BeBeC-2026-S02

CONSTRUCTION OF A 64-CHANNEL MULTICHANNEL SYSTEM FOR ACOUSTIC IMAGING BASED ON AUDIO EQUIPMENT

Lucas Bogaz[1] and William D'Andrea Fonseca[1,2]

[1]Acoustical Engineering (EAC), Federal University of Santa Maria (UFSM),
[2]Graduate Program in Architecture, Urbanism, and Landscaping (PPGAUP-UFSM),
Av. Roraima nº 1000, Cidade Universitária, Bairro Camobi,
Santa Maria, RS, Brazil

Abstract

This paper presents the development, characterization, and experimental validation of a 64-channel acquisition system for acoustic imaging based on commercial audio interfaces, external AD/DA converters, and optical ADAT expansion. The study addresses the feasibility of using professional audio equipment as a cost-effective alternative to dedicated multichannel measurement front-ends, with particular emphasis on inter-channel temporal alignment, spectral consistency, phase stability, and synchronization repeatability. Electrical loopback measurements were performed using impulsive signals and logarithmic sine sweeps, allowing impulse responses, frequency response functions, and sample-level offsets to be estimated under different hardware topologies, clocking strategies, and operating systems. The results show that successful multichannel acquisition alone is insufficient for beamforming applications: the decisive requirement is the temporal invariance of the inter-channel delays. Configurations relying exclusively on ADAT clock recovery may exhibit converter-level offsets and, in some cases, non-repeatable latency variations. Conversely, BNC word-clock synchronization preserved internal alignment within each interface group, leaving only stable interface-level offsets that could be compensated through loopback-based time-alignment correction. The proposed architecture was further validated with a 61-microphone arc array in anechoic conditions, producing acoustic maps consistent with simulated point-spread functions. The results demonstrate that cost-effective audio-based architectures can support reliable acoustic imaging when synchronization, loopback referencing, and systematic signal correction are incorporated into the measurement methodology.

1 Introduction

The construction of high-channel-count acquisition systems constitutes a central requirement in several branches of experimental acoustics, particularly when the spatial structure of the sound field must be measured, interpreted, or reconstructed. Applications such as acoustic imaging, beamforming, source localization, holography, and three-dimensional directionality assessment depend not only on the quality of the individual transducers, but also on the simultaneous and coherent acquisition of a large number of signals distributed over space. In this context, microphone arrays provide a methodological framework through which the acoustic field can be sampled at multiple positions, allowing spatial information that would be inaccessible to a single sensor to be inferred from inter-channel amplitude, phase, and time-delay relationships [9, 14].

Acoustic imaging techniques are especially sensitive to the architecture of the acquisition system. Since the position of a sound source is commonly estimated from phase differences or time-of-arrival differences among array channels, measured signals must maintain a stable temporal relationship throughout the acquisition process. Any uncertainty in the relative timing between channels is therefore translated into an uncertainty in the reconstructed source map. For a given time mismatch Δt, the corresponding phase deviation increases with frequency according to $\Delta\phi = 2\pi f \Delta t$, where f is the frequency of the analysis. Hence, even delays corresponding to only a few samples may become relevant at higher frequencies, producing phase errors capable of degrading localization accuracy, widening the main lobe, increasing sidelobe levels, or introducing spurious acoustic sources in the reconstructed image.

1.1 Motivation, research gap, and objectives

The main research gap addressed in this work lies between two contrasting realities. On the one hand, commercial systems dedicated to acoustic measurement offer high reliability and well-controlled synchronization, but their cost may limit their dissemination. On the other hand, professional audio equipment provides a cost-effective and scalable alternative, but its use in acoustic imaging requires careful validation, since its performance with respect to inter-channel coherence, latency stability, and spectral uniformity is not necessarily specified for this purpose.

Accordingly, this study addresses the construction and experimental characterization of a 64-channel multichannel acquisition system for acoustic imaging, implemented using commercially available audio interfaces and external AD/DA converters. Conceived for microphone-array measurements, the proposed system is evaluated with particular attention to temporal synchronization, inter-channel coherence, and spectral consistency. Instead of regarding audio equipment as a straightforward replacement for dedicated measurement hardware, the study adopts a characterization-driven perspective, through which the practical feasibility, intrinsic limitations, and measurable deviations of the architecture are critically assessed within the requirements of acoustic imaging applications.

The specific objectives of the study are as follows:

- to implement a 64-channel acquisition architecture based on professional audio interfaces and external AD/DA converters expanded through the ADAT protocol;
- to configure and compare different synchronization strategies among the devices composing the acquisition system, including ADAT clock recovery and BNC word-clock distribution;

- to characterize the temporal alignment, repeatability, and inter-channel delay behavior of the acquisition chain through electrical loopback measurements;
- to assess the spectral consistency and the delay-induced phase behavior of the acquisition chain; and
- to evaluate the feasibility of the proposed system as a cost-effective platform for multichannel acoustic measurements, including loopback-based time-alignment correction procedures and its validation in an acoustic-imaging application.

By addressing these objectives, the article contributes to the development of accessible experimental infrastructures for acoustic imaging, providing both a practical implementation strategy and a methodological assessment of the conditions under which audio-based multichannel systems may be employed in scientific acoustic measurements [11]. Furthermore, the article itself may serve as a guideline for the design and construction of similar architectures.

2 Theoretical background

This section outlines the main theoretical concepts underlying the development and assessment of the proposed multichannel acquisition system. The discussion is restricted to the topics most directly related to the present study. Further theoretical and technical details can be found in the referenced literature.

2.1 Digital audio architectures for coherent acquisition

Microphone-array processing relies on the preservation of inter-channel time and phase relationships, since these quantities encode the spatial structure of the acoustic field. In high-channel-count systems, the relevant issue is therefore not merely the availability of multiple inputs, but whether the acquisition architecture preserves the temporal conditions required for coherent spatial analysis.

Commercial audio interfaces and external converters provide an accessible route toward large-scale acquisition systems. In this context, the ADAT optical protocol is particularly attractive because it enables modular channel expansion, allowing multiple conversion units to be integrated into dense architectures such as 64-channel configurations. From an implementation standpoint, this makes audio-oriented hardware a practical candidate for experimental infrastructures that demand scalability without the complexity or cost typically associated with dedicated instrumentation.

Scalability alone, however, is insufficient to establish scientific suitability. In multidevice arrangements, all converters must operate from a common temporal reference so that the recorded signals correspond to the same sampling instants. Otherwise, fixed offsets, drift, jitter, or non-repeatable latency may alter the relative timing among channels and compromise the coherence of the acquired data.

2.1.1 Implications for acoustic imaging

Any distortion introduced at the acquisition stage may propagate directly into localization, beamforming, and field reconstruction. Array-based methods assume that inter-channel differences

arise from sound propagation; when the hardware introduces additional timing or phase deviations, these artifacts may be interpreted as physical information rather than as acquisition-induced errors.

For this reason, the assessment of multichannel audio hardware must extend beyond nominal recording performance. What must be demonstrated is the preservation of temporal alignment, repeatability, and spectral consistency under representative measurement conditions. Within this framework, the present study addresses both the feasibility of a cost-effective acquisition strategy and the methodological principles required for the conception of comparable architectures.

2.2 Digital clock synchronization

In digital audio systems, the sampling clock defines the temporal reference for analog-to-digital and digital-to-analog conversion. Whereas a single device typically operates from its internal clock, multidevice architectures require a shared reference to ensure coherent sampling across all channels. Without such synchronization, relative timing errors may arise, leading to phase instability, clicks, sample slips, or non-repeatable latency.

In professional systems, this common reference may be distributed either through a dedicated word-clock connection, typically implemented via BNC cables, or through digital audio links that embed timing information, such as ADAT, AES/EBU, or S/PDIF, see Figure 4. In both cases, coherent operation depends on all devices being locked to the same temporal basis, so that the acquired signals correspond to the same sampling instants.

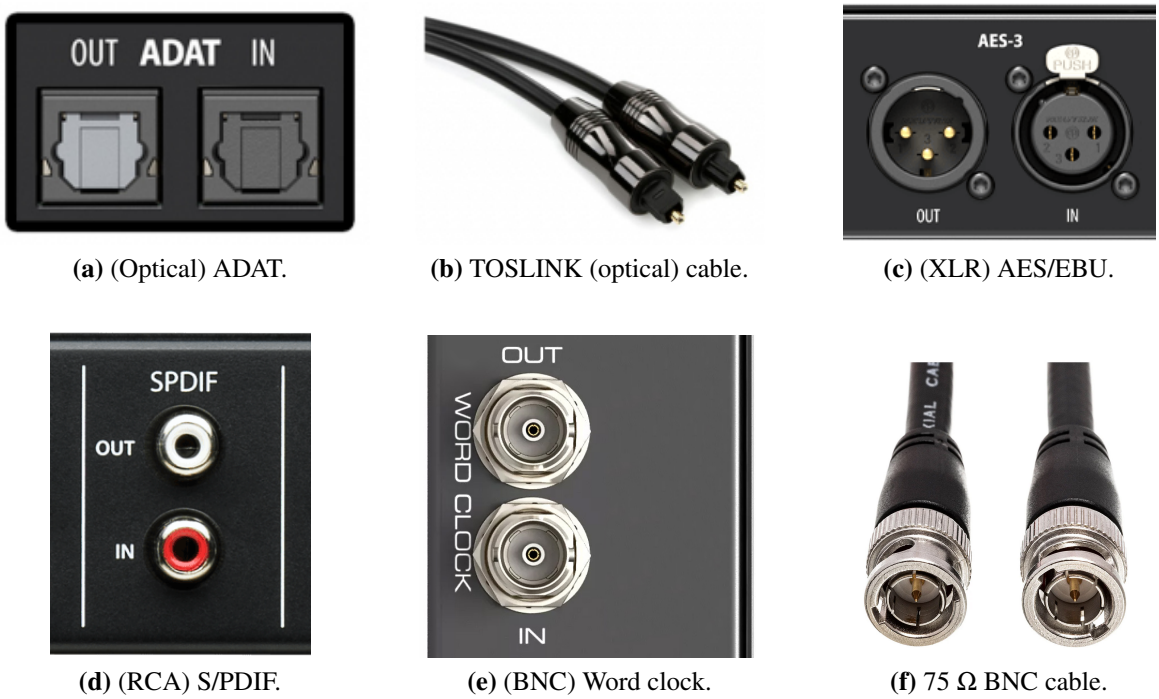
In this context, it is useful to distinguish between jitter and drift. Jitter denotes short-term deviations of clock edges from their ideal temporal positions, whereas drift refers to slower variations in clock frequency over time. Although both phenomena degrade synchronization quality, they do so differently: jitter primarily affects short-term sampling accuracy, whereas drift progressively alters the relative timing between devices during extended or repeated acquisitions [7, 12, 15].

2.3 ADAT optical interface

The ADAT optical interface, originally derived from the Alesis Digital Audio Tape system, became a widely adopted protocol for expanding digital audio systems through optical transmission via TOSLINK[1] connections [1]. In its conventional single-speed configuration, ADAT transmits eight channels of uncompressed digital audio at 24-bit resolution and sampling rates up to 48 kHz through a single optical TOSLINK cable. In practical implementations, optical TOSLINK cables are commonly kept within short lengths, often around a few meters, in order to preserve reliable optical transmission. At higher sampling rates, the number of available channels is reduced through sample multiplexing schemes, commonly referred to as S/MUX. Thus, a single ADAT optical connection typically carries eight channels at 44.1 kHz or 48 kHz, four channels at 88.2 kHz or 96 kHz, and two channels at 176.4 kHz or 192 kHz.

In the present study, this limitation is particularly relevant because the number of simultaneous channels constitutes a central requirement for acoustic imaging. Since reducing the channel count would compromise the intended array configuration, the analysis is restricted to sampling rates compatible with eight-channel ADAT transmission, namely $f_s = 44.1$ kHz and $f_s = 48$ kHz.

[1] TOSLINK, originally developed by Toshiba, is a standardized optical fiber connection used for the transmission of digital audio signals. In this interface, the signal is transmitted by light rather than by an electrical conductor, which provides galvanic isolation between devices and reduces susceptibility to electromagnetic interference.



(a) (Optical) ADAT. **(b)** TOSLINK (optical) cable. **(c)** (XLR) AES/EBU.

(d) (RCA) S/PDIF. **(e)** (BNC) Word clock. **(f)** 75 Ω BNC cable.

Figure 1: Digital audio interconnection formats, connectors, and clock: ADAT (optical), TOSLINK (optical) cable, AES/EBU (XLR), S/PDIF (RCA), word-clock (BNC), and 75 Ω BNC cable. [4, 19, 24].

The ADAT protocol organizes the transmitted data in frames associated with successive sampling instants. Each frame contains audio samples from up to eight channels, each represented with 24-bit resolution, in addition to synchronization and auxiliary information. A simplified representation of this structure is shown in Figure 2. Since the optical stream carries both audio data and embedded timing information, no separate clock line is required when synchronization is performed directly through ADAT. The receiver recovers the clock from the incoming optical stream, typically through a Phase-Locked Loop (PLL) [6, 13]. Consequently, the quality and stability of this clock recovery process may influence the temporal alignment among channels and devices.

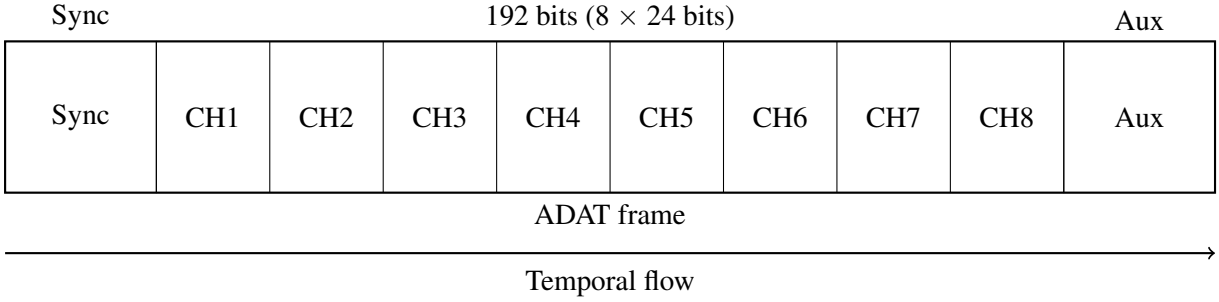


Figure 2: Simplified representation of an ADAT frame containing synchronization information, eight 24-bit audio channels, and auxiliary data.

2.4 Latency and temporal alignment

Latency is the time interval between the occurrence of an analog event and its representation in the digital domain, or conversely between a digital signal and its analog reconstruction. In multichannel acquisition systems, the absolute latency of the system is usually less critical than the relative latency among channels. If all channels exhibit the same delay, the temporal structure of the measured sound field is preserved. However, if different channels exhibit different delays, the acquired signals become artificially misaligned. For example, at a sampling rate of 48 kHz, one sample corresponds to approximately 20.8 µs; thus, a relative delay of 3 samples between two channels corresponds to about 62.5 µs. Assuming a sound speed of 343 m/s, this delay is equivalent to a propagation-path difference of approximately 21.4 mm, which is sufficient to affect phase-based localization and beamforming results.

These offsets may be fixed or variable. Fixed offsets can, in principle, be measured and compensated during post-processing, provided that they remain invariant over time and across measurement sessions. Variable offsets, on the other hand, are considerably more problematic, since they undermine the repeatability of the measurement system and may require a loopback reference for each relevant acquisition group or even for each converter.

The phase deviation introduced by a temporal delay Δt at a frequency f is given by [17]

$$\Delta\phi = 2\pi f \Delta t. \tag{1}$$

If the delay corresponds to N samples at a sampling rate f_s, then

$$\Delta t = \frac{N}{f_s}, \tag{2}$$

and therefore

$$\Delta\phi = 2\pi f \frac{N}{f_s}. \tag{3}$$

As an illustrative example, a delay of one sample at $f_s = 44.1$ kHz corresponds to

$$\Delta t = \frac{1}{44100} \approx 2.2676 \times 10^{-5}\ \text{s} \approx 22.68\ \mu\text{s}.$$

At $f = 10$ kHz, this delay produces a phase deviation of

$$\Delta\phi = 2\pi \cdot 10000 \cdot 2.2676 \times 10^{-5} \approx 1.424\ \text{rad} \approx 81.6^\circ.$$

This result shows that even a one-sample delay may become highly significant at high frequencies. For this reason, sample-level misalignments cannot be neglected in experimental multichannel acoustic measurements, especially when phase information is used for spatial processing.

