



UNIVERSITÀ DI PARMA

ARCHIVIO DELLA RICERCA

University of Parma Research Repository

An Innovative Architecture of Full-digital Microphone Arrays Over A2B Network for Consumer Electronics

This is the peer reviewed version of the following article:

Original

An Innovative Architecture of Full-digital Microphone Arrays Over A2B Network for Consumer Electronics / Pardini, D.; Rocchi, N.; Toscani, A.; Binelli, M.; Chiorboli, G.; Farina, A.; Cattani, L.. - In: IEEE TRANSACTIONS ON CONSUMER ELECTRONICS. - ISSN 0098-3063. - 68:3(2022), pp. 200-208. [10.1109/TCE.2022.3187453]

Availability:

This version is available at: 11381/2927271 since: 2024-10-03T14:44:04Z

Publisher:

IEEE

Published

DOI:10.1109/TCE.2022.3187453

Terms of use:

Anyone can freely access the full text of works made available as "Open Access". Works made available

Publisher copyright

note finali coverpage

(Article begins on next page)

An Innovative Architecture of Full-digital Microphone Arrays Over A²B Network for Consumer Electronics

D. Pinardi, N. Rocchi, A. Toscani, M. Binelli, G. Chiorboli, A. Farina, L. Cattani

Abstract— Microphone arrays of various sizes and shapes are currently employed in consumer electronics devices such as speakerphones, smart TVs, smartphones, and headphones. In this paper, a full-digital, planar microphone array is presented. It makes use of digital Micro Electro-Mechanical Systems (MEMS) microphones, connected through the Automotive Audio Bus (A²B). A clock propagation model for A²B networks, developed in a previous work, was employed to estimate the effects of jitter and delay on microphone arrays. It will be shown that A²B allows for a robust data transmission, while ensuring deterministic latency and channels synchronization, thus overcoming the signal integrity issues which usually affect MEMS capsules. The microphone positioning is also discussed since it greatly affects the spatial accuracy of beamforming. Numerical simulations were performed on four regular geometries to identify the optimal layout in terms of number of capsules and beamforming directivity. An A²B planar array with equilateral triangle geometry and four microphones, three in the vertices and one in the center, was built. Experimental measurements were performed, obtaining an excellent matching with numerical simulations. Finally, the concept of an array of arrays (meta-array) is presented, designed by combining several triangular units and analyzed through numerical simulations.

Index Terms—Automotive Audio Bus (A²B), automotive applications, beamforming, consumer products, digital MEMS microphones, meta-arrays, microphone arrays, planar arrays, spatial audio, triangular arrays.

I. INTRODUCTION

Microphone arrays are employed in consumer electronic devices for several applications, such as teleconferencing, speakerphone units, smartphones, and smart TVs. They are mainly used for dereverberation or echo reduction [1]–[4] and speech recognition [5]. Other consumer electronics applications may include Active Noise Control (ANC) [6], [7], multi-channel audio recording [8] and beamforming [9]. When designing such multichannel systems, the type of capsules and data transmission architecture are critical aspects, as they have a significant influence on the audio quality. The geometry of the array, the number of capsules and their layout, instead, mostly affect spatial accuracy and working frequency range of beamforming. As a result, we see a steadily

growing demand for systems capable of supporting more and more channels. Examples of massive multichannel microphone arrays can be found in [10]–[12].

Regarding the choice of capsule type, most of today’s solutions employ analog microphones, which may provide high-quality audio signals. On the other hand, they entail bulky wiring and noise immunity problems, particularly in case of long wires connecting the capsules with the Analog-to-Digital (A/D) converters. In addition, analog capsules, pre-amplifiers, and A/D converters contribute to increase the cost of the system and design complexity. Conversely, low-cost digital MEMS microphones are more robust to electrical noise, usually at the price of worse acoustic performances (dynamic range and Signal-to-Noise Ratio).

In this paper, a full-digital solution for multichannel arrays is presented, overcoming most of the above-mentioned limitations. Although the usage of MEMS capsules is not new [13]–[15] the adoption of the A²B bus [16], [17] is significantly innovative. In fact, most of the existing solutions employing digital MEMS capsules make use of the I²S or TDM interfaces [18], [19], which do not provide robust transmission of digital signals over distance, thus limiting the length of the cabling to a few centimeters. An A²B network, instead, can transport up to 32 channels through a series of distributed nodes connected in daisy chain up to 40 m. The connection between nodes is made with a single low-cost Unshielded Twisted Pair (UTP) cable, which can also carry power supply together with data. Each node can include transducers such as, but not limited to, microphones and accelerometers. To manage the data flow, a single low-cost A²B transceiver is required on each node, reducing dramatically the cost of the network. At the same time, it offers several advantages for building microphone arrays such as low deterministic latency of just two samples (e.g., less than 50 μ s at 48 kHz) and clock synchronization. Finally, yet important, the A²B bus is expandable: the number of capsules of an array, or in general the complexity of an A²B network, can be increased just by adding additional nodes to the daisy chain. This allows to cover large areas by distributing the nodes, and to build modular arrays, or “meta-arrays”, by combining multiple nodes into a single physical device.

