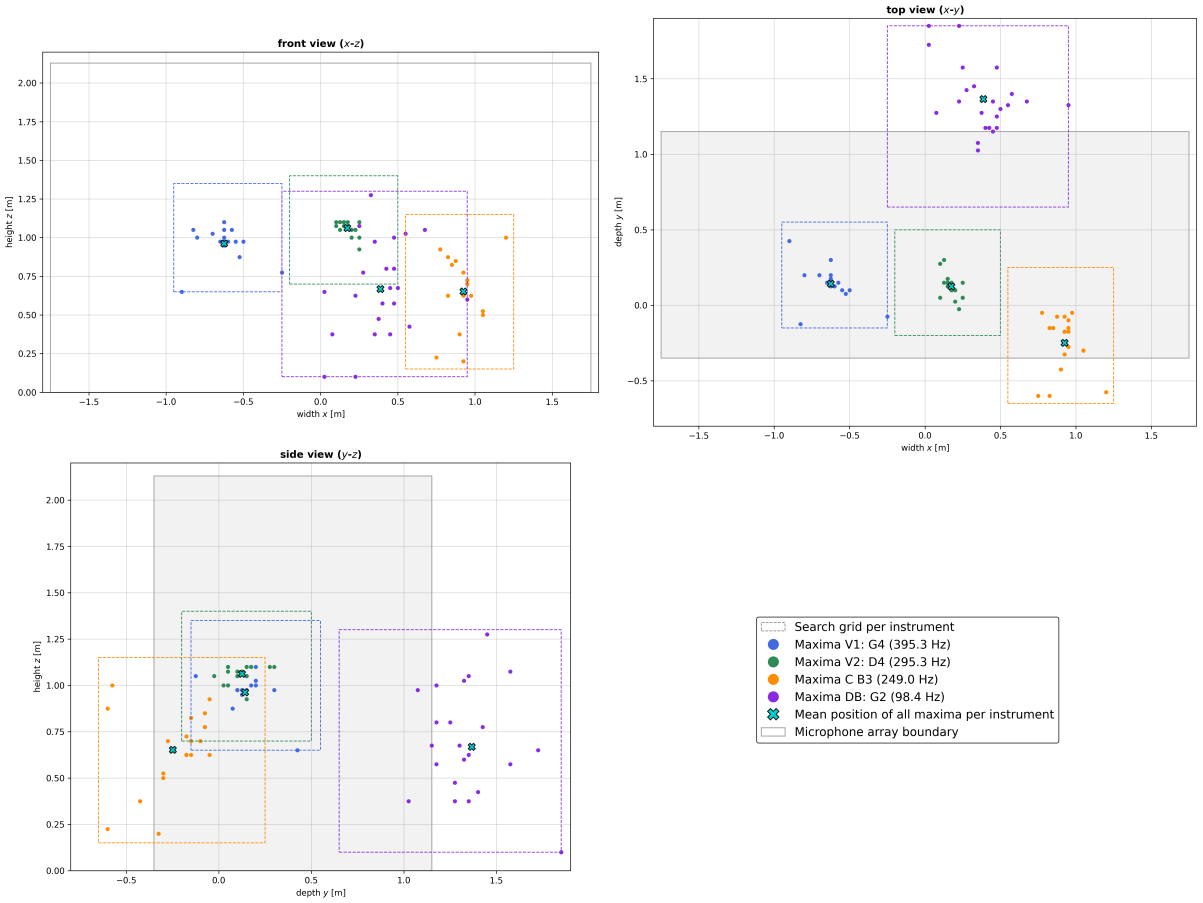

Figure 4: Frequency-dependent maxima found in each instrument-specific search grid with the mean position of all maxima per instrument.

For all investigated peak frequencies, a mean position per instrument was calculated from the found maxima, indicated by an *X* in Figure 4. Although all search grids are larger than the