2.5 Acoustic imaging and beamforming

Acoustic imaging comprises a set of experimental and computational techniques intended to reconstruct the spatial distribution of acoustic sources from measurements performed over a finite number of receiving positions. In this context, the measured sound field is interpreted not only through its spectral content, but also through the spatial information contained in the relative

amplitudes and phases observed among the sensors. Among the most established techniques for acoustic imaging, three approaches are commonly emphasized: acoustic holography, sound intensity methods and beamforming. Acoustic holography is generally associated with near-field measurements and allows the reconstruction of the sound field over a surface close to the source region. Sound intensity methods, in turn, are based on the estimation of the active acoustic energy flow. Beamforming differs from these approaches by enabling the localization and mapping of sources from array measurements, including applications in the far-field, for which the incident wavefronts can often be approximated as planar, or near-field.

The fundamental principle of beamforming is the coherent combination of signals acquired by a microphone array in order to emphasize sound arriving from a prescribed position or direction while attenuating contributions from other regions of space. Thus, beamforming can be understood as a spatial filter, whose selectivity is determined by the array geometry, the number of sensors, the frequency range, the assumed propagation model, and the distance between the source and the array. When the source is sufficiently far from the array, the impinging wavefront may be represented by a plane wave. Conversely, for near-field problems, the curvature of the wavefront becomes relevant and the source is more appropriately represented by a monopole model. Therefore, the choice between a point-source or plane-wave formulation depends on the relation between the source distance, the array aperture, and the wavelength of interest.

The simplest interpretation of beamforming is given by the delay-and-sum formulation [16]. Consider an array composed of M omnidirectional microphones located at positions $\mathbf{x}_m = (x_m, y_m, z_m)$, with $m = \{1, 2, \ldots, M\}$. For a candidate source position $\mathbf{x}_p$ in a predefined scanning region, the distance between the focus point and the mth microphone is

$$r'_m = \left|\mathbf{x}_m - \mathbf{x}_p\right|, \tag{4}$$

in which the distance between the focus point and the array reference position is denoted by $r_p = \left|\mathbf{x}_p\right|$. The relative delay required to focus the array at $\mathbf{x}_p$ is then given by

$$\Delta_m = \frac{r_p - r'_m}{c}, \tag{5}$$

in which c represents the speed of sound. In the time domain, the delay-and-sum beamformer output may be written as

$$b(t) = \frac{1}{M} \sum_{m=1}^{M} w_m \, p_m(t - \Delta_m), \tag{6}$$

in which $p_m(t)$ is the pressure signal measured by the mth microphone and w_m is a spatial weighting factor. For near-field focusing, this weighting may also include a correction for spherical spreading, so that microphone signals are both phase-aligned and amplitude-compensated before summation. If the assumed focus point coincides with the actual source position, the microphone signals are coherently added and the beamformer output is maximized. If the focus point does not correspond to a true source location, the phase alignment is imperfect and destructive interference reduces the beamformer response.

For acoustic imaging, the scanning region is discretized into a grid of candidate source positions. For each grid point, a propagation model is used to compute the expected phase and amplitude relationship between a hypothetical source at that point and each microphone in the

array. This model is commonly expressed through a steering vector $\mathbf{g}(\mathbf{x}_p, \omega)$, defined as

$$\mathbf{g} = \begin{bmatrix} g_1 & g_2 & \cdots & g_M \end{bmatrix}^{\mathrm{T}}, \tag{7}$$

whose entries contain the transfer functions between the candidate source position and the microphone positions. In free-field, a monopole-based steering function may be written as

$$g_m(\mathbf{x}_p, \omega) = \frac{\mathrm{e}^{-\mathrm{j}k\left|\mathbf{x}_m - \mathbf{x}_p\right|}}{4\pi\left|\mathbf{x}_m - \mathbf{x}_p\right|}, \tag{8}$$

in which $k = 2\pi f/c$ is the acoustic wavenumber. This expression accounts for both propagation delay and spherical spreading from the candidate source position to the sensor. In practice, the steering vector represents the acoustic signature that would be observed by the array if a source were located at a given scanning point. Beamforming therefore compares the measured data with this modeled signature and assigns higher values to positions whose predicted propagation behavior is more consistent with the measured sound field.

Conventional Beamforming (CB) is a frequency-domain method usually formulated from the cross-spectral matrix of the microphone signals [5, 22]. In practice, the time signals acquired by the M microphones are first divided into finite data blocks, optionally windowed, and then transformed into the frequency domain by means of the Fast Fourier Transform (FFT). Thus, for each frequency bin ω_ℓ, the complex pressure of the mth microphone may be written as

$$P_m(\omega_\ell) = \mathrm{FFT}\left\{p_m[n]\right\}_\ell, \tag{9}$$

in which $p_m[n]$ is the discrete-time pressure signal measured by the mth microphone and ℓ denotes the frequency-bin index. The beamforming procedure is then carried out independently for each frequency, or over a selected frequency band after an appropriate spectral averaging or band integration. Let the vector of complex pressures at a given frequency be

$$\mathbf{p}(\omega_\ell) = \begin{bmatrix} P_1(\omega_\ell) & P_2(\omega_\ell) & \cdots & P_M(\omega_\ell) \end{bmatrix}^{\mathrm{T}}. \tag{10}$$

The cross-spectral matrix $\mathbf{C}(\omega_\ell)$ is then defined as

$$\mathbf{C}(\omega_\ell) = \frac{1}{2}\left\langle \mathbf{p}(\omega_\ell)\,\mathbf{p}^{\mathrm{H}}(\omega_\ell) \right\rangle, \tag{11}$$

in which $(\cdot)^{\mathrm{H}}$ denotes the Hermitian transpose and $\langle\cdot\rangle$ indicates spectral averaging over data blocks. This matrix contains the auto-spectral terms along its main diagonal and the cross-spectral terms outside the diagonal, thereby preserving the relative magnitude and phase relationships between all microphone pairs. Consequently, the cross-spectral matrix summarizes the spatial coherence information required to estimate the direction or position from which sound energy is arriving.

For a given scanning position, CB estimates the source power by projecting the measured cross-spectral information onto the steering vector associated with that position. The corresponding

array power response is

$$A(\mathbf{x}_p, \omega_\ell) = \frac{\mathbf{g}^{\mathrm{H}}(\mathbf{x}_p, \omega_\ell)\, \mathbf{C}(\omega_\ell)\, \mathbf{g}(\mathbf{x}_p, \omega_\ell)}{\left|\mathbf{g}(\mathbf{x}_p, \omega_\ell)\right|^4}, \tag{12}$$

in which the denominator provides the appropriate normalization with respect to the steering vector norm. By evaluating Equation (12) over all points of the scanning grid, an acoustic map is obtained for each frequency bin or frequency band of interest. Peaks in this map indicate regions whose modeled propagation characteristics are most consistent with the measured array data. The resulting image may then be represented as a color map and, when appropriate, superimposed on an optical image of the measurement scene, allowing the acoustic contribution of different regions to be visually interpreted.

The spatial resolution and dynamic range of a conventional beamforming map are strongly governed by the array aperture, microphone distribution, frequency, and source distance. Larger apertures generally improve angular resolution, whereas insufficient spatial sampling may lead to spatial aliasing at high frequencies. In addition, the sidelobe structure of the array point-spread function (PSF) may introduce artifacts in the acoustic image, especially when strong sources coexist with weaker ones. Despite these limitations, CB remains one of the most widely used methods in acoustic imaging due to its conceptual clarity, computational efficiency, and robustness in practical measurements.

2.5.1 Point-spread function of the array

Following the conventional beamforming formulation introduced above, the spatial performance of an array is evaluated through its point-spread function (PSF) [23]. In acoustic imaging, the PSF represents the response produced by the array and by the imaging algorithm when the acoustic field is generated by an ideal point source located at a known position. Therefore, it provides a direct description of the spatial selectivity of the measurement system, including the width of the main lobe, the distribution of sidelobes, the effective resolution, and the possible occurrence of spatial artifacts.

In the present analysis, the PSF was obtained by simulating the acoustic field generated by a monopole source and processing the resulting microphone signals with CB. Let $\mathbf{x}_\mathrm{s}$ denote the position of the simulated point source. Using the free-field monopole model introduced in Equation (8), the synthetic pressure vector measured by the array may be written as

$$\mathbf{p}_\mathrm{s}(\omega) = q(\omega)\, \mathbf{g}(\mathbf{x}_\mathrm{s}, \omega), \tag{13}$$

in which $q(\omega)$ is the complex source strength and $\mathbf{g}(\mathbf{x}_\mathrm{s}, \omega)$ is the steering vector associated with the true source position. For a deterministic point-source field, the corresponding cross-spectral matrix is given by

$$\mathbf{C}_\mathrm{s}(\omega) = \mathbf{p}_\mathrm{s}(\omega)\, \mathbf{p}_\mathrm{s}^{\mathrm{H}}(\omega). \tag{14}$$

Substitution of Equation (14) into the CB formulation of Equation (12) yields the frequency-

dependent array response to the simulated point source:

$$A_{\mathrm{PSF}}(\mathbf{x}_p,\omega) = \frac{\mathbf{g}^{\mathrm{H}}(\mathbf{x}_p,\omega)\,\mathbf{C}_{\mathrm{s}}(\omega)\,\mathbf{g}(\mathbf{x}_p,\omega)}{\left|\mathbf{g}(\mathbf{x}_p,\omega)\right|^4}. \tag{15}$$

The normalized point-spread function is then defined as

$$\mathrm{PSF}(\mathbf{x}_p,\omega) = \frac{A_{\mathrm{PSF}}(\mathbf{x}_p,\omega)}{\max\limits_{\mathbf{x}_p\in\Omega} A_{\mathrm{PSF}}(\mathbf{x}_p,\omega)}, \tag{16}$$

in which Ω denotes the scanning domain. This normalization makes the maximum response equal to unity and allows the spatial pattern of the array to be compared across different frequencies independently of the absolute source amplitude. For graphical representation, the normalized response is usually expressed in decibels (dB) according to

$$\mathrm{PSF}_{\mathrm{dB}}(\mathbf{x}_p,\omega) = 10\log_{10}\left[\mathrm{PSF}(\mathbf{x}_p,\omega)\right]. \tag{17}$$

3 Methodological development

This section presents the methodological development adopted for the construction and evaluation of the proposed 64-channel acquisition system. The procedures were designed to compare the investigated configurations both temporally and spectrally, with emphasis on inter-channel alignment, repeatability, and synchronization stability.

As discussed in the theoretical background, the ADAT protocol imposes a direct relation between the sampling rate and the number of channels transmitted simultaneously. Since a high channel count is essential for beamforming and microphone-array measurements, configurations that reduce the number of available channels were not considered. Therefore, the system was evaluated primarily at $f_{\mathrm{s}} = 44.1$ kHz and $f_{\mathrm{s}} = 48$ kHz[2]. All measurements were conducted with Matlab [25] combined with ITA-Toolbox [3].

3.1 Acquisition systems

The experimental investigation was conducted on both MacOS and Windows platforms, using audio-interface configurations[3] defined according to the compatibility constraints of each operating system. Under MacOS, two configurations were evaluated. The first consisted of a hybrid arrangement comprising one PreSonus FireStudio Lightpipe [18] and one RME Digiface USB [21], whereas the second employed two RME Digiface USB interfaces, thereby allowing a fully RME-based configuration to be assessed under the same operating system,

[2]This variation was necessary because, when the PreSonus FireStudio Lightpipe was used only as the clock master without a FireWire connection, its clock operated at 48 kHz. In more elaborate setups, dedicated clock-distribution hardware connected via BNC word-clock lines may also be employed to provide a centralized timing reference for multiple devices.

[3]Although dedicated 64-channel ADAT-to-Dante (audio-interface) solutions exist, they were not accessible in the laboratory, which motivated the construction and assessment of a multi-interface architecture.

observe Figure 3. In both configurations, each audio interface was connected to four Behringer ADA8200 converters [2], resulting in a 64-channel acquisition architecture.

Under Windows, the measurements were restricted to the dual-RME configuration, since the FireWire-based PreSonus interface could not be used in that environment. In selected tests, the PreSonus interface was employed only as a clock-generation device via BNC, allowing its role in the synchronization strategy to be evaluated independently of its use as an acquisition interface. Table 1 summarizes the main equipment used in the experimental configurations of the proposed architecture.

(a) PreSonus FireStudio Lightpipe (audio-interface).

(b) Behringer ADA8200 (AD/DA).

(c) RME Digiface USB (audio-interface).

Figure 3: Acquisition systems: PreSonus FireStudio Lightpipe [18], Behringer ADA8200 [2], and RME Digiface USB [21].

Additional configurations were also evaluated, including the use of a (previous) Behringer ADA8000 converter. This configuration was relevant in identifying that the combined use of ADA8000 and ADA8200 units may require additional loopback references, since different converter models may exhibit different latencies and slightly different frequency responses.

Table 1: Main equipment used in the experimental multichannel acquisition configurations.

Equipment	General description	Relevant features	Role in the system
PreSonus FireStudio Lightpipe [18]	FireWire digital audio interface for multichannel optical audio transmission.	Up to 32 channels of ADAT or S/-MUX optical input/output; 24-bit operation; word-clock input/output via BNC.	Audio interface in the hybrid MacOS configuration and clock-generation device in selected tests.
RME Digiface USB [21]	USB digital audio interface based on optical ADAT connectivity.	Four optical ADAT input and output ports; up to 32 input and 32 output channels in ADAT mode; operation under MacOS and Windows.	Audio interface in both hybrid and dual-RME configurations.
Behringer ADA8200 [2]	Eight-channel AD/DA converter and microphone preamplifier unit with optical ADAT connectivity.	Eight analog inputs and outputs; 24-bit conversion at 44.1 kHz or 48 kHz; ADAT optical input/output; external synchronization via ADAT or BNC word-clock input.	External converter used to expand each audio interface by eight analog channels per ADAT connection.

Table 2: List of components employed in the system assembly.

Item	Quantity	Application
PreSonus FireStudio Lightpipe	01	Audio interface
RME Digiface USB	02	Audio interface
Behringer ADA8200	08	AD/DA converters
8U Rack Case 8U x 40 cm or deeper	01	System assembly
Standard rack power strip	01	Electrical connections
IEC C13 power cord	09	Electrical connections
TOSLINK cables	16	ADAT IN/OUT connections
75 Ω BNC cables, 30 cm each	08	Clock connections
BNC T-shape adapter MX 75 Ω	08	Clock connections
75 Ω BNC termination	01	Clock connections
USB 2.0 type A to B male cable	02	Interface connections
Fireware 800/400 cable	01	Interface connections
Standard rack screews (usually M5 or M6 x 15 mm)	40−64	Equipment mounting
Notebook (Windows and Mac OS)	02	System control
Velcro and nylon cable ties	–	Rack organization
Identification and fastening tapes	–	Rack organization
Organization rack panels (blanking, venting etc)	00−06	Rack organization

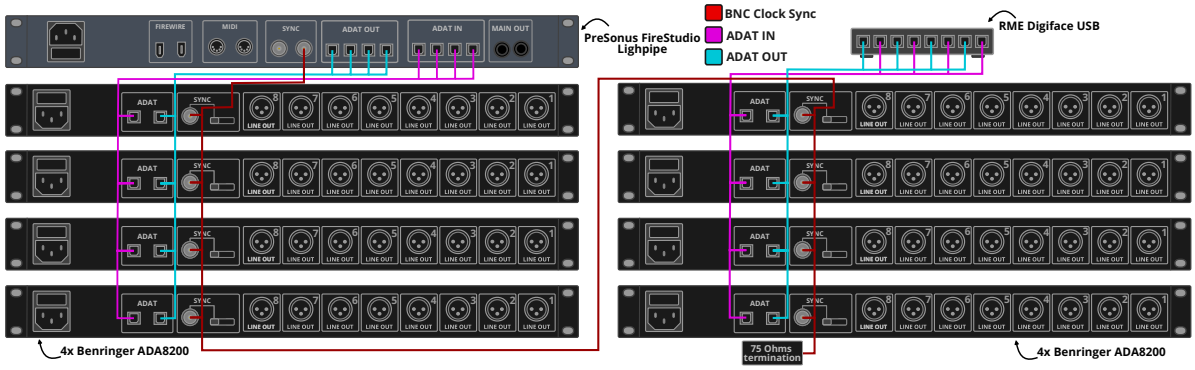
Consequently, when heterogeneous converter units are included in the same acquisition chain, the adjustment procedure must account not only for temporal offsets, but also for possible spectral differences among devices.

The assembled acquisition system is shown in Figure 4, where different combinations of audio interfaces and ADAT converters can be observed. Due to the large number of channels, a hierarchical channel-numbering scheme was adopted. Each ADA8200 unit was treated as a

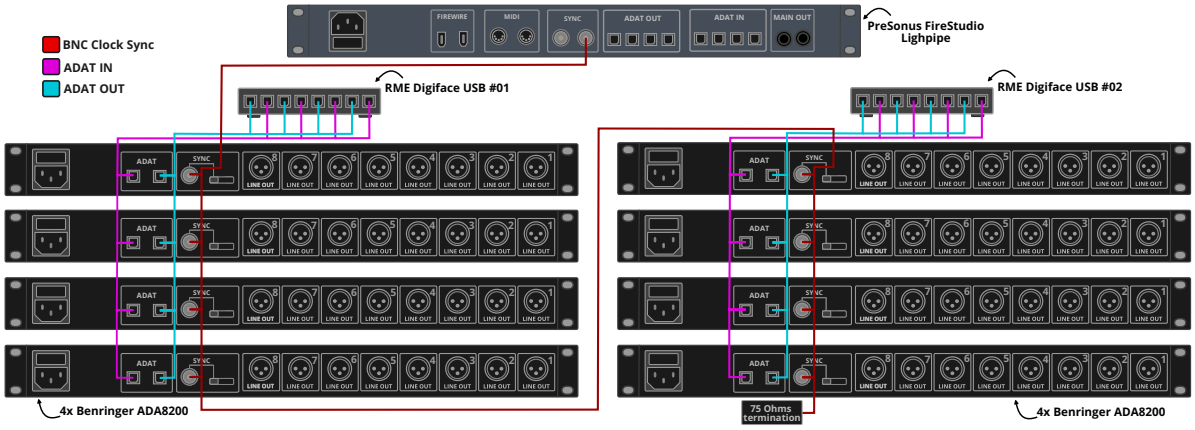
(a) Front view and frontal connections.

(b) Rear view and rear connections.

Figure 4: Acquisition system assembled in an 8U rack case on each side (40 cm depth).



(a) MacOS configuration.



(b) Windows and MacOS configurations.

Figure 5: Connection diagrams of the evaluated acquisition systems.

channel group[4]. The complete system was assembled in an 8U rack case[5]. This arrangement facilitated transport, cable organization, physical protection, and repeatability of the connection layout. Table 2 presents a detailed list of the components used in the system assembly, in order to enable its replication by interested researchers.

The complete connection diagrams for the MacOS and Windows configurations are shown in Figure 5. These diagrams summarize the interface topology, the ADAT optical interconnections, and the clock-distribution scheme adopted in each evaluated setup.

3.2 Clock configuration

Two synchronization strategies were evaluated experimentally: recovery of the clock from the ADAT optical stream and distribution through dedicated BNC word-clock connections. In all cases, a single device was configured as the master clock, while the remaining units operated as slaves. The comparison therefore focused on the practical performance of the two clock-distribution schemes, rather than on different synchronization topologies.

Figure 6 presents the Behringer ADA8200 sync configuration switch and the physical components employed in the word-clock configuration (BNC T-connectors and 75 Ω termination). In this BNC configuration, proper termination was required to reduce reflections along the clock-distribution line and to ensure stable signal transmission.

(a) Sync section of the Behringer ADA8200.

(b) 75 Ω BNC T-connector with two female ports and one male port.

(c) 75 Ω BNC termination.

Figure 6: Physical components used for clock synchronization for the Behringer ADA8200 converters and in the cascaded BNC word-clock connections.

3.3 Experimental procedure

The experimental procedure was designed to evaluate the temporal behavior of the acquisition system under different operating systems, interface combinations, and clocking strategies. Nearly all measurements were carried out by means of electrical loopback tests, in which an analog output channel is physically connected to the input channels through XLR cabling. In this configuration, the same excitation signal is fed back into the system, allowing the relative timing

[4]Instead of numbering the channels sequentially from 1 to 64, they were renumbered from 1.1 to 8.8. The first digit refers to the ADA8200 unit, whereas the second digit refers to the channel within that unit. For example, channel 50 corresponds to channel 7.2.

[5]A rack unit is a standardized measure used to define the height of rack-mounted equipment. Each unit corresponds to 1.75”, approximately 44.45 mm, in height. Standard rack width is 19.00”, approximately 482.60 mm, whereas depth depends on the specific system configuration [20].

between input channels to be assessed directly from the recorded waveforms. Figure 7 illustrates the adopted loopback connection scheme.

The first channel of the first ADA8200 unit, denoted as Channel 1.1, was adopted as the temporal reference in all tests. Using passive splitter cables, the output signal was simultaneously routed to Channel 1.1 and to each of the remaining inputs, so that the reference channel was always acquired jointly with the other channels under identical loopback conditions. In this analysis, only the acquisition chain was evaluated in terms of synchronization; the signal-generation path was kept fixed throughout the measurements and was therefore treated as a common excitation reference, with no independent synchronization analysis performed on the output channels.

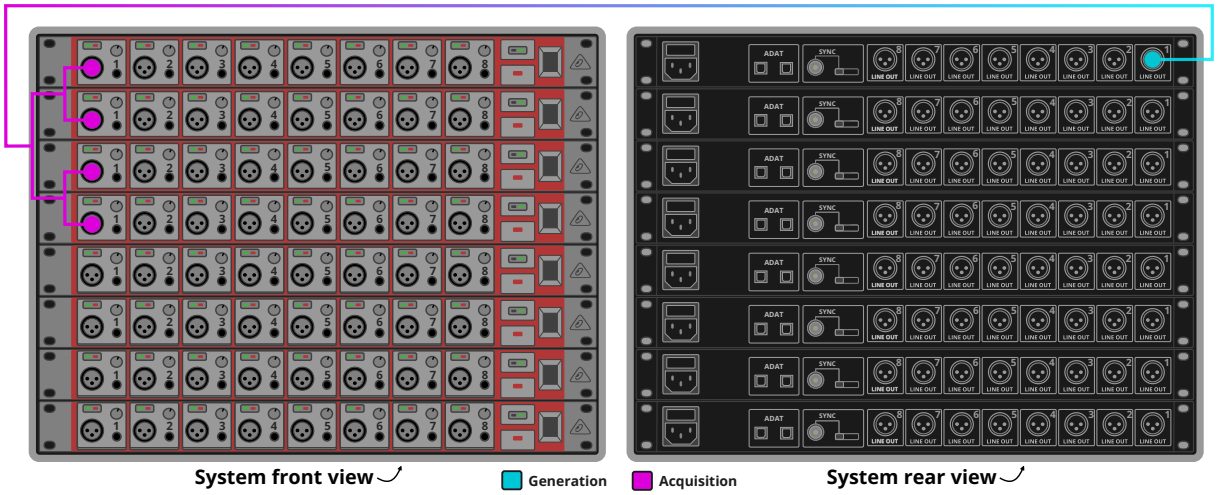


Figure 7: Loopback connection diagram — example showing measurements of Channels 1.1 (reference), 2.1, 3.1, and 4.1 using Output 1.1 as the reference source.

The investigated configurations were organized into eight experimental groups:

1. RME Digiface USB with ADAT clocking on MacOS;
2. RME Digiface USB with BNC word-clock synchronization on MacOS;
3. RME Digiface USB and PreSonus FireStudio Lightpipe with ADAT clocking on MacOS;
4. RME Digiface USB and PreSonus FireStudio Lightpipe with BNC word-clock synchronization on MacOS;
5. two RME Digiface USB interfaces with BNC word-clock synchronization on MacOS;
6. RME Digiface USB with ADAT clocking on Windows;
7. RME Digiface USB with BNC word-clock synchronization on Windows; and
8. two RME Digiface USB interfaces with BNC word-clock synchronization on Windows.

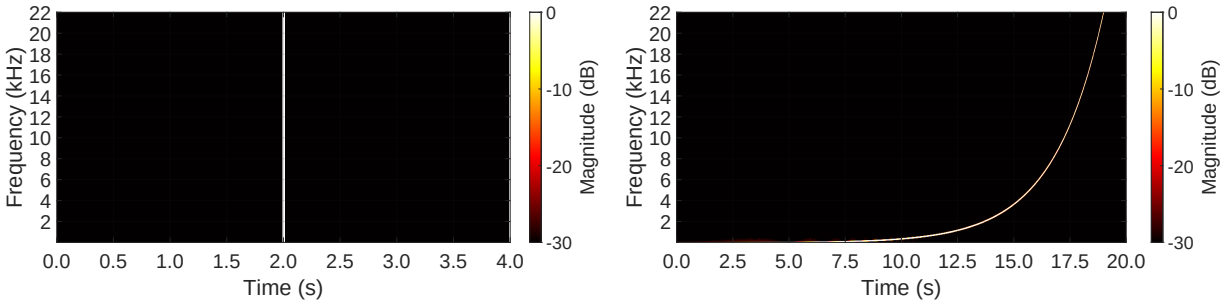
For each configuration, successive measurements were performed under controlled perturbations of the system. These perturbations included removing and reconnecting optical cables, removing and reconnecting BNC cables, turning ADAT converters off and on, restarting the

clock source, changing the sampling rate and then returning to the original value, and repeating measurements without any physical or virtual disconnection. The purpose of these tests was to determine whether the observed sample offsets were fixed, repeatable, and compensable, or whether they varied over time and after reconnection procedures.

3.3.1 Test signals

Impulse and logarithmic sine-sweep signals were used as test signals, their respective spectrograms are shown in Figure 8. Impulsive excitation allows for the direct visualization of temporal offsets in measured responses, making it suitable for identifying misalignments at the sample-level. Logarithmic sweeps, in turn, allow the estimation of impulse responses and frequency response functions with improved spectral resolution and signal-to-noise ratio.

Tonal signals were not used as the primary method for temporal-alignment assessment. Although tones may be useful for evaluating phase at a specific frequency, they can conceal integer-period delays. For example, at 1 kHz and $f_s = 44.1$ kHz, a delay of approximately 44 samples corresponds to nearly one complete period. Under this condition, two signals may appear to be in phase in a time-domain plot despite being delayed by one full cycle. Therefore, broadband test signals were preferred for temporal characterization.



(a) Impulse signal.

(b) Logarithmic sine-sweep signal (5 Hz – 22 kHz).

*Figure 8: Spectrogram of the test signals (*nWin = 1024*,* nOver = 512*, and* nFFT = 1024*).*

3.4 Data processing

The acquired signals of all channels were processed in order to estimate inter-channel delays, impulse responses, magnitude responses, and phase deviations. For each measurement, the impulse peak of each channel was compared with the reference channel, and the delay was expressed in samples. This procedure enabled the construction of tables summarizing the temporal alignment of the acquisition system under each tested condition.

Frequency response functions were obtained from the measured excitation and response signals. The resulting magnitude responses were normalized at 1 kHz when appropriate, allowing channel-to-channel spectral differences to be more clearly observed. Phase responses were analyzed in order to relate measured temporal offsets to frequency-dependent phase deviations, according to Equation (3).

3.4.1 Experimental delay estimation during acoustic measurements

Since the 64-channel system is composed of two 32-channel acquisition groups, two electrical loopback channels were used as temporal references. Channels 1 to 31 correspond to the microphone signals acquired by the first audio interface, while Channel 32 corresponds to the first loopback. Similarly, Channels 33 to 63 correspond to the microphone signals acquired by the second audio interface, while Channel 64 corresponds to the second loopback. Thus, the measured signals were divided into two acquisition groups, each one associated with its own loopback reference.

Let $x_m[n]$ denote the signal acquired by the mth microphone channel, and let $\ell_1[n]$ and $\ell_2[n]$ denote the loopback signals associated with the first and second acquisition groups, respectively. The microphone-channel sets are defined as

$$\mathscr{G}_1 = \{1,2,\ldots,31\} \qquad \text{and} \qquad \mathscr{G}_2 = \{33,34,\ldots,63\}, \tag{18}$$

whereas $\ell_1[n] = x_{32}[n]$ and $\ell_2[n] = x_{64}[n]$.

Two complementary correction procedures were considered. The first one, referred to here as the delay-based approach, estimates only the relative temporal offset between the two acquisition groups. The second one, referred to as the deconvolution-based approach, corrects the recorded acoustic signals by deconvolving each microphone group by its corresponding loopback signal. The distinction between these two procedures is important because they have different fields of application. The delay-based approach can be used in passive beamforming measurements once the inter-group delay has been estimated, whereas the deconvolution-based approach is only applicable when the source under analysis reproduces the same known sweep used as the electrical reference.

In the delay-based approach, the inter-group offset was estimated from the impulse responses of the two loopback channels. Let $s[n]$ be the reference sweep signal used in the electrical loopback characterization. The impulse responses associated with the two loopbacks were estimated by deconvolution as

$$h_{\ell_1}[n] = \ell_1[n] \oslash s[n] \qquad \text{and} \qquad h_{\ell_2}[n] = \ell_2[n] \oslash s[n], \tag{19}$$

in which $\oslash$ denotes deconvolution. The peak positions of these impulse responses were then obtained as

$$n_1 = \arg\max_n \left|h_{\ell_1}[n]\right| \qquad \text{and} \qquad n_2 = \arg\max_n \left|h_{\ell_2}[n]\right|. \tag{20}$$

The relative delay between the two acquisition groups, expressed in samples, is therefore

$$\Delta n = n_1 - n_2\,. \tag{21}$$

If $\Delta n > 0$, the peak of the second loopback occurs earlier than that of the first loopback, indicating that the second acquisition group is advanced with respect to the first one. In this case, the first $|\Delta n|$ samples of the second group are discarded. Conversely, if $\Delta n < 0$, the first group is advanced, and the first $|\Delta n|$ samples of the first group are discarded. This procedure enforces a common temporal reference between the two acquisition blocks before beamforming.

Equivalently, the trimming operation can be written in a compact form by defining

$$d_1 = \max(n_1, n_2) - n_1 \qquad \text{and} \qquad d_2 = \max(n_1, n_2) - n_2 \,, \tag{22}$$

where d_1 and d_2 are the number of samples removed from the beginning of the first and second acquisition groups, respectively. The aligned microphone signals are then given by

$$\tilde{x}_m[n] = x_m[n + d_1] \,, \qquad m \in \mathscr{G}_1 \,, \tag{23}$$

and

$$\tilde{x}_m[n] = x_m[n + d_2] \,, \qquad m \in \mathscr{G}_2 \,. \tag{24}$$

This is the most appropriate interpretation for passive acoustic imaging, since the source signal is not assumed to be known. In normal passive beamforming applications, the array only records the sound field generated by the source under investigation. Therefore, no sweep is emitted by the measurement system toward the acoustic source; the sweep is used only as an electrical diagnostic signal for estimating the temporal offset between acquisition groups.

The deconvolution-based approach follows a different rationale. In this case, the microphone signals recorded by each group are corrected by deconvolution with the corresponding loopback signal. Thus, for the first acquisition group,

$$\hat{h}_m[n] = x_m[n] \oslash \ell_1[n] \,, \qquad m \in \mathscr{G}_1 \,, \tag{25}$$

whereas for the second acquisition group,

$$\hat{h}_m[n] = x_m[n] \oslash \ell_2[n] \,, \qquad m \in \mathscr{G}_2 \,. \tag{26}$$

The resulting signals $\hat{h}_m[n]$ represent impulse-response estimates in which the excitation signal and the acquisition-chain response associated with each interface group are compensated through the corresponding loopback. In the frequency domain, this operation may be written as

$$\hat{H}_m(\omega) = \frac{X_m(\omega)}{L_g(\omega)} \,, \qquad m \in \mathscr{G}_g \,, \quad g \in \{1, 2\} \,, \tag{27}$$

where $X_m(\omega)$ is the spectrum of the mth microphone signal and $L_g(\omega)$ is the spectrum of the loopback associated with the corresponding acquisition group[6]. In practical implementations, a regularized division may be used in order to avoid numerical instabilities at frequencies where the loopback spectrum has low magnitude.

This deconvolution-based correction is only valid when the source also reproduces the same known sweep used as the reference signal. Under this condition, the acoustic path between the loudspeaker and the microphones is embedded in the measured response, and deconvolution yields an impulse-response representation suitable for subsequent frequency-domain analysis. This situation was adopted exceptionally in the experimental validation of the present article, in

[6] In the discrete implementation, the involved time-domain vectors are preconditioned so that the FFT-based spectral division represents a linear deconvolution rather than a circular operation. This includes, when necessary, using compatible vector lengths, appropriate zero-padding, regularized spectral division, and subsequent selection of the physically meaningful portion of the resulting impulse response.

which the same electrical sweep was supplied to the loudspeaker in order to generate a controlled and repeatable acoustic excitation. It should not, however, be interpreted as the standard operating condition of passive beamforming.

3.5 Operational systems

It is also important to present methodological information regarding the operating systems used, since their operation and driver management are performed differently. When using the Windows operating system, it is necessary to install the RME MADIface driver, which consists of the executables `madifaceusb_x64` and `TotalMixFX_x64`. The former is responsible for configuring the RME interfaces, while the latter manages the playback and acquisition channels. In this case, both interfaces are simultaneously recognized, with no additional configuration required for their combined use. Matlab (as well as other software) recognizes the driver as a single device, in which the playback and acquisition channels correspond to the concatenation of both devices — that is, the RME interface defined as Device 1 contains Channels 1 to 32, while the interface defined as Device 2 contains Channels 33 to 64. Figure 9 presents details regarding the aforementioned drivers and their simplified use.