physical volume of the respective instrument, some maxima for violin V1 and the double bass lie on the boundary of the grid, indicating that the true sound pressure maximum for these frequencies could not be reliably located within the expected region.
The frequency dependence of the maxima positions was investigated by computing the Euclidean distance between each frequency-dependent maximum and the corresponding instrument mean position, as shown in Figure 5. The double bass exhibits the largest deviations, particularly at low frequencies. This can be attributed to the physical size of the instrument and the fact that lower frequencies are radiated by the vibrating body rather than the strings and f-holes alone. An important consideration is also that at higher frequencies, sound radiation becomes increasingly directional, making it more difficult to reliably identify the maximum.
For the two violins, the distances generally remain below 0.2 m, with the exception of the three highest frequencies of violin V1 (5532.8 Hz, 5928.1 Hz, and 6323.4 Hz), which exceed 0.3 m. These peaks have low levels of 22.6 dB, 20.5 dB, and 20.9 dB respectively, compared to a maximum of 60.0 dB at the fundamental frequency of 395.3 Hz, resulting in a low signal-to-noise ratio that negatively affects the localization accuracy.
For the cello, a different problem arises: the f-holes point away from the array and the instrument body is located at the edge of the array coverage, making the cello less well-centered than violin V2, which shows the smallest deviations of all instruments.

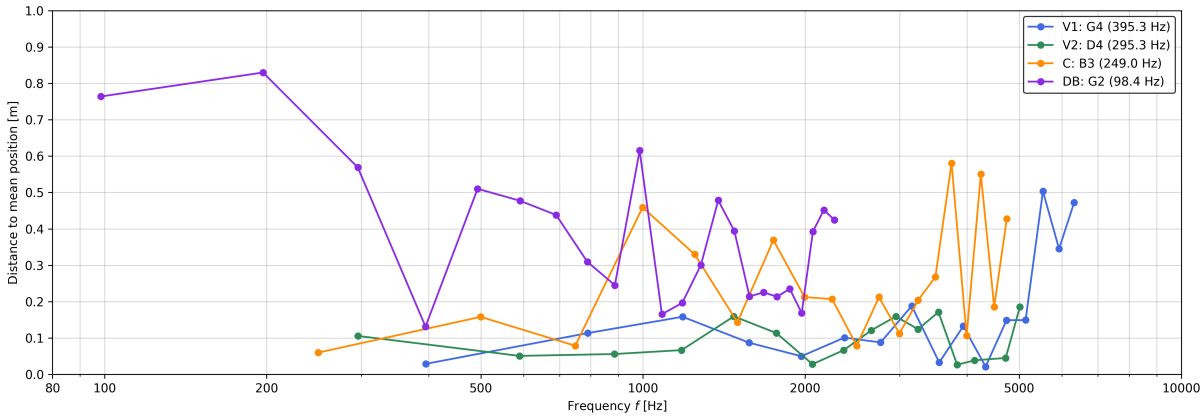

Figure 5: Distance between each frequency-dependent maximum and the mean position of all maxima per instrument.

It should be noted that the localization approach assumes each instrument to behave as a single monopole source. In reality, string instruments radiate sound from multiple regions — including the strings, the f-holes, and the vibrating body — each with frequency-dependent contributions and directional characteristics. This simplification introduces uncertainty in the mean position estimate for all instruments, and is particularly pronounced for the double bass due to its physical size.

4.3 Comparison of the spectra

Figures 6–9 show the investigated frequency spectra for each instrument. All figures share the same structure: the top subplot displays the frequency spectrum of the solo measurement averaged over the instrument-specific time interval (see Figure 3) and all 95 microphones (black),

the synthetic input signal constructed from the sum of all four solo measurements (magenta), and the beamformed signal obtained using CLEAN-T, reconstructed at the instrument-specific