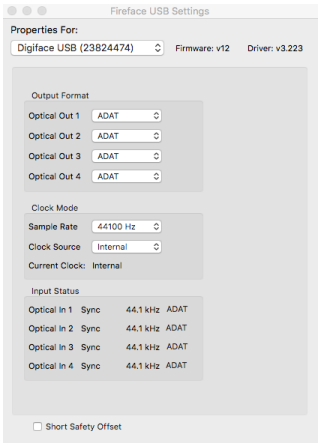
Under MacOS, each interface is recognized individually, which also enables the use of the FireWire PreSonus interface. Therefore, it becomes necessary to create a virtual device, referred to as an *aggregated device*. This procedure can be performed through the "Audio MIDI Setup" utility, in which a name must be assigned to the new device and the audio interfaces to be simultaneously used must be selected. In this case, it is possible to freely modify the order of the devices, as well as rename their respective channels. The software used for playback and acquisition will then recognize a new device with the chosen name, composed of the selected interfaces. The aggregated device configuration can be observed in Figure 10. It is important to emphasize that, similarly to the Windows case, the appropriate audio drivers must also be installed in order for the operating system to correctly recognize the connected interfaces. For documentation purposes, the versions of the operating systems and drivers are detailed in Table 3.

Table 3: System software and driver versions.

Operating system	Matlab	RME driver	PreSonus driver
Windows 11, version 25H2	R2023b	v1.02	not applicable
MacOS High Sierra, version 10.13.6	R2020a	v3.223	4.2.1.464375875
MacOS Tahoe, version 26.2	R2025a	v4.27	not applicable

4 Results and discussion

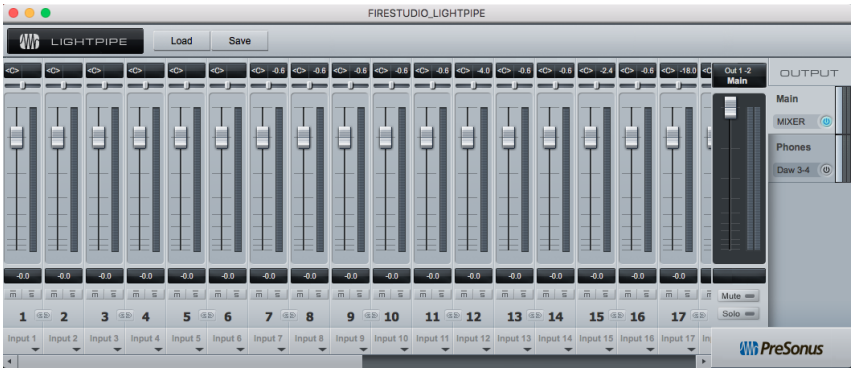
This section presents the experimental results obtained from the acquisition configurations evaluated. The discussion focuses on temporal alignment, frequency-response consistency, synchronization stability, and the practical implications of using ADAT or BNC clocking in a 64-channel acoustic imaging system. A dedicated acoustic imaging application using the proposed system is discussed in Section 5.



(a) RME Madiface USB.

(b) RME TotalMix FX.

(c) PreSonus Universal Control.



(d) PreSonus FireStudio Lightpipe.

Figure 9: Graphical User Interface of RME and Presonus drivers.

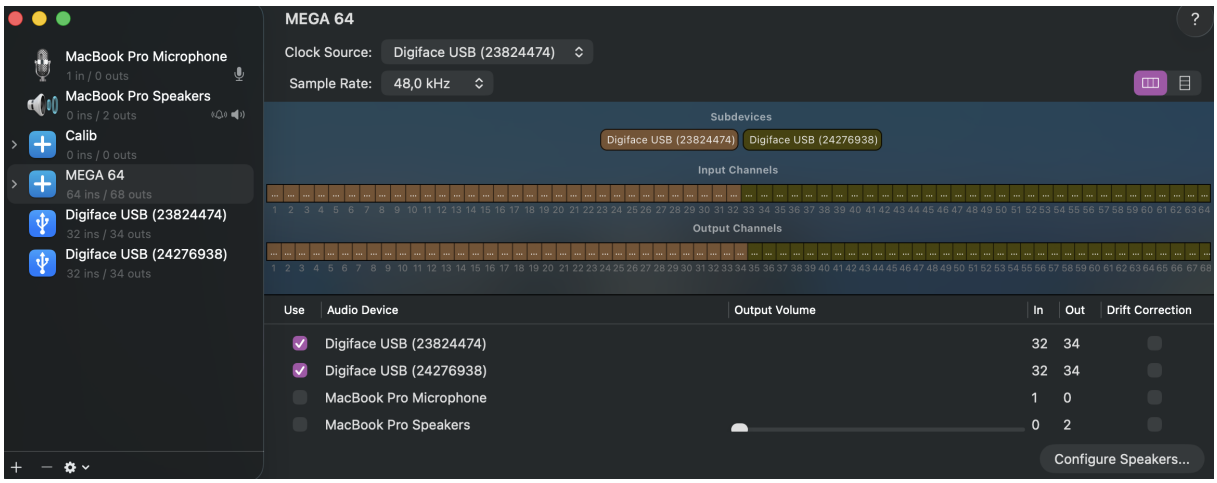
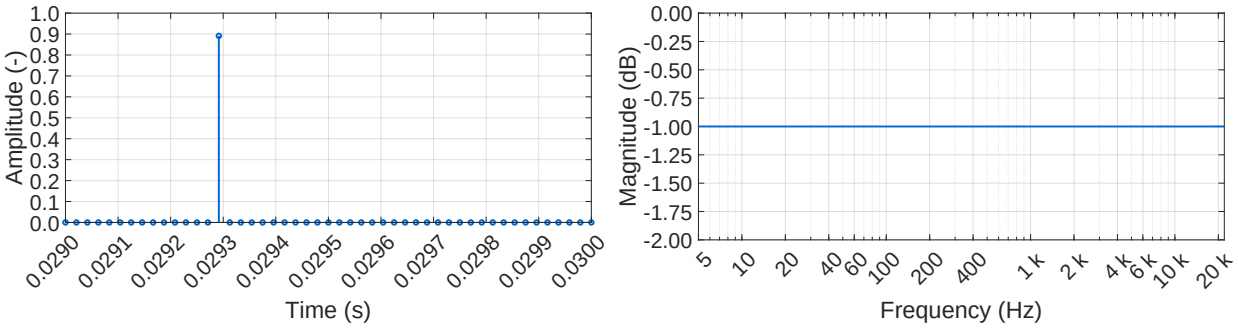


Figure 10: MacOS audio device settings.

4.1 Preliminary loopback measurements

Preliminary loopback measurements were performed in order to verify the basic behavior of the acquisition chain. Figure 11 shows a representative optical-loopback ADAT measurement in the time and frequency domains. These measurements were used to verify signal integrity, gain consistency, and the suitability of the measurement procedure before evaluating the complete multichannel configurations.

It is important to emphasize that, in this case, there is no signal conversion between the measured optical output and input, that is, the signal is generated digitally and also acquired in the digital domain. For this reason, the time-domain plot was generated with emphasis on the discrete digital samples. The obtained result corresponds to a perfectly shifted impulse, corroborating the analysis that the optical protocol did not introduce any significant influence on the measurements.



(a) Time-domain response. **(b)** Frequency-domain response.

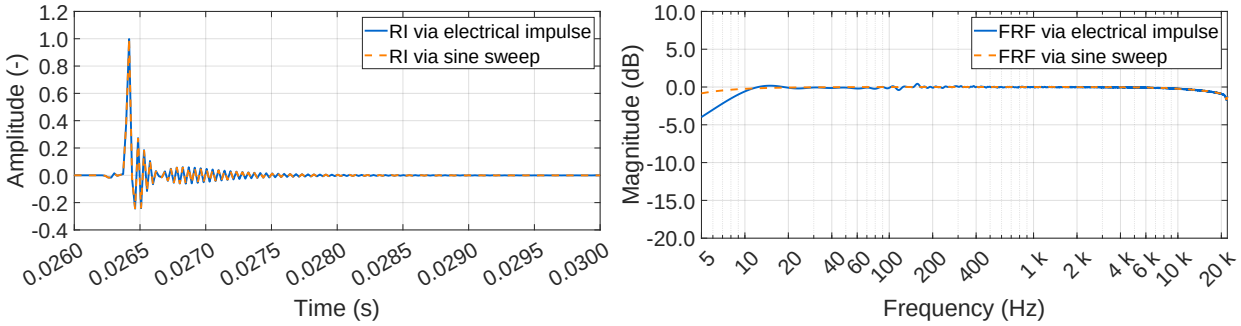
Figure 11: Representative optical-loopback ADAT measurement used for preliminary verification of the acquisition chain.

Furthermore, before comparing impulse responses between channels and converters, it is necessary to establish a relationship between the test signals and the quality and reliability of the results obtained. In general acoustic measurements, logarithmic sine sweeps are commonly adopted, with impulse responses being obtained through deconvolution. However, in the context of electrical measurements, it is also appropriate to perform tests using impulsive signals, as presented in Section 3.3.1.

Accordingly, Figure 12 illustrates impulse-response estimates obtained from electrical loopback measurements performed through the XLR output and input connections, using both impulsive excitation and logarithmic-sweep deconvolution. The results indicate that both test signals proved to be suitable for the proposed applications. Therefore, logarithmic sine sweeps were adopted for frequency response function (FRF) measurements, whereas impulsive signals were also adopted for delay measurements.

4.2 Sample-level temporal offsets

Figure 13 contrasts two representative temporal-alignment scenarios. When comparing channels routed through a single AD/DA converter and audio interface, the impulse responses exhibit strict alignment within the discrete-time resolution. Conversely, cross-device evaluations reveal



(a) Normalized time-domain response after temporal alignment.

(b) Magnitude of the frequency response function, windowed and normalized at 1 kHz.

Figure 12: Comparison between impulse-response estimates obtained from impulsive excitation and logarithmic-sweep measurements (electrical).

measurable sample-level offsets between distinct hardware units. Although these temporal discrepancies may appear negligible, their resulting phase deviations amplify significantly at higher frequencies, in accordance with Equation (3). Consequently, this comparative analysis establishes a critical baseline for characterizing the synchronization integrity across the entire multichannel acquisition architecture.

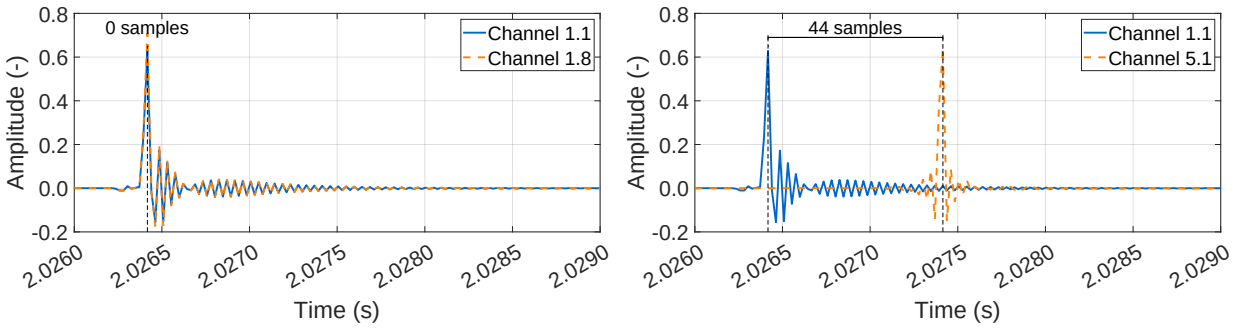
Still regarding the importance of selecting appropriate test signals for result analysis, Figure 14 presents the same type of measurement shown in Figure 13, but now using a 1 kHz sinusoidal test signal. It can be observed that, coincidentally, the 44-sample delay between the impulse responses corresponds exactly to one complete cycle of the frequency used in the measurement, giving the false impression that the signals are aligned. This result reinforces the need for impulse- or sweep-based procedures when validating multichannel acquisition systems for acoustic imaging.

The same results may also be interpreted through the deconvolution between the compared channels. Figure 15 shows the corresponding phase spectra obtained from the deconvolution of the channel pairs presented in the previous plots (Figure 13). When the impulse responses are inherently aligned, as in the case of channels associated with the same AD/DA converter, the phase spectrum remains essentially null. By contrast, when a relative delay of 44 samples is present, the phase response exhibits the periodic pattern characteristic of a comb filter, in agreement with the theoretical behavior discussed earlier.

The three figures are based on illustrative measurements comparing Channels 1.1 and 5.1, connected to the RME and PreSonus interfaces, respectively, in a MacOS environment with ADAT clock synchronization. These boundary conditions are stated only for contextual clarity, since their synchronization behavior is examined in detail in the subsequent results. Accordingly, the selected channels and operating conditions should be interpreted as examples of the measurement and analysis procedures, rather than as a complete assessment of this configuration.

In order to avoid redundancy in the presentation of plots such as Figure 13 for all analyzed channels, tables were developed containing the types of measurements performed and their respective results, including both informative and interpretative analyses. Table 4 presents the results for all analyzed scenarios, along with their specifications and corresponding analyses regarding the obtained results and system viability.

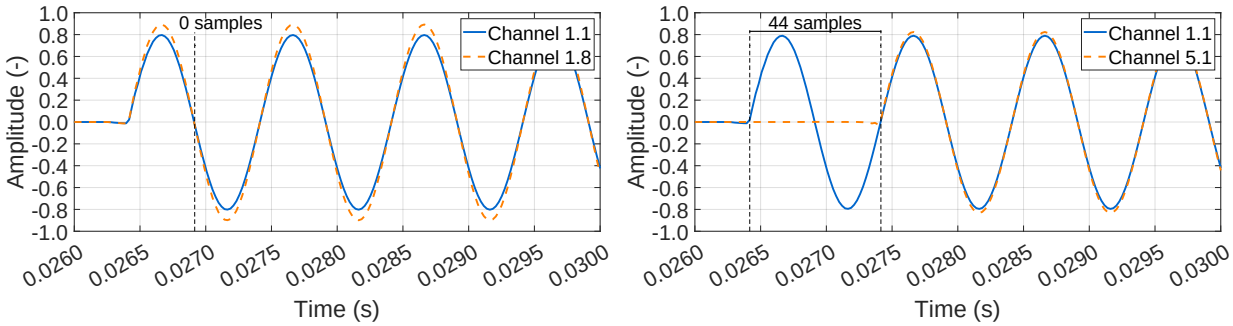
More specifically, Table 5 shows the relative sample delays among the measured channels for all experiments and boundary conditions analyzed. The channels measured in each scenario



(a) Temporally aligned responses.

(b) Responses with sample-level temporal offsets.

Figure 13: Time-domain comparison of measured impulse responses with and without sample-level temporal offsets.



(a) Apparently aligned 1 kHz tonal signals.

(b) Delayed 1 kHz tonal signals that may appear aligned because of an integer-period offset.

Figure 14: Limitation of tonal signals for temporal-alignment assessment. Integer-period delays may be visually masked in narrowband analysis.

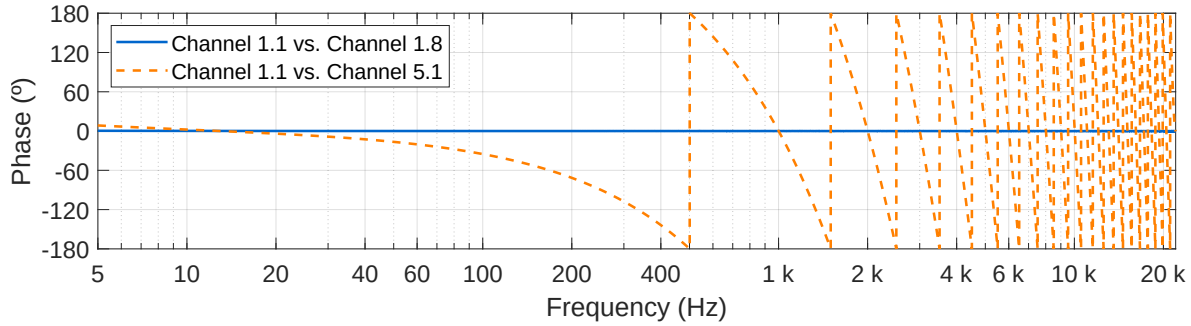


Figure 15: Phase deviation associated with measured temporal delays, obtained from sweep-based deconvolution.

are listed in Table 4, and the delay of the first channel is always zero, since it serves as the reference channel. The rows in the table are divided into groups by horizontal bars, each bar indicating a physical variation in the measurement conditions. That is, measurements 1 to 10

were performed sequentially under an initial configuration. Intentional perturbations were then introduced: power cycling the audio interfaces and the ADA8200 converters, disconnecting and reconnecting TOSLINK cables, BNC cables, and USB cables, changing sampling rates, modifying the clock source, as well as the synchronization attempt procedures described in Section 4.4.3. After each set of perturbations, a new sequence of measurements (e.g., 11 to 20) was acquired. This process was repeated successively, generating consecutive blocks of measurements, allowing the assessment of temporal invariance of the delays and the effect of physical reconfigurations on the inter-channel offsets.