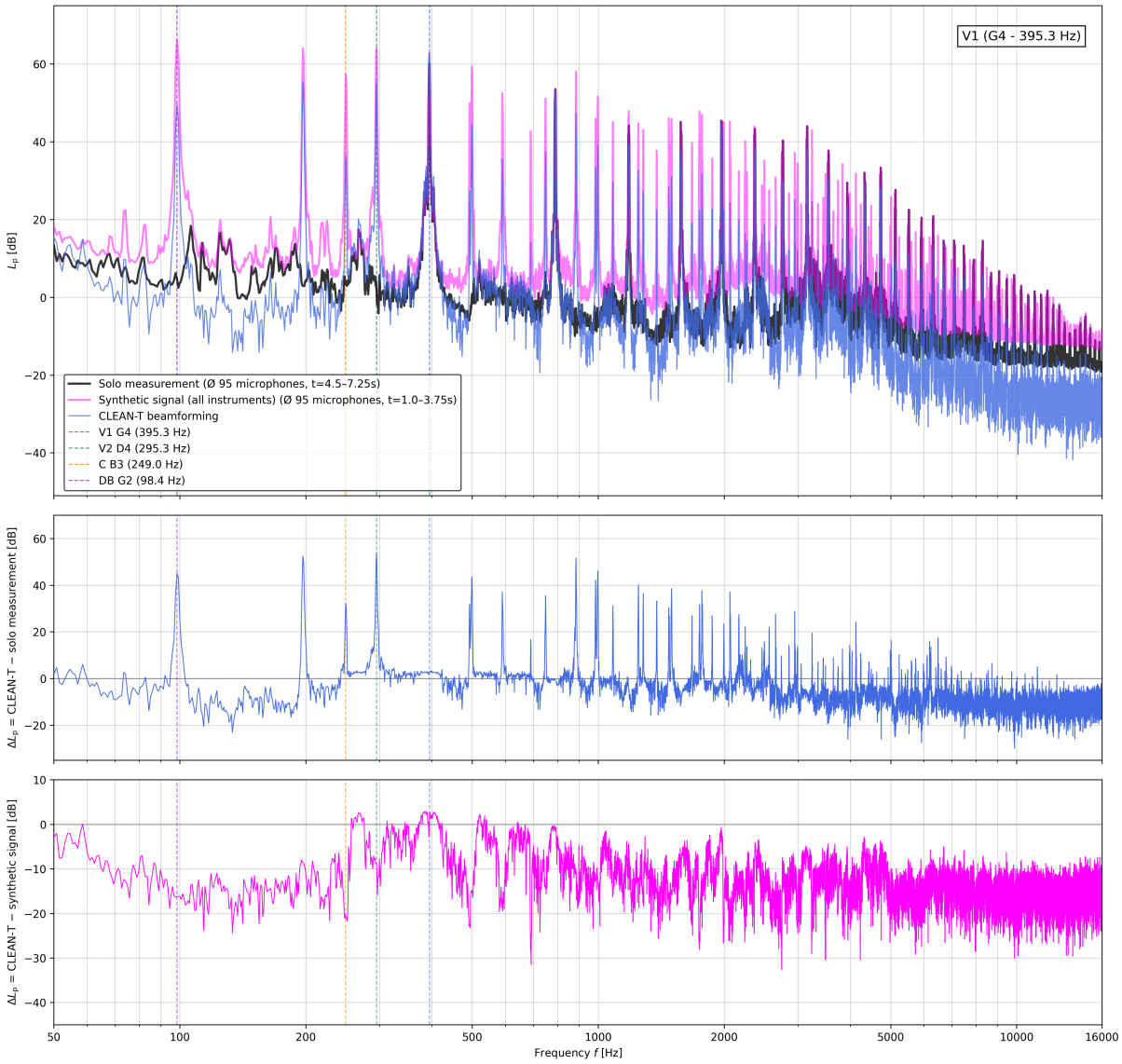

Figure 6: Violin V1: Top — frequency spectra of the solo measurement, the synthetic signal (sum of all instruments), and the CLEAN-T beamformed signal. Middle — difference spectrum ΔL_p between the CLEAN-T beamformed signal and the solo measurement. Bottom — difference spectrum ΔL_p between the CLEAN-T beamformed signal and the synthetic signal. Vertical dashed lines indicate the fundamental frequencies of each instrument.

mean position determined from all frequency-dependent maxima as described in subsection 4.2 (instrument-specific color). The middle subplot shows the difference spectrum ΔL_p of the CLEAN-T beamformed signal minus the solo measurement, serving as an indicator of how

well the target instrument was reconstructed. The bottom subplot shows the difference spectrum ΔL_p between the CLEAN-T beamformed signal and the synthetic signal of all instruments combined, indicating the degree to which contributions from the other three instruments were suppressed. Four vertical dashed lines mark the fundamental frequencies of each instrument in their respective colors.