4.3 Amplitude differences among channels

The frequency responses of the Behringer ADA8200 converters were evaluated through loopback measurements. Figure 16 presents the transfer functions between each measured channel and the reference channel, defined as Channel 1.1. These measurements allow channel-dependent magnitude deviations to be identified and provide a basis for evaluating whether equalization or calibration adjustments are necessary prior to acoustic imaging measurements. For the results shown, the gain controls of all channels on all converter units were set as close as possible to the 12 o'clock position.

From the obtained FRF results, it can be observed that the converter channels exhibit a substantially flat frequency response within the audible range from 20 Hz to 20 kHz. When directly compared to the reference channel, the measured responses presented amplitude deviations of up to approximately 5 dB among channels. These variations are mainly attributed to differences in channel input sensitivity among the converters and, inevitably, to intrinsic positioning errors of the gain controls due to human factors. However, such variations primarily produce vertical offsets in the curves and can therefore be compensated through indirect adjustment of the microphones connected to the converters. After normalization of the measured responses through this adjustment procedure, the residual variations among channels remained below 1 dB within the envelope of the presented curves.

4.4 Temporal stability

Temporal stability was evaluated by comparing repeated measurements performed with and without physical or virtual perturbations of the acquisition system. In this context, the relevant metrological criterion is not merely the presence of inter-channel delays, but their invariance over time. A fixed delay, even when measurable at the sample level, can be estimated and compensated through loopback references. Conversely, a non-repeatable delay prevents the establishment of a stable temporal reference and therefore compromises the reliability of subsequent compensation procedures.

This distinction is particularly important in a 64-channel architecture assembled from multiple converter and interface groups. In such a system, temporal errors may arise at different hierarchical levels, including individual channels, converter units, ADAT groups, and audio interfaces. Therefore, the interpretation of the measured delays must consider not only their magnitude, but also whether they are associated with converter-level misalignment, interface-level latency, or non-repeatable synchronization behavior.

The experimental results show that some configurations exhibited sample-level offsets among channels or converter groups. When these offsets remained invariant after system initialization,

Table 4: Summary of the inter-channel delay experiments. Channel 1.1 was adopted as the reference channel in all cases.

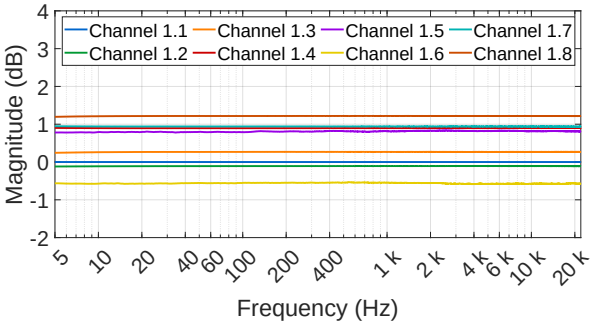
Exp.	SUT	OS	Clock	Channels	Result	Feasibility
1	RME	MacOS	ADAT	1.1 / 2.1 / 3.1 / 4.1	Delay of approximately ±1 sample among channels of each converter, invariant over time.	Partial
2	RME	MacOS	BNC	1.1 / 2.1 / 3.1 / 4.1	Channels synchronized in all measurements and boundary conditions.	Yes
3	RME + PS	MacOS	ADAT	1.1 | 5.1 / 6.1 / 7.1	Delay of approximately ±1 sample among channels of each converter, in addition to interface latency differences; not invariant over time.	No
4	RME + PS	MacOS	BNC	1.1 | 5.1 / 6.1 / 7.1	Sample difference attributed exclusively to latency between interfaces, with no converter-level misalignment within each interface; invariant over time.	Yes
5	RME + RME	MacOS	BNC	1.1 / 2.1 | 5.1 / 6.1	Sample difference attributed exclusively to latency between interfaces, with no converter-level misalignment within each interface; invariant over time.	Yes
6	RME	Windows	ADAT	1.1 / 2.1 / 3.1 / 4.1	Delay of approximately ±1 sample among channels of each converter, invariant over time.	Partial
7	RME	Windows	BNC	1.1 / 2.1 / 3.1 / 4.1	Channels synchronized in all measurements and boundary conditions.	Yes
8	RME + RME	Windows	BNC	1.1 / 2.1 | 5.1 / 6.1	Sample difference attributed exclusively to latency between interfaces, with no converter-level misalignment within each interface; invariant over time.	Yes

The feasibility categories adopted in Table 4 are interpreted as follows:

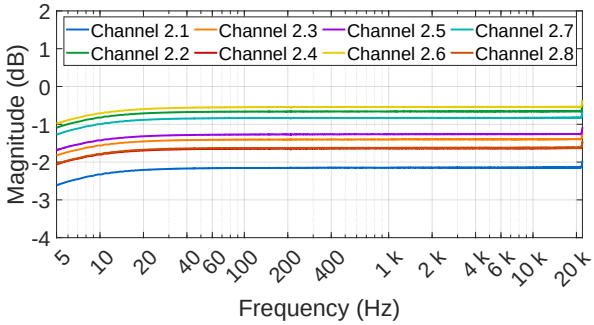
- **Yes**: the configuration requires only one loopback per ADA8200 group, or one loopback per audio interface, depending on the system topology;
- **Partial**: the configuration requires one loopback per ADA8200 converter; and
- **No**: the configuration does not exhibit temporal invariance, remaining susceptible to sample-level variations associated with the synchronization process.

Table 5: Delay measurement results in samples (each column corresponds to one experiment).

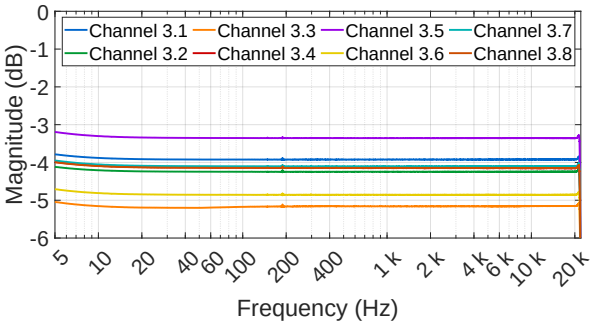
Meas	Exp 1	Exp 2	Exp 3	Exp 4	Exp 5	Exp 6	Exp 7	Exp 8
1	0 0 1 0	0 0 0 0	0 11 11 11	0 9 9 9	0 0 2 2	0 0 -1 0	0 0 0 0	0 0 -46 -46
2	0 0 1 0	0 0 0 0	0 -28 -28 -28	0 9 9 9	0 0 2 2	0 0 -1 0	0 0 0 0	0 0 -46 -46
3	0 0 1 0	0 0 0 0	0 -28 -28 -29	0 9 9 9	0 0 2 2	0 0 -1 0	0 0 0 0	0 0 -46 -46
4	0 0 1 0	0 0 0 0	0 -29 -29 -29	0 9 9 9	0 0 2 2	0 0 -1 0	0 0 0 0	0 0 -46 -46
5	0 0 1 0	0 0 0 0	0 -30 -29 -30	0 9 9 9	0 0 2 2	0 0 -1 0	0 0 0 0	0 0 -46 -46
6	0 0 1 0	0 0 0 0	0 -30 -30 -30	0 9 9 9	0 0 2 2	0 0 -1 0	0 0 0 0	0 0 -46 -46
7	0 0 1 0	0 0 0 0	0 -30 -30 -30	0 9 9 9	0 0 2 2	0 0 -1 0	0 0 0 0	0 0 -46 -46
8	0 0 1 0	0 0 0 0	0 -31 -31 -31	0 9 9 9	0 0 2 2	0 0 -1 0	0 0 0 0	0 0 -46 -46
9	0 0 1 0	0 0 0 0	0 -31 -31 -31	0 9 9 9	0 0 2 2	0 0 -1 0	0 0 0 0	0 0 -46 -46
10	0 0 1 0	0 0 0 0	0 -31 -31 -31	0 9 9 9	0 0 2 2	0 0 -1 0	0 0 0 0	0 0 -46 -46
11	0 0 0 0	0 0 0 0	0 -31 -31 -32	0 25 25 25	0 0 2 2	–	–	0 0 -46 -46
12	0 0 0 0	0 0 0 0	0 -32 -32 -32	0 25 25 25	0 0 2 2	–	–	0 0 -46 -46
13	0 0 0 0	0 0 0 0	0 -32 -32 -32	0 25 25 25	0 0 2 2	–	–	0 0 -46 -46
14	0 0 0 0	0 0 0 0	0 -32 -32 -32	0 25 25 25	0 0 2 2	–	–	0 0 -46 -46
15	0 0 0 0	0 0 0 0	0 -33 -32 -33	0 25 25 25	0 0 2 2	–	–	0 0 -46 -46
16	0 0 0 0	0 0 0 0	0 -33 -33 -33	0 25 25 25	0 0 2 2	–	–	0 0 -46 -46
17	0 0 0 0	0 0 0 0	0 -33 -33 -33	0 25 25 25	0 0 2 2	–	–	0 0 -46 -46
18	0 0 0 0	0 0 0 0	0 -34 -34 -34	0 25 25 25	0 0 2 2	–	–	0 0 -46 -46
19	0 0 0 0	0 0 0 0	0 -34 -34 -34	0 25 25 25	0 0 2 2	–	–	0 0 -46 -46
20	0 0 0 0	0 0 0 0	0 -34 -34 -34	0 25 25 25	0 0 2 2	–	–	0 0 -46 -46
21	0 -1 0 0	0 0 0 0	0 -35 -35 -35	0 25 25 25	0 0 2 2	–	–	0 0 -46 -46
22	0 -1 0 0	0 0 0 0	0 -35 -35 -35	0 25 25 25	0 0 2 2	–	–	0 0 -46 -46
23	0 -1 0 0	0 0 0 0	0 -36 -36 -36	0 25 25 25	0 0 2 2	–	–	0 0 -46 -46
24	0 -1 0 0	0 0 0 0	0 -36 -36 -36	0 25 25 25	0 0 2 2	–	–	0 0 -46 -46
25	0 -1 0 0	0 0 0 0	0 -36 -36 -36	0 25 25 25	0 0 2 2	–	–	0 0 -46 -46
26	0 -1 0 0	0 0 0 0	0 -37 -36 -37	0 25 25 25	0 0 4 4	–	–	0 0 -46 -46
27	0 -1 0 0	0 0 0 0	0 -37 -37 -37	0 25 25 25	0 0 4 4	–	–	0 0 -46 -46
28	0 -1 0 0	0 0 0 0	0 -37 -37 -37	0 25 25 25	0 0 4 4	–	–	0 0 -46 -46
29	0 -1 0 0	0 0 0 0	0 -37 -37 -37	0 25 25 25	0 0 8 8	–	–	0 0 -46 -46
30	0 -1 0 0	0 0 0 0	0 -37 -37 -38	0 25 25 25	0 0 8 8	–	–	0 0 -46 -46
31	0 0 0 -1	0 0 0 0	–	0 44 44 44	0 0 103 103	–	–	0 0 338 338
32	0 0 0 -1	0 0 0 0	–	0 24 24 24	0 0 2 2	–	–	0 0 -574 -574
33	0 0 0 -1	0 0 0 0	–	0 170 170 170	0 0 2 2	–	–	0 0 338 338
34	0 0 0 -1	0 0 0 0	–	0 12 12 12	0 0 2 2	–	–	0 0 2 2
35	0 0 0 -1	0 0 0 0	–	0 5 5 5	0 0 2 2	–	–	0 0 50 50
36	0 0 0 -1	0 0 0 0	–	0 12 12 12	0 0 2 2	–	–	0 0 -46 -46
37	0 0 0 -1	0 0 0 0	–	0 5 5 5	–	–	–	–
38	0 0 0 -1	0 0 0 0	–	0 6 6 6	–	–	–	–
39	0 0 0 -1	0 0 0 0	–	0 7 7 7	–	–	–	–
40	0 0 0 -1	0 0 0 0	–	–	–	–	–	–
41	0 1 1 1	0 0 0 0	–	–	–	–	–	–
42	0 1 1 1	0 0 0 0	–	–	–	–	–	–
43	0 1 1 1	–	–	–	–	–	–	–
44	0 1 1 1	–	–	–	–	–	–	–
45	0 1 1 1	–	–	–	–	–	–	–
46	0 1 1 1	–	–	–	–	–	–	–
47	0 1 1 1	–	–	–	–	–	–	–
48	0 1 1 1	–	–	–	–	–	–	–
49	0 1 1 1	–	–	–	–	–	–	–
50	0 1 1 1	–	–	–	–	–	–	–



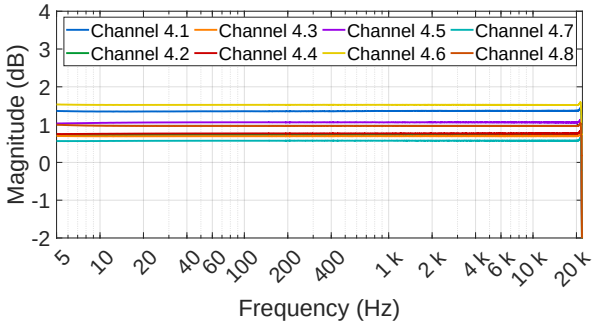
(a) Behringer ADA8200 No. 1.



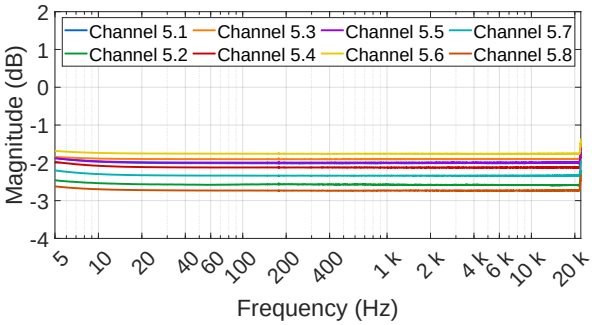
(b) Behringer ADA8200 No. 2.



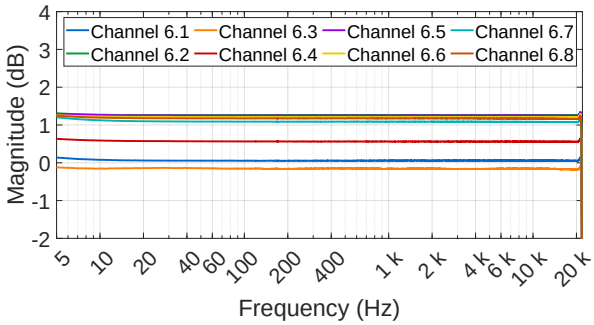
(c) Behringer ADA8200 No. 3.



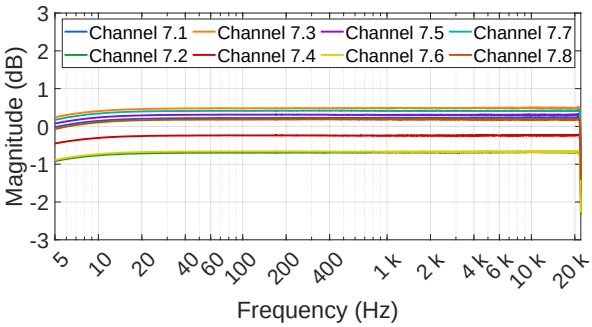
(d) Behringer ADA8200 No. 4.



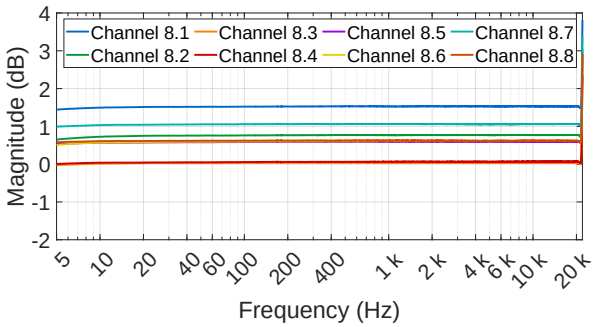
(e) Behringer ADA8200 No. 5.



(f) Behringer ADA8200 No. 6.



(g) Behringer ADA8200 No. 7.



(h) Behringer ADA8200 No. 8.

Figure 16: Transfer functions between the channels of the Behringer ADA8200 converters and the reference channel, defined as Channel 1.1.

the corresponding configuration was considered partially viable, since the delays could be compensated by assigning loopback references to the appropriate hardware groups. However, when the offsets varied over time, compensation became unreliable, independently of the nominal magnitude of the observed delay. The results summarized in Table 4 therefore indicate that the feasibility of each topology depends primarily on delay invariance rather than on the mere occurrence of successful multichannel acquisition.

4.4.1 Influence of the synchronization strategy

The comparison between ADAT-based and BNC word-clock synchronization clarifies the origin of the temporal behaviors observed in the previous analysis. Both strategies enabled functional digital audio transfer, but they differed substantially in their suitability for repeatable acoustic measurements. ADAT synchronization has the practical advantage of embedding clock information within the optical data stream itself, thereby reducing cabling complexity. However, because the clock must be recovered by the optical receiver and associated PLL circuitry, residual temporal-alignment uncertainties may occur among converter units.

In the configurations based on a single RME interface using ADAT synchronization, namely Experiments 1 and 6, sample-level offsets of approximately ± 1 sample were observed among converter groups. These offsets were generally invariant after system initialization. Although this condition does not correspond to perfect inter-converter alignment, it remains experimentally manageable when one loopback reference is used for each Behringer ADA8200 converter. Under this condition, the loopback channels provide the necessary information to estimate and compensate fixed converter-level delays during post-processing.

A less favorable behavior was observed when the PreSonus and RME interfaces were combined through ADAT synchronization, as in Experiment 3. In this case, the measured delays did not remain invariant over time. The temporal non-invariance can be observed in Table 5, particularly in rows 1 to 10, in which delay variations occur between the reference channel and the measured channels. In addition, the channels connected to the PreSonus interface exhibited an alternation of approximately ± 1 sample. Because these variations were not repeatable, this configuration did not provide a stable basis for delay compensation and was therefore considered unsuitable for repeatable acoustic measurements.

In contrast, BNC word-clock synchronization exhibited a more robust temporal behavior in all evaluated configurations. As indicated by Experiments 2, 4, 5, 7, and 8 in Table 4, the converter groups associated with each audio interface remained internally aligned, regardless of the operating system employed or the number of interfaces used. In configurations involving more than one audio interface, fixed latency differences between interfaces were still observed. However, these differences remained temporally invariant[7] throughout the measurements and were attributed to interface-level latency rather than to the individual ADAT converters.

Therefore, for the proposed 64-channel acquisition system, BNC word-clock synchronization represents the preferable strategy. Its main advantage is not the complete elimination of latency offsets, but the transformation of the synchronization problem into a stable and compensable

[7]The observed invariance is associated with the initialization conditions of the system. Variations in latency among the audio interfaces may naturally occur after power cycling or reconnecting the system. Nevertheless, once the interfaces were initialized and operating without physical modifications to the system configuration, the measured latency offsets remained invariant throughout the performed measurements.

one. Under this topology, converter groups remain internally aligned, while residual latency differences occur primarily between audio-interface groups. As a result, a single loopback reference per audio-interface group becomes sufficient, reducing experimental complexity and increasing the reliability of acoustic-imaging measurements.

4.4.2 Absence of proper clock synchronization

Measurements performed without a properly defined synchronization strategy, that is, without clock synchronization either via ADAT or via BNC, showed that clocking problems may lead not only to delays, but also to time-varying amplitude and phase responses. Such behavior is incompatible with acoustic-imaging applications, since the measured differences among channels can no longer be attributed solely to the acoustic field.

Figure 17 shows the visual indication of failed synchronization attempts between the ADA8200 converters and the RME Digiface USB. Blinking synchronization indicators reveal that the converters were repeatedly attempting to lock to the clock source. The figure is implemented as an animation, and playback controls are provided so that the user can start and pause the visualization.

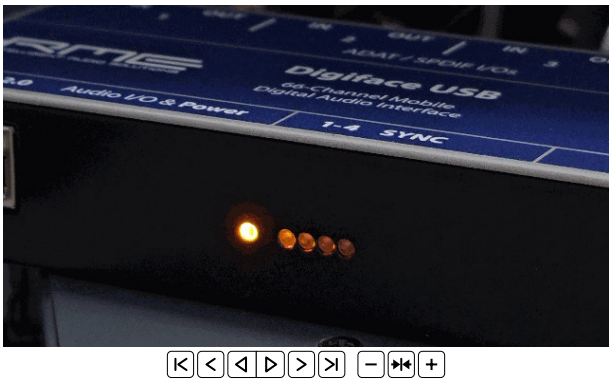
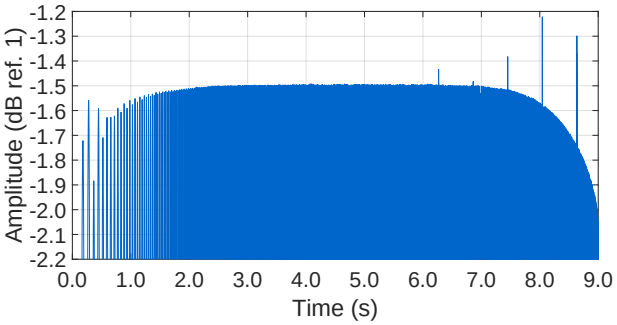


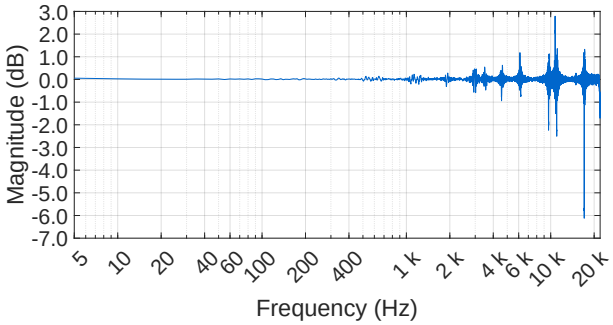
Figure 17: Visual indication of synchronization attempts between the ADA8200 converters and the RME Digiface USB. Blinking indicators denote unsuccessful synchronization attempts. The sequence is presented as an animation with playback controls.

The absence of proper synchronization produces spurious signal peaks in the time domain, consequently degrading the frequency response of the system. This behavior can be observed in both domains, see Figure 18. In the time-domain response, abrupt discontinuities and irregular peaks become evident, whereas the corresponding frequency response exhibits substantial spectral inconsistencies and magnitude deviations when compared to the synchronized conditions previously presented.

Additionally, a qualitative correspondence could be observed between the blinking synchronization indicators shown in Figure 17 and the appearance of spurious pulses in the measured signals. Visually, the temporal occurrence of such artifacts appeared to coincide with the repeated synchronization attempts performed by the converters. However, no direct measurement or synchronization-monitoring procedure was implemented to formally establish this relationship, and therefore this observation should be interpreted only as an indicative qualitative assessment.



(a) Time-domain response obtained from a sweep measurement.



(b) Magnitude of the frequency response function, normalized at 1 kHz.

Figure 18: Representative measurement obtained under improper clock synchronization, showing temporal and spectral inconsistencies.

4.4.3 Synchronization-attempt procedures

Two procedures were tested in an attempt to eliminate sample-level offsets associated with PLL locking and clock recovery.

- **Procedure 1: simultaneous restart of the ADAT converters.** This procedure was intended to initialize the clock-recovery process simultaneously in all converters:
  1. all ADAT converters were simultaneously turned off using the main power switch, while the audio interfaces and remaining devices remained powered;
  2. the system was left idle for 30 to 60 seconds;
  3. all ADAT converters were simultaneously turned on using the main power switch;
  4. the system was left idle for another 30 to 60 seconds; and
  5. new measurements were performed.
- **Procedure 2: restart of the clock-generation condition.** This procedure was intended to perturb the master-clock generation and force a new synchronization state:
  1. the serial connection was disconnected and Matlab was closed;
  2. the sampling rate and clock rate were changed using the control software of the master-clock interface;
  3. the system was left idle for 30 to 60 seconds;
  4. the original sampling rate and clock rate were restored;
  5. the system was left idle for another 30 to 60 seconds;
  6. Matlab was reopened and the serial connection was re-established; and
  7. new measurements were performed.

Neither procedure consistently eliminated the observed sample-level offsets. This result indicates that the offsets are associated with the synchronization and clock-recovery behavior of the system rather than with a simple deterministic initialization sequence. Consequently, the most reliable strategy is not to depend on a reconnection procedure, but to adopt a synchronization topology that yields invariant delays and to measure the remaining offsets through loopback references.

4.5 General discussion

The findings confirm that the proposed 64-channel system is viable for acoustic imaging, provided the synchronization method meets the temporal-stability demands of multichannel acquisition. The decisive aspect, as established, is not the mere presence of inter-channel delays but their invariance over time. Fixed offsets can be corrected via loopback-based adjustment procedures; non-repeatable fluctuations, however, undermine phase consistency across separate measurement sessions.

When relying on ADAT clocking, the specific hardware/software combination (interfaces, converters, OS, and clock master) plays a critical role. With a single RME interface, offsets remained constant after startup — enabling compensation through one loopback per ADA8200 converter, albeit with added adjustment effort. Mixing different interfaces (e.g., RME plus PreSonus) under ADAT produced unstable behavior: delays varied between runs and even across channels within the same converter group, rendering reliable compensation impractical for repeatable measurements.

Conversely, BNC word-clock synchronization proved markedly more stable. Across all tested configurations, converter groups stayed internally aligned regardless of OS or interface count. Any residual latencies were fixed, interface-specific, and unrelated to ADAT converter errors. As a result, the adjustment procedure becomes straightforward: one loopback reference per audio interface (i.e., per ADAT converter group) suffices.

For the specific 64-channel system implemented in this work, which relies on BNC word-clock synchronization with two audio interfaces (each connected to four ADA8200 converters), only one loopback channel per interface was required — totaling two loopbacks. Consequently, 62 usable channels remain available for acoustic imaging. It should be noted that this number depends on the chosen synchronization strategy. Had ADAT clocking been adopted, one loopback channel would have been required for each individual ADA8200 converter, resulting in eight loopbacks in total and reducing the number of usable channels to 56. Thus, the synchronization method affects not only temporal stability, but also the effective channel count of the acquisition system.

The experimental results also demonstrated the importance of selecting appropriate excitation signals for synchronization validation and temporal analysis. Broadband excitation signals, particularly impulsive signals and logarithmic sine sweeps, proved significantly more suitable than tonal excitations for identifying inter-channel temporal offsets. As experimentally demonstrated, sinusoidal signals may conceal delays corresponding to integer multiples of the analyzed frequency period, producing misleading indications of temporal alignment at specific frequencies. In contrast, analyses based on impulse responses and sweep deconvolution directly reveal the actual temporal displacements between channels and their corresponding phase behavior across the frequency spectrum. Therefore, adjustment procedures intended for acoustic-imaging

applications should preferentially employ broadband excitations and impulse-response-based analyses.

Finally, the measurements performed under intentionally improper synchronization conditions demonstrated that the absence of a stable clock reference may produce not only inter-channel delays, but also severe temporal and spectral inconsistencies, including spurious impulsive artifacts and non-repeatable deviations in the system frequency response. Such effects reinforce the importance of experimentally validating synchronization stability prior to practical acoustic-imaging measurements, especially in large-scale multichannel systems based on cascaded digital converters.

5 Acoustic imaging application

For the experimental application of the proposed multichannel acquisition system, an arc-shaped microphone array was assembled and used as a spatial sampling aperture for acoustic imaging. The array consisted of 61 microphones distributed along a circular arc in the XZ plane, with radius $R = 1$ m and angular positions uniformly spanning the interval from 0° to 90°. Thus, although the acquisition system provided 64 available channels, only 61 channels were assigned to the microphones composing the array, while the remaining channels were reserved for loopback and reference measurements. Observe Figure 19 where the experimental chain is depicted. The adopted channel distribution was defined as follows:

- Channels 1 to 31: microphones 1 to 31;
- Channel 32: loopback 1;
- Channels 33 to 62: microphones 32 to 61;
- Channel 63: reference microphone, positioned outside the array; and
- Channel 64: loopback 2.

5.1 Array geometry

The arc geometry was defined from the angular coordinate β, measured from the horizontal plane toward the positive z-axis. For an array placed in the XZ plane, the position of the m-th microphone may be written as

$$\mathbf{r}_m = \begin{bmatrix} R\cos\beta_m & 0 & R\sin\beta_m \end{bmatrix}^{\mathrm{T}} \quad \text{with} \quad m = \{1, 2, \ldots, 61\}. \tag{28}$$

Since the array spans a quarter of a circle, the total angular aperture is

$$\Delta\beta_{\mathrm{tot}} = 90^\circ = \frac{\pi}{2}\ \mathrm{rad}. \tag{29}$$

Consequently, the corresponding arc length is

$$L_{\mathrm{arc}} = R\,\Delta\beta_{\mathrm{tot}} = 1 \cdot \frac{\pi}{2} = 1.5708\ \mathrm{m}. \tag{30}$$

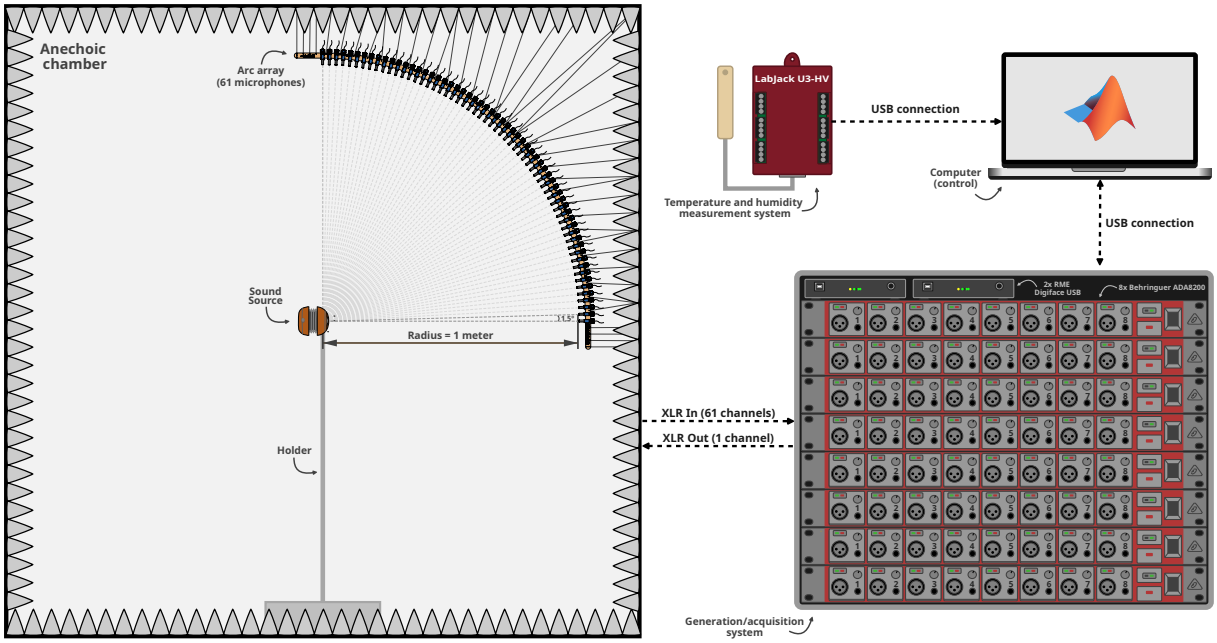


Figure 19: Experimental chain of the microphone array application (61 microphones, X-Mini sound source, and small anechoic chamber).

With $M = 61$ microphones uniformly distributed along this aperture, the angular spacing between adjacent microphones is

$$\Delta\beta = \frac{\Delta\beta_{\text{tot}}}{M-1} = \frac{90^\circ}{60} = 1.5^\circ = 0.02618 \text{ rad}. \tag{31}$$

The corresponding inter-microphone spacing along the arc is therefore

$$d_{\text{arc}} = R\Delta\beta = 1 \cdot 0.02618 = 0.02618 \text{ m}. \tag{32}$$

Equivalently, the chord distance between two adjacent microphones is

$$d_{\text{chord}} = 2R\sin\left(\frac{\Delta\beta}{2}\right) = 2\sin\left(\frac{0.02618}{2}\right) = 0.02618 \text{ m}, \tag{33}$$

which is practically identical to d_{arc} for this small angular increment.