Violin V1 Figure 6 shows the results for violin V1. In the middle subplot, no prominent peak is visible at the fundamental frequency of V1 (395.3 Hz), indicating that the reconstruction at this frequency was successful. The beamformed signal closely follows the solo measurement at the fundamental and at higher harmonics or overtones of V1, and the difference remains near 0 dB across much of the relevant frequency range. However, large positive peaks are visible in the middle subplot at the fundamental frequencies of the other instruments — most prominently at 98.4 Hz (double bass), 249.2 Hz (cello) and 295.3 Hz (violin V2) — indicating that these contributions were not successfully suppressed and represent leakage in the reconstructed signal. Compared to the synthetic input signal (magenta), the absolute levels of these peaks are reduced, confirming a partial but incomplete separation. The bottom subplot reflects the relative contribution of each instrument to the synthetic signal: values close to or above 0 dB indicate frequency ranges where the target instrument dominates the synthetic signal. Positive values can additionally arise from the fact that the sound pressure reconstructed at a single point near the source may exceed the spatially averaged level of the synthetic signal across all 95 microphones. Strongly negative values indicate frequency ranges dominated by the other instruments, as is particularly evident for the violins below their own fundamental frequencies.

Violin V2 The results for violin V2 in Figure 7 are qualitatively similar to those of V1. The fundamental frequency of V2 (295.3 Hz) is well reconstructed, as confirmed by the near-zero difference at this frequency in the middle subplot. Leakage from the other instruments is again visible as positive peaks in the middle subplot, particularly at the double bass fundamental (98.4 Hz) and the cello fundamental (249.2 Hz), though these are reduced by 5 to 10 dB compared to the synthetic signal. Notably, the leakage at the cello fundamental (249.2 Hz) is approximately 10 dB larger in the V2 reconstruction than in that of V1. This is consistent with the fact that violin V2 was positioned adjacent to the cello, whereas V1 was the most spatially distant from it — suggesting that the spatial proximity between instruments has a measurable effect on the degree of separation achievable as supported by earlier studies [6].
For both violins, the bottom subplot shows strongly negative values below 200 Hz, reflecting that the cello and the double bass dominate the synthetic signal in this frequency range. Near the fundamental frequencies of the violins, the difference slightly exceeds 0 dB, while a small negative dip is visible directly at the fundamental frequency itself. This behavior is not observed for the cello and the double bass, and its cause remains unclear.