The spatial-aliasing limit of the arc array may be estimated from the spatial Nyquist criterion, according to which the spacing between adjacent sensors should not exceed half of the acoustic wavelength. By adopting the adjacent arc spacing as the characteristic spatial sampling interval, the maximum frequency before the onset of spatial aliasing is approximated by

$$f_{\text{alias}} = \frac{c_0}{2\,d_{\text{arc}}}. \tag{34}$$

For $c_0 = 343$ m/s and $d_{\text{arc}} = 0.02618$ m, this gives

$$f_{\text{alias}} = \frac{343}{2 \cdot 0.02618} \approx 6550.8 \text{ Hz}. \tag{35}$$

Thus, the theoretical spatial-aliasing threshold of the implemented arc array is approximately 6.55 kHz. In practice, frequencies slightly below this value are preferable for robust acoustic imaging, since the effective spatial bandwidth may also be influenced by the finite aperture, the nonuniform angular projection of the array with respect to different source directions, positioning tolerances, microphone phase mismatch, and the signal-to-noise ratio of the measurement.

5.2 Experimental arrangement and measurement conditions

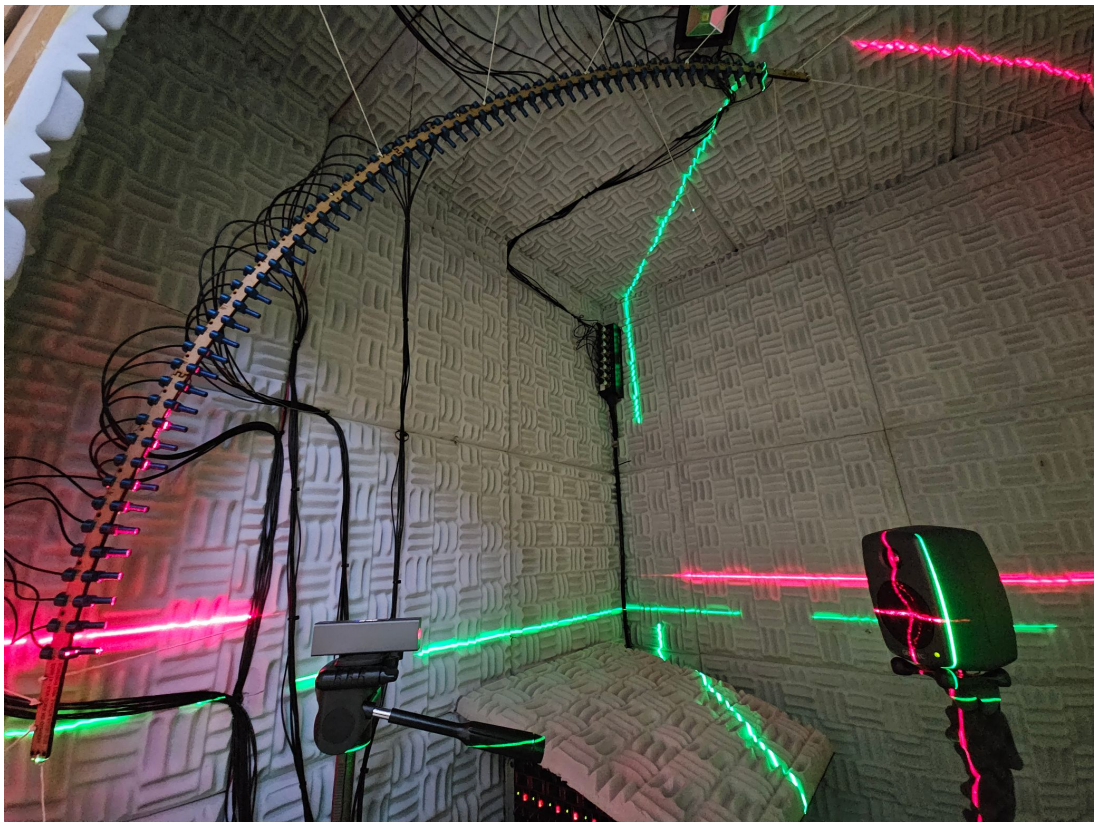
The physical microphone array used in the experimental validation was constructed using Panasonic WM61 electret microphone capsules coupled to dedicated signal-conditioning circuits. This architecture was adopted to obtain a compact and low-cost measurement system while preserving the basic requirements for multichannel acoustic acquisition, namely simultaneous sensing, adequate electrical conditioning of the microphone outputs, and compatibility with subsequent frequency-domain beamforming processing. The resulting array was experimentally assessed in a small anechoic chamber, allowing the measurements to be performed under controlled acoustic conditions and reducing the influence of reflections from the surrounding environment.

Two sound sources were employed during the experimental campaign: an X-Mini mini loudspeaker and a Genelec 8010A active monitor. The electrical excitation signal supplied to the sources covered the frequency range from 180 Hz to 20000 Hz, thereby allowing the array response and the resulting acoustic images to be evaluated over a broad spectral interval. In addition, recordings were performed both with logarithmic sweep signals and with (broadband) flat-noise excitation, so that the proposed deconvolution-based and delay-based methods could be systematically assessed under comparable conditions. The X-Mini was selected due to its compact dimensions, so that its radiation could be used as an experimental approximation of a monopole-like source, particularly in comparison with the simulated point-spread functions (PSFs). The experimental instrumentation is documented in detail in Figures 20 and 21. In total, five measurement conditions were considered, as listed below:

1. X-Mini aligned with the microphone at 0°;
2. X-Mini aligned with the microphone at 45°;
3. X-Mini aligned with the microphone at 90°;
4. Genelec 8010A with the low-frequency driver aligned with the microphone at 0°; and
5. Genelec 8010A with the tweeter aligned with the microphone at 0°.

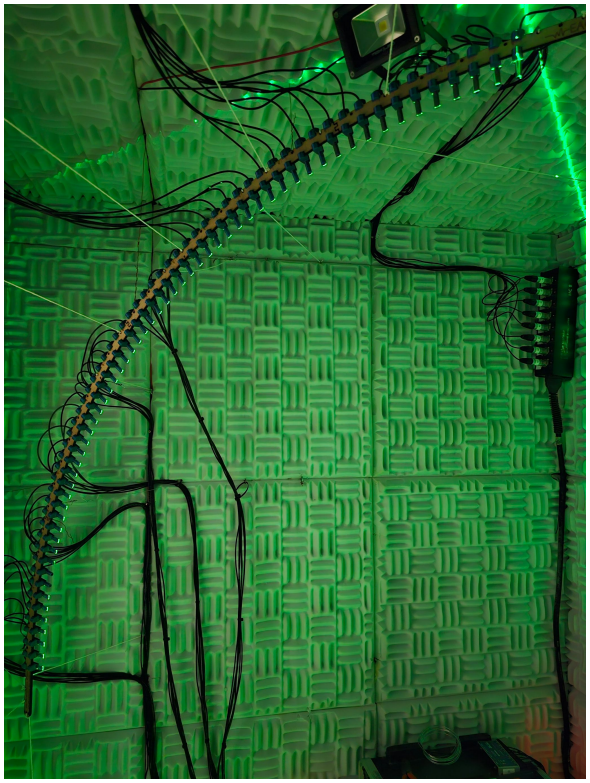
The experimental acoustic maps obtained from these measurements are presented together with the corresponding PSFs. This joint presentation facilitates the interpretation of the results by allowing the measured beamforming maps to be compared directly with the expected spatial response of the array. In this way, source localization, main-lobe shape, sidelobe distribution, and possible imaging artifacts can be more clearly assessed for each experimental condition.

(a) X-Mini measurement aligned at 0º.

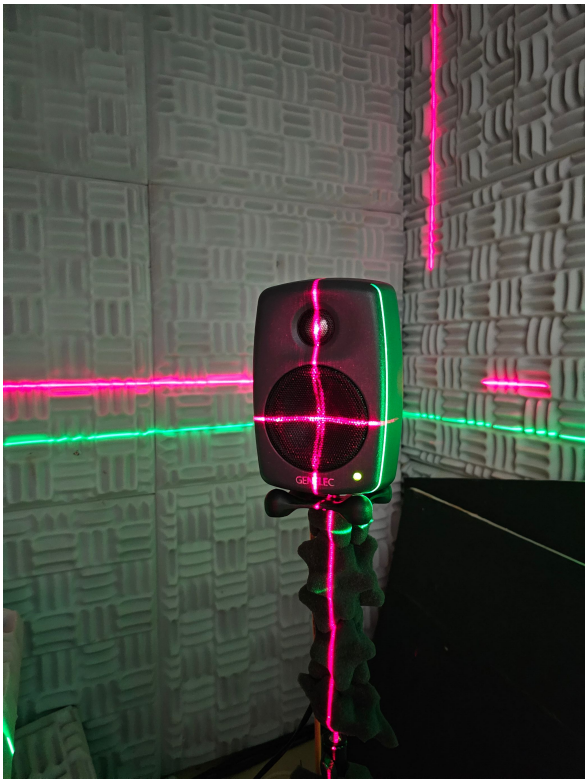


(b) Positioning of the Genelec with driver aligned at 0º.

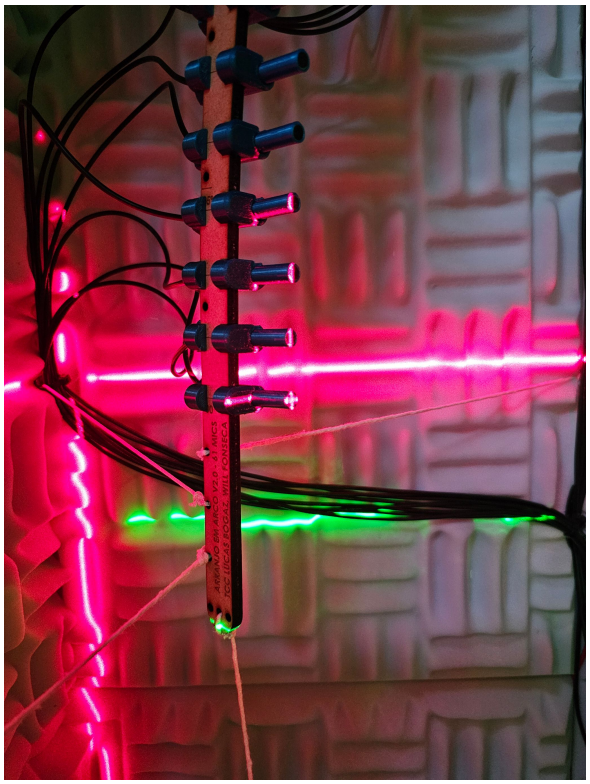
Figure 20: Pictures of the experimental instrumentation (Part 1).



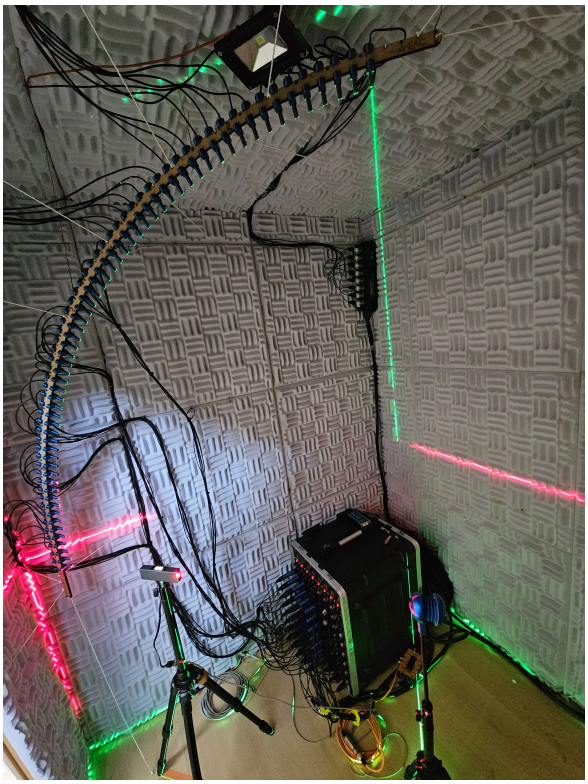
(a) Overall alignment of the arc using a laser.



(b) Front view of the aligned Genelec.

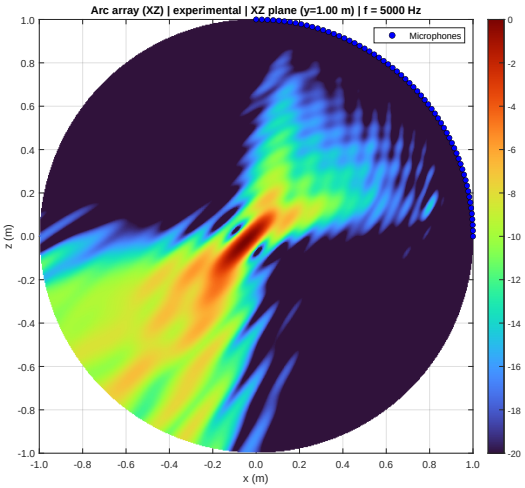


(c) Alignment of the microphone at 0º of the array.

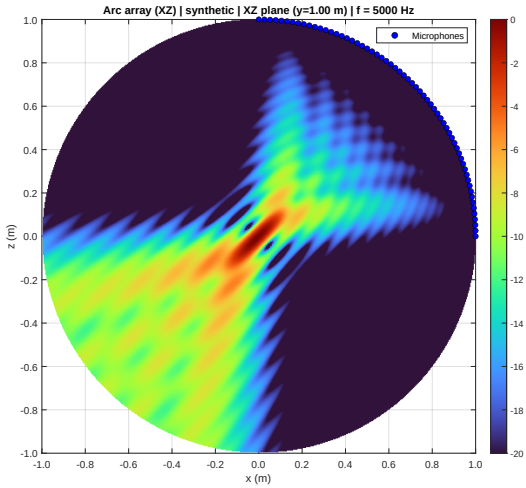


(d) Alignment of the X-Mini at 0º.

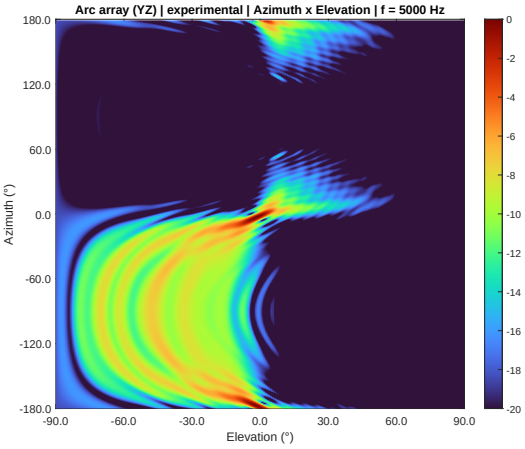
Figure 21: Pictures of the experimental instrumentation (Part 2).



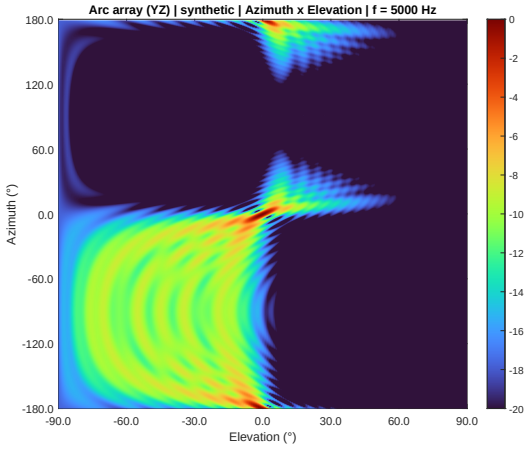
(a) Experimental acoustic map (5 kHz).



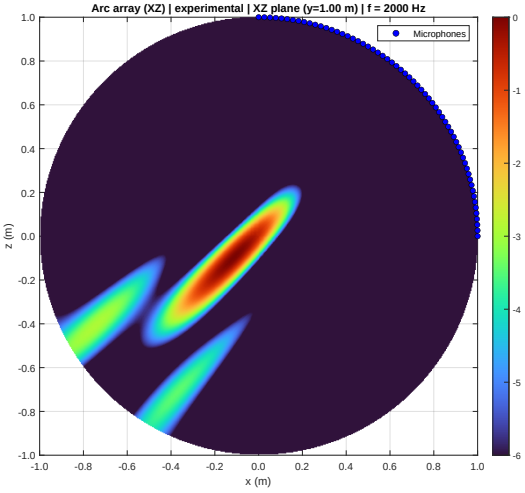
(b) Simulated PSF (5 kHz).



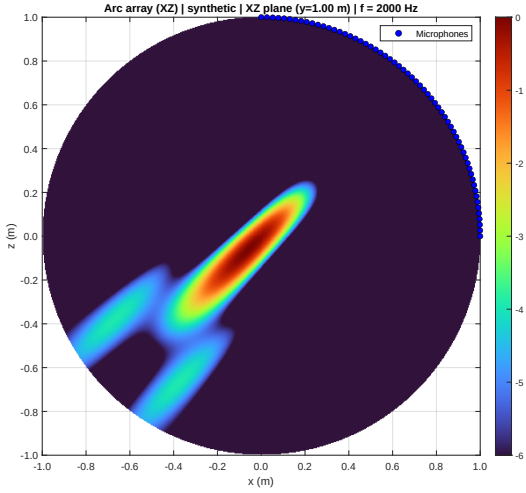
(c) Experimental acoustic map (5 kHz).