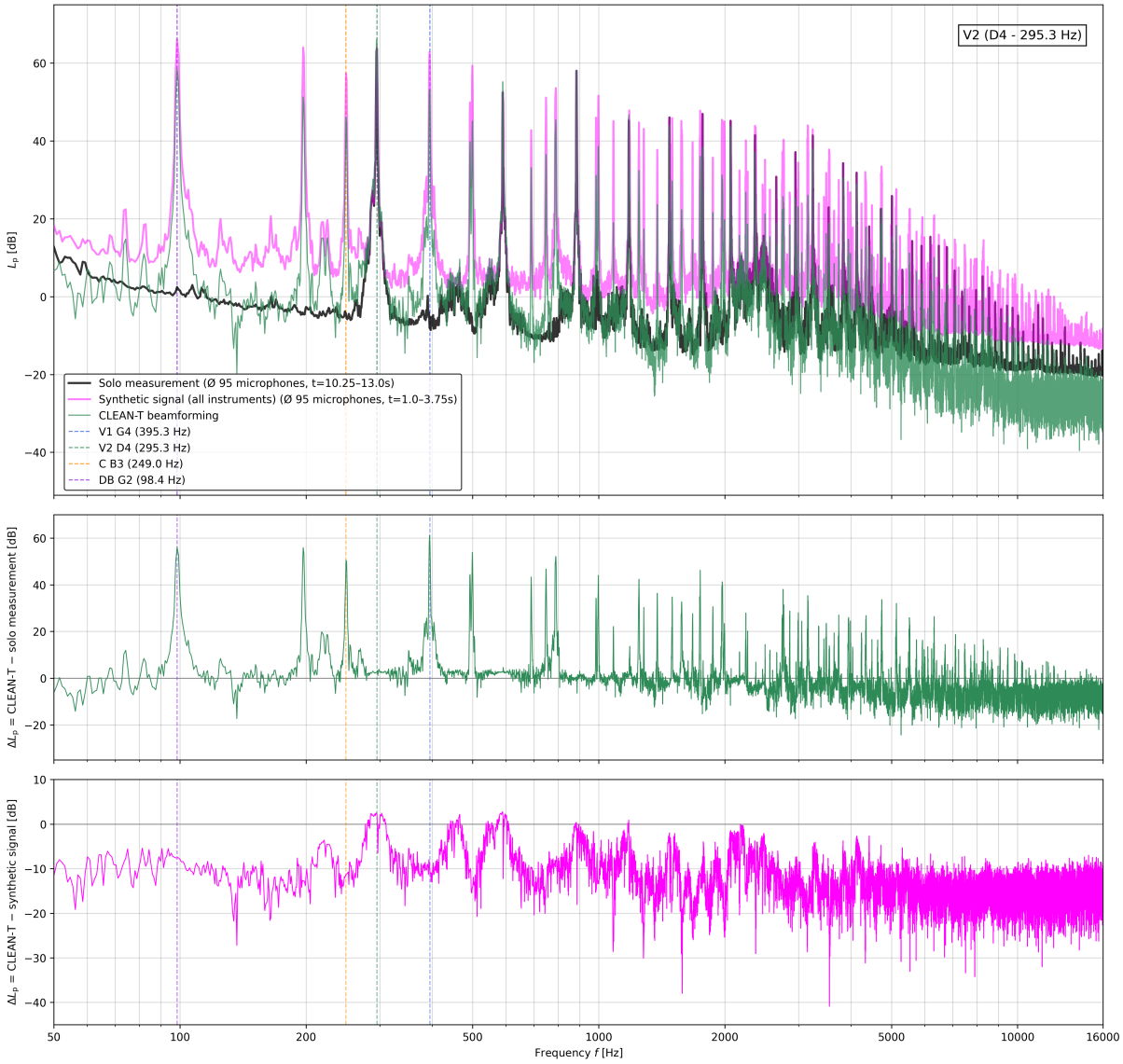

Figure 7: Violin V2: Frequency spectra and difference spectra as in Figure 6.

Cello The results for the cello in Figure 8 are broadly consistent with the violin results. The cello's fundamental frequency (249.2 Hz) is successfully reconstructed, with the difference in the middle subplot remaining close to 0 dB at this frequency. However, a striking feature of the cello result is the very large positive peak at approximately 196.8 Hz in the middle subplot, reaching nearly 60 dB. This corresponds to the second harmonic of the double bass (2×98.4 Hz $\approx$ 196.8 Hz) and indicates that the spatial mean position determined for the cello coincides closely with a strong emission point of this harmonic component of the double bass. The fundamental of the double bass at 98.4 Hz, by contrast, was better suppressed. The bottom subplot shows predominantly negative values across the frequency range, as the cello is the quietest of the four instruments, closely followed by violin V1 (see Table 1), and therefore contributes the least to the synthetic signal. Near the fundamental frequency and its second

harmonic, the difference is positive, indicating successful reconstruction at these frequencies. Notably, positive values are also visible between 100 and 200 Hz, which could indicate either residual contributions from the double bass in the reconstructed cello signal, or radiation from the cello body itself in this frequency range — the latter being supported by the visible peak in the solo measurement (black) at approximately 20 dB in this region.