(d) Simulated PSF (5 kHz).

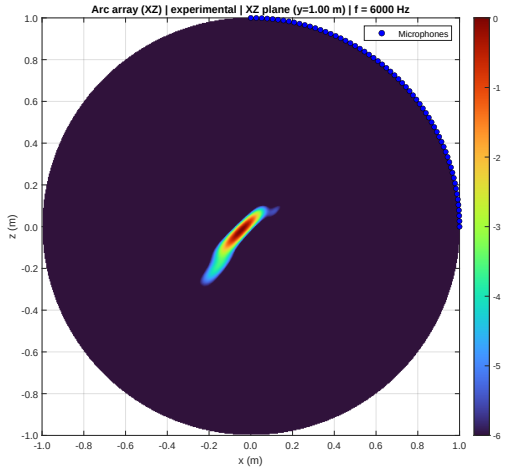


(e) Experimental acoustic map (2 kHz).

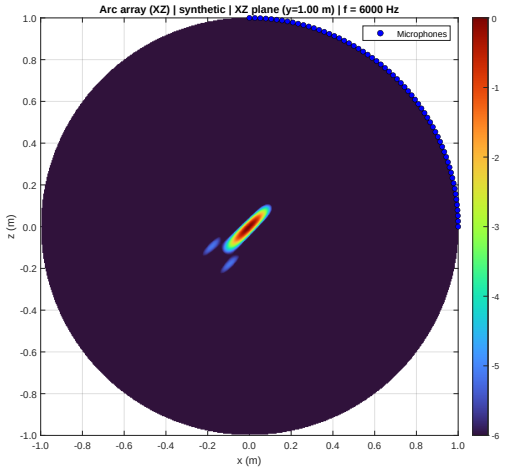


(f) Simulated PSF (2 kHz).

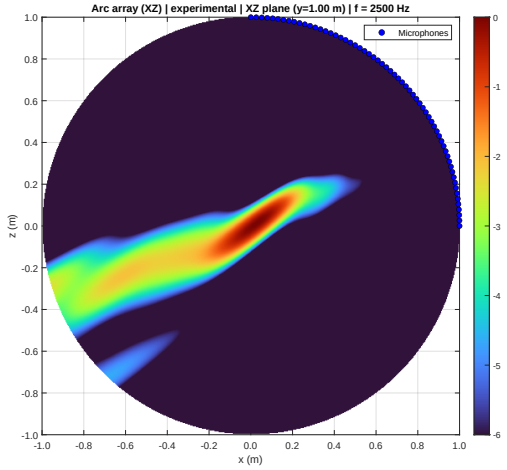
Figure 22: Comparison between the experimental acoustic maps and the simulated point-spread function (X-Mini at 45° *for deconvolution-based method).*



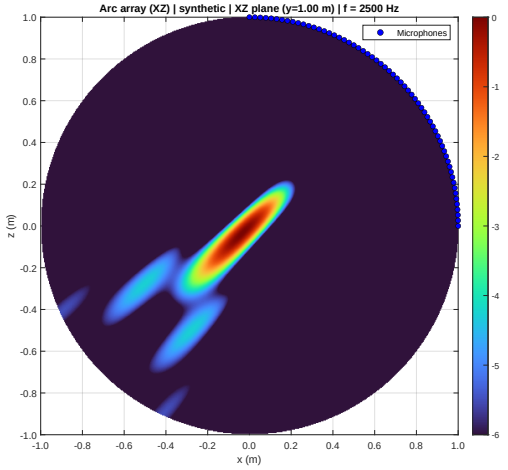
(a) Experimental acoustic map (6 kHz, flat noise, X-Mini at 45°, and delay-based method).



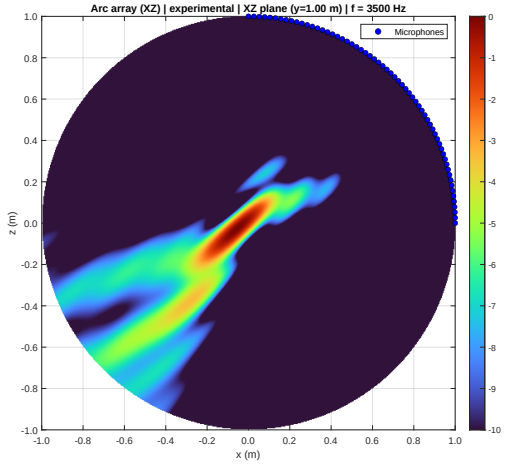
(b) Simulated PSF (6 kHz).



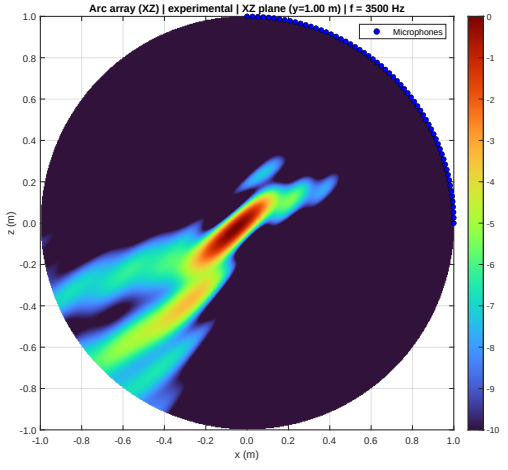
(c) Experimental acoustic map (2.5 kHz, flat noise, Genelec, and delay-based method).



(d) Simulated PSF (2.5 kHz).



(e) Experimental acoustic map (3.5 kHz, flat noise, X-Mini at 45°, and delay-based method).



(f) Experimental acoustic map (3.5 kHz, sweep sine, X-Mini at 45°, and delay-based method).

Figure 23: Comparison between the experimental acoustic maps and the simulated point-spread function (X-Mini and Genelec).

5.3 Results and discussion

The experimentally acquired data were processed using both the deconvolution-based approach and the delay-based procedure, so as to enable a consistent comparison between the two strategies under identical measurement conditions. The resulting acoustic images[8] are presented in Figures 22 and 23, which illustrate the beamforming performance of the arc array for the different source configurations considered in the experimental campaign. These figures allow the spatial resolution, the side-lobe structure, and the overall agreement with the corresponding synthetic PSFs to be visually assessed.

For the delay-based, the time-alignment stage was carried out using solely the two loopback channels recorded during the measurements. For example, in the present case, the first loopback signal exhibited its maximum correlation at sample index 910045, whereas the second loopback signal presented its peak at sample index 909999. The relative delay between the two acquisition blocks is therefore given by

$$\Delta n = n_1 - n_2 = 910045 - 909999 = 46 \text{ samples},$$

which corresponds to a time offset of approximately 0.9583 ms at a sampling frequency of 48 kHz. The compensation of this mismatch was implemented by trimming 46 samples from the beginning of the second microphone block, thereby enforcing a common temporal reference for all channels before the frequency-domain beamforming step.

It is observed that, once the delays estimated from the two loopback signals are properly applied, the deconvolution-based formulation and the delay-based procedure yield identical beamforming maps for the experimental dataset, with no observable differences between the resulting acoustic images. Consequently, the choice between the deconvolution and delay-based implementations may be guided primarily by practical considerations, such as computational cost and ease of integration into existing measurement workflows, rather than by discrepancies in the resulting acoustic images. Moreover, in many practical applications beamforming is deployed as a purely passive "sonar-like" technique, in which an external excitation signal is not available and deconvolution-based schemes cannot readily be applied, further reinforcing the relevance of delay-based processing in real-world scenarios.

6 Final remarks

The present study established a methodological framework for the experimental characterization and deployment of a 64-channel multichannel acquisition system based on commercial audio multi-interface architecture. Beyond the physical assembly of the architecture, its main technical contribution lies in the systematic evaluation of temporal and spectral behavior under different synchronization topologies. For acoustic imaging, such characterization is essential, since spatial reconstruction depends directly on inter-channel coherence, temporal alignment, spectral consistency and phase stability. Even sample-level discrepancies may produce relevant high-frequency phase deviations, which affect the reliability of microphone-array measurements.

The experimental results confirm that professional audio equipment can be employed in demanding multichannel acoustic applications, provided that its limitations are explicitly char-

[8]The post-processing was carried out in Matlab using BEAMAP Toolbox [8].

acterized and compensated. The most critical restriction observed was the occurrence of non-repeatable latencies in configurations relying exclusively on optical clock recovery. Conversely, the external BNC word-clock distribution provided more stable temporal behavior, allowing residual offsets to be treated through loopback testing and post-processing correction. Thus, the feasibility of the proposed system is conditioned not only by the number of available channels, but by the repeatability and controllability of the acquisition chain.

From a broader engineering perspective, the validation of this cost-effective architecture contributes to making experimental acoustic imaging more accessible and reproducible. By combining modular off-the-shelf hardware with systematic synchronization and calibration procedures, the proposed framework enables high-density microphone-array measurements while reducing dependence on dedicated and financially restrictive platforms. This approach does not eliminate the need for rigorous characterization; rather, it shows that scientific reliability can be achieved when the limitations of the instrumentation are incorporated into the measurement methodology.

Future developments may include expanded array geometries, automated time-alignment routines, refined amplitude and phase compensation procedures, and the integration of higher-level beamforming algorithms. Additional validation using different sound sources and more demanding measurement scenarios would further clarify the operational limits of the system. In this perspective, the proposed architecture provides a technically grounded foundation for accessible experimental acoustic imaging, in which cost-effective instrumentation becomes viable through careful synchronization, systematic signal correction, and engineering control.

7 Acknowledgements

The authors would like to thank the support and infrastructure provided by the Acoustical Engineering Program (EAC) [10] and the Graduate Program in Architecture, Urbanism, and Landscaping (PPGAUP), both from the Federal University of Santa Maria (UFSM), Brazil.

References

[1] A. Bailey. *Network Technology for Digital Audio*. Focal Press Music Technology Series. Focal Press, Oxford, 2001. ISBN 978-0240515885.

[2] Behringer. “ADA8200 — Audiophile 8 In / 8 Out ADAT Audio Interface with Midas Mic Preamplifiers.”, 2026. URL https://www.behringer.com/en/products/0800-AAB, accessed: 2026-05-13.

[3] M. Berzborn, R. Bomhardt, J. Klein, J.-G. Richter, and M. Vorländer. “The ITA-Toolbox: An Open Source Matlab Toolbox for Acoustic Measurements and Signal Processing.”, 2017. URL http://publications.rwth-aachen.de/record/687308.

[4] Digital Sound & Music. “5.1.3 Audio Data Streams and Transmission Protocols.”, 2020. URL https://digitalsoundandmusic.com/5-1-3-audio-data-streams-and-transmission-protocols/, accessed: 2026-05-13.

[5] R. P. Dougherty. "Beamforming in Acoustic Testing." In *Aeroacoustic Measurements* (edited by T. J. Mueller), Experimental Fluid Mechanics. Springer, Berlin, Heidelberg, 2002. ISBN 978-3662050583. doi: 10.1007/978-3-662-05058-3_2.

[6] C. Dunn and M. J. Hawksford. "Is the AES/EBU/SPDIF digital audio interface flawed?" *Journal of the Audio Engineering Society*, 40(10), 800–811, 1992. ISSN 0004-7548. URL https://aes.org/publications/elibrary-page/?id=6773.

[7] J. Dunn. "Jitter: Specification and assessment in digital audio equipment." In *Proceedings of the 93rd AES Convention*. San Francisco, CA, USA, 1992. URL https://www.nanophon.com/audio/jitter92.pdf.

[8] W. D'A. Fonseca. "BEAMAP: Beamforming Multiple Analysis Platform." Matlab Toolbox, 2007. URL https://github.com/eac-ufsm/beamap, (computational toolbox for beamforming and acoustic imaging).

[9] W. D'A. Fonseca. *Beamforming Considering Acoustic Diffraction over Cylindrical Surfaces*. PhD Dissertation, Graduate Program in Mechanical Engineering, Federal University of Santa Catarina, Florianópolis, SC, Brazil, 2013. URL https://repositorio.ufsc.br/xmlui/handle/123456789/107608.

[10] W. D'A. Fonseca, E. Brandão, P. H. Mareze, V. S. Melo, R. A. Tenenbaum, C. dos Santos, and D. X. da Paixão. "Acoustical Engineering: a complete academic undergraduate program in Brazil." *The Journal of the Acoustical Society of America*, 152(2), 1180–1191, 2022. ISSN 0001-4966, 1520-8524. doi: 10.1121/10.0013570.

[11] W. D'A. Fonseca, P. H. Mareze, F. R. Mello, and C. C. Fonseca. "Teaching Acoustical Beamforming via Active Learning Approach." In *9th Berlin Beamforming Conference (BeBeC)*, BeBeC-2022-D4. Berlin, Germany, 2022. URL https://www.bebec.eu/fileadmin/bebec/downloads/bebec-2022/papers/BeBeC-2022-D04.pdf.

[12] C. G. Frandsen. "Clock and synchronization in system 6000." Technical note, TC Electronic, 2004. URL https://images.thomann.de/pics/atg/atgdata/document/manual/155826_manual.pdf.

[13] F. M. Gardner. *Phaselock Techniques*. John Wiley & Sons, Hoboken, NJ, 3rd edition, 2005. ISBN 978-0471430636. doi: 10.1002/0471732698.

[14] D. H. Johnson and D. E. Dudgeon. *Array Signal Processing: Concepts and Techniques*. Prentice Hall PTR, 1993. ISBN 978-0130485137.

[15] G. Kashiwa. "Clocking, Jitter and the Digidesign 192 I/O Audio Interface." *Digidesign Technical White Paper*, 2001. URL https://www.yumpu.com/en/document/view/26551038/clocking-jitter-and-the-digidesign-192-i-o-audio-interface-pdf#google_vignette.

[16] R. Merino-Martinez, P. Sijtsma, M. Snellen, T. Ahlefeldt, J. Antoni, C. J. Bahr, D. Blacodon, D. Ernst, A. Finez, S. Funke, T. Geyer, S. Haxter, G. Herold, X. Huang, Q. Leclère, A. Malgoezar, U. Michel, T. Padois, A. Pereira, C. Picard, E. Sarradj, H. Siller, D. Simons, and C. Spehr. "A review of acoustic imaging methods using phased microphone arrays." *CEAS Aeronautical Journal*, 10(1), 197–230, 2019. doi: 10.1007/s13272-019-00383-4.

[17] A. V. Oppenheim and R. W. Schafer. *Discrete-Time Signal Processing*. Pearson, Upper Saddle River, NJ, 3rd edition, 2010. ISBN 978-0131988422.

[18] PreSonus Audio Electronics, Inc. "FireStudio Lightpipe — 32x32 ADAT Lightpipe to FireWire Recording System.", 2024. URL https://www.presonus.com/products/firestudio-lightpipe, accessed: 2026-05-13.

[19] PreSonus Audio Electronics, Inc. "Quantum 2626 — 26x26 Thunderbolt 3 Ultra-Low-Latency Audio Interface.", 2024. URL https://www.presonus.com/products/quantum-2626, accessed: 2026-05-13.

[20] Rack Solutions. "What is "U" height? Rack Units explained.", 2020. URL https://www.racksolutions.com/news/blog/what-is-u-height-rack-units-explained/?srsltid=AfmBOorEhLko1etskOLo5pLcgWDmgANffDrvyO0ZkvuMwGKp0c479oyE, accessed: 2026-05-13.

[21] RME Audio. "Digiface USB — 66-Channel 192 kHz USB Audio Interface.", 2026. URL https://rme-audio.de/digiface-usb.html, accessed: 2026-05-13.

[22] P. Sijtsma. "Experimental techniques for identification and characterisation of noise sources." In *Advances in Aeronautics & Applications – VKI Lecture Series*. Rhode-Saint-Genèse, Belgium, 2004. URL https://reports.nlr.nl/server/api/core/bitstreams/9108ade9-4d4a-4d4a-a979-ddec2ff3079f/content, also in Netherlands Aerospace Centre (NLR) Technical Reports (NLR-TP-2004-165).

[23] P. Sijtsma. "CLEAN Based on Spatial Source Coherence." *International Journal of Aeroacoustics*, 6(4), 357–374, 2007. doi: 10.1260/147547207783359459.

[24] Synthax Audio UK. "What Is ADAT? Synthax Explains.", 2023. URL https://www.synthax.co.uk/latest/2023/06/05/what-is-adat-synthax-explains/, accessed: 2026-05-13.

[25] The MathWorks, Inc. *Matlab*. MathWorks, Natick, Massachusetts, USA, 2024. URL https://www.mathworks.com/products/matlab.html.