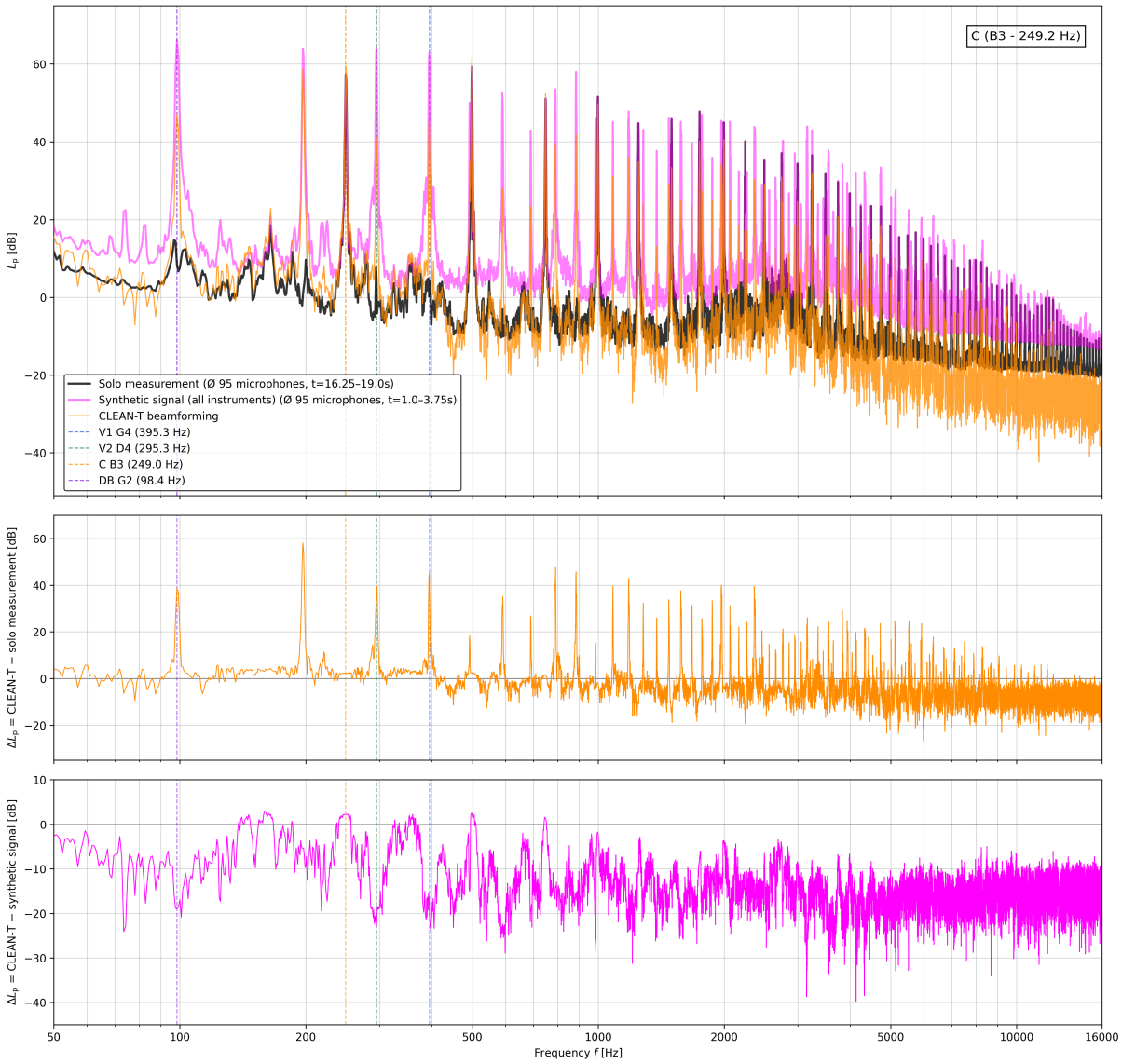

Figure 8: Cello C: Frequency spectra and difference spectra as in Figure 6.

Double bass The double bass results are shown in Figure 9. Among all four instruments, the double bass exhibits the largest spatial spread of frequency-dependent maxima (see Figure 4), a consequence of its physical size and the directional characteristics of its low-frequency radiation. The assumption of a single monopole source position is therefore the least accurate for this instrument, which introduces additional uncertainty in the beamforming reconstruction.

Despite this, the middle subplot shows that the double bass fundamental (98.4 Hz) and much of the higher-frequency content are reconstructed with differences close to 0 dB, representing the best self-reconstruction among all four instruments. The most prominent failure of the isolation is at the cello fundamental frequency (249.2 Hz), where a large positive peak of approximately 45 dB is visible in the middle subplot. This is attributed to the spectral proximity of the cello (249.2 Hz) to the double bass (98.4 Hz), as the cello plays the next lowest fundamental frequency of all instruments. A second notable peak occurs near 498 Hz, corresponding to the second harmonic of the cello (2×249.2 Hz ≈ 498 Hz), which again leaks into the double bass channel. These observations consistently indicate that spectral and spatial proximity between instruments are the primary factors limiting the separation quality.

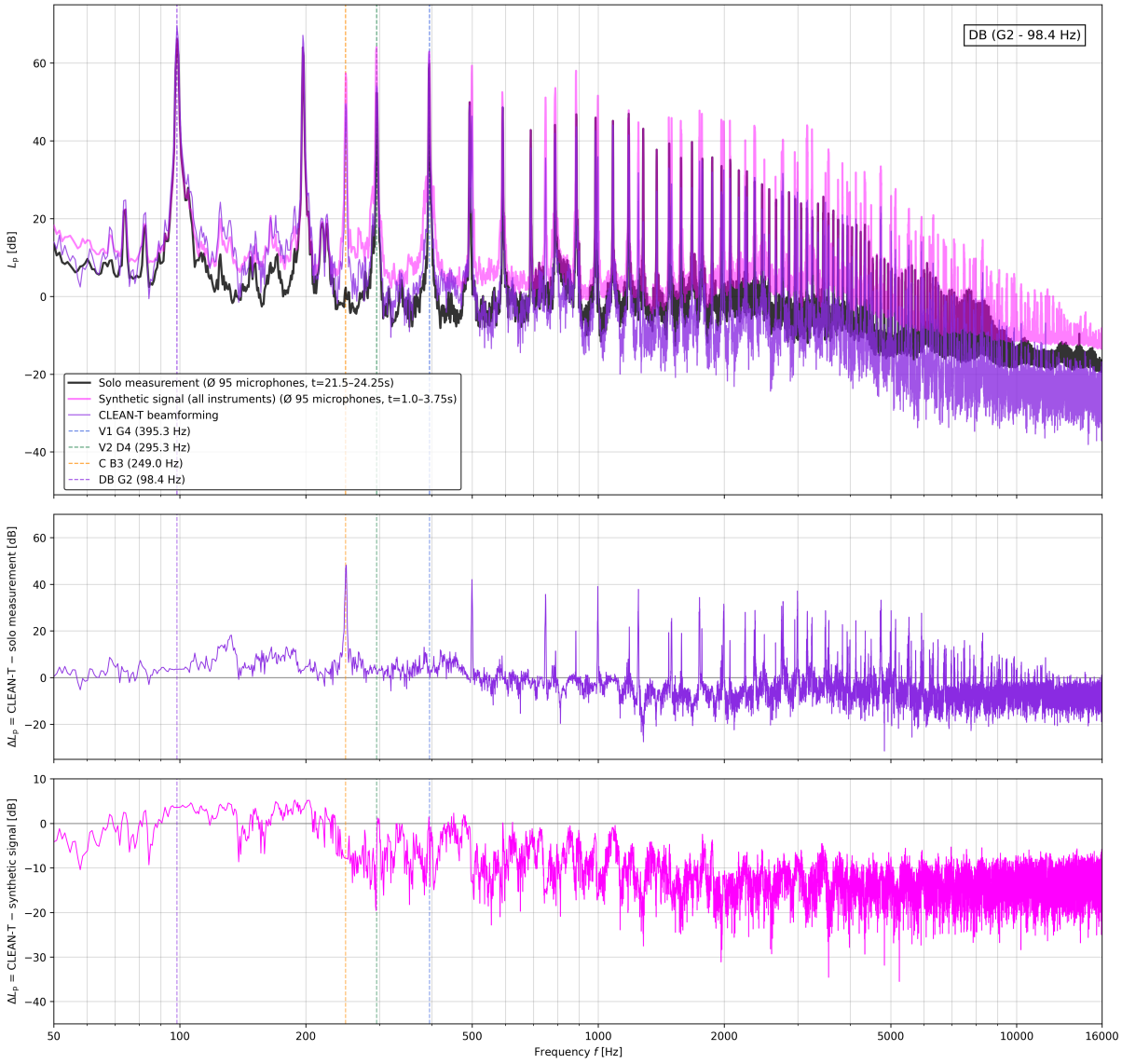

Figure 9: Double bass DB: Frequency spectra and difference spectra as in Figure 6.

The bottom subplot reflects the dominant contribution of the double bass to the synthetic signal at low frequencies, with values close to or above 0 dB below 230 Hz. At higher frequencies, the values become increasingly negative, consistent with the trend observed for all instruments in this frequency range.

General observations Across all four instruments, the most significant sources of incomplete separation are the fundamental frequencies of spatially or spectrally adjacent instruments, with each instrument exhibiting specific limitations as discussed above. In general, the separation quality is primarily limited by the low fundamental frequencies of the cello and the double bass, since the spatial resolution of a microphone array decreases with decreasing frequency, which limits the separability of low-frequency sources. Additionally, the directional radiation characteristics of string instruments contribute to the incomplete separation, as the sound field at the reconstruction point may not accurately represent the full spectral content of the instrument. Finally, despite the use of an anechoic environment, minor reflections caused by the instruments and music stands present during the measurement may have introduced additional disturbances in the recorded signals.

5 CONCLUSIONS

This study investigated the extent to which a microphone array combined with beamforming-based signal separation can be used to isolate individual instruments from a small string ensemble. A three-dimensional array of 95 microphones was designed and its geometry was selected based on a leakage analysis using the point spread function. The measurement was carried out in the anechoic room at TU Berlin with two violins, a cello, and a double bass, each playing a note of the G major chord in succession. Frequency-domain delay-and-sum beamforming was used to localize each instrument, and the CLEAN-T algorithm was subsequently applied at the mean position of all frequency-dependent maxima, determined from the fundamental frequency and the prominent harmonics of each instrument, to extract separate audio tracks from the joint recording.

The results show that the fundamental frequencies of all four instruments were reconstructed with reasonable accuracy. However, complete separation could not be achieved: contributions from adjacent instruments remained visible at their respective fundamental frequencies in all reconstructed signals. The degree of incomplete separation was found to depend primarily on the spatial and spectral proximity between instruments as well as on the directional radiation characteristics of the instruments. In general, the separation quality was most limited for the cello and the double bass, attributed to their low fundamental frequencies and the limited spatial resolution of the array at these frequencies.

Several limitations and directions for future work can be identified. First, the reconstruction was performed at a single mean position per instrument. Since string instruments are not point sources — radiating sound from the strings, the resonance body, and the f-holes simultaneously — the use of multiple reconstruction points per instrument, potentially frequency-dependent, could improve the separation quality. Second, the directivity of the instruments was not accounted for in the array geometry selection or in the beamforming reconstruction. Incorporating instrument-specific directivity models into both the leakage analysis and the signal separation could yield more accurate results. Third, the present study assumed stationary instruments,

whereas in a real performance the position and orientation of each instrument changes continuously. Tracking the instrument motion and adapting the reconstruction points accordingly represents an important step towards a practical application. Finally, all measurements were conducted in an anechoic environment. The performance of the method in acoustically more challenging conditions, such as a concert hall with significant reverberation, remains to be investigated.
Furthermore, in a real orchestral setting instruments of the same section typically play in groups and are positioned together spatially. When multiple instruments of the same type play in unison, their combined sound pressure level dominates the synthetic signal in the shared frequency range, which could potentially improve the separation quality for that instrument group. Investigating the performance of the method for such grouped instrument sections represents a natural extension of the present work.

References

[1] A. Alexandridis, A. Griffin, and A. Mouchtaris. "Capturing and reproducing spatial audio based on a circular microphone array." *Journal of Electrical and Computer Engineering*, 2013, 718574, 2013. URL https://doi.org/10.1155/2013/718574.

[2] J. Berg, T. Johannesson, M. Löfdahl, and A. Nykänen. "In-ear vs. loudspeaker monitoring for live sound and the effect on audio quality attributes and musical performance." In *Proceedings of the 142nd AES Convention*. Berlin, Germany, 2017. Convention Paper 9798.

[3] P. Coleman, P. J. B. Jackson, J. Francombe, and M. Brookes. "Audio object separation using microphone array beamforming." In *Proc. AES 138th Int. Convention, Warsaw, Poland*. 2015.

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

[5] J. Fischer, V. Valeau, and L.-E. Brizzi. "Beamforming of aeroacoustic sources in the time domain: An investigation of the intermittency of the noise radiated by a forward-facing step." *Journal of Sound and Vibration*, 383, 464–485, 2016.

[6] S. Gareayaghi, I. Hagenmaier, H. Henze, and G. Herold. "Filtering of separate audio tracks from synchronously playing musical instruments using beamforming in the time domain." In *Proceedings of the 9th Berlin Beamforming Conference*. 2022.

[7] E. Sarradj. "A generic approach to synthesize optimal array microphone arrangements." In *Proceedings of the 6th Berlin beamforming conference*, volume 29, page 5. 2016.

[8] E. Sarradj, M. Czuchaj, M. Dittrich, and E. Jansen. "Train pass-by noise source characterization and separation tools for cost-effective vehicle certification - innovative separation techniques: Theoretical description and validation testing campaign proposal, Deliverable D2.2." Technical report, TRANSIT Consortium, 2022.

[9] 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.

[10] B. D. Van Veen and K. M. Buckley. "Beamforming: A versatile approach to spatial filtering." *IEEE ASSP Magazine*, 5(2), 4–24, 1988. URL https://doi.org/10.1109/53.665.

[11] H. Vogel. "A better way to construct the sunflower head." *Mathematical biosciences*, 44(3-4), 179–189, 1979.