The synchronization of the clock signals is mandatory for

(Corresponding author: N. Rocchi.)

D. Pinardi, N. Rocchi, A. Toscani, M. Binelli, G. Chiorboli, A. Farina are with the Department of Engineering and Architecture of University of Parma, Parma, 43125 Italy (email: daniel.pinardi@unipr.it, nicholas.rocchi@unipr.it, andrea.toscani@unipr.it, giovanni.chiorboli@unipr.it, angelo.farina@unipr.it).

L. Cattani is with ASK Industries S.p.A., Reggio Emilia, 42124 Italy (email: luca.cattani@askgroup.global).

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

ensuring optimal beamforming of transducer arrays. For this reason, the effects of the clock propagation in an A²B network [20] have been studied, evaluating how the latency and the jitter could affect the beamforming of an A²B microphone array. The clock propagation was modelled with uncorrelated normal distributions, whose mean values (μ) and standard deviations (σ) represent deterministic latency and random jitter. The values of μ and σ were assessed by means of experimental measurements. By applying the clock model to numerical simulations of the array, it will be shown that the A²B bus is optimal for building microphone arrays, since the effects of clock propagation and clock reconstruction are negligible in almost every practical application.

The geometry is another critical aspect in the design of a microphone array. In fact, the position of the capsules greatly affects the spatial sampling of the sound field. This has an impact on the robustness and stability of the beamforming filters and on the performance of the array in terms of spatial accuracy. Many solutions have been already explored, mostly spherical and cylindrical [21]–[26], and examples of two-dimensional planar microphone arrays can be found in [27]–[30] with circular, radial or random distribution of the capsules.

In this paper, numerical simulations of planar arrays having regular polygon geometry with different number of vertexes were carried out, with the aim of optimizing the beamforming accuracy and the number of microphones.

Capsules have been positioned on the vertices of triangular, square, pentagonal, and hexagonal geometries, and the theory that a central capsule can improve the spatial accuracy of the beamforming was exploited, as suggested in [31]. It will be shown that the equilateral triangular array with four capsules, three at vertices and one in the center, is the most efficient solution with respect to square, pentagonal, or hexagonal arrays with number of capsules ranging between 4 and 7. Finally, a prototype of triangular array with four capsules over A²B bus was designed and built. Measurements were performed, showing an excellent matching between numerical and experimental results.

Aforementioned geometries were preferred, as they allow exploiting the A²B bus expandability, an innovative feature of the described solution. In the proposed design, several arrays can be attached in daisy-chain and combined to form more complex and more performing structures, the meta-arrays, which are currently under development by the same authors. The paper concludes presenting a preliminary study of a planar meta-array, constituted by seven triangular units, designed accordingly to the previous results, and positioned side by side. The beamforming accuracy of the meta-array was analyzed through numerical simulations also considering the effects of the clock propagation.

II. BEAMFORMING THEORY

The raw signals captured by the capsules of a microphone array can be combined to obtain arbitrary directivity beams, called virtual microphones. This is generally known as encoding or beamforming. In this work, such operation is

performed with a linear processing, which makes use of a matrix of Finite Impulse Response (FIR) filters, computed with the regularized Kirkeby [32] inversion:

$$H_{m,v,k} = [C_{m,d,k}^* \cdot C_{m,d,k} + \beta_k \cdot I_{m,m}]^{-1} \cdot [C_{m,d,k}^* \cdot A_{d,v} \cdot e^{-j\pi k}] \quad (1)$$

where $m = [1, \dots, M]$ is the capsule index, $v = [1, \dots, V]$ is the virtual microphone index, k is the frequency index, $d = [1, \dots, D]$ is the index of the Directions-of-Arrival (DoA) of the sound wave; the matrix C is the complex response of each capsule m for each direction d , the matrix A defines the frequency independent amplitude of the target directivity patterns, $e^{-j\pi k}$ introduces a latency that ensures filters causality, the dot (\cdot) is the scalar product, I is the identity matrix, $[\]^*$ denotes the conjugate transpose, $[\]^{-1}$ denotes the pseudo-inverse, β is a frequency-dependent regularization parameter [33], and it represents the most significant improvement of the Kirkeby method over the traditional Tikhonov regularization, where instead β is constant.

The grid of DoA employed for simulating and measuring the array responses is a spherical t-design geometry [34], [35], of order $t = 21$, consisting in a total of $D = 240$ directions uniformly distributed over a unit-radius sphere.

A unidirectional virtual microphone ($V = 1$) was encoded, centered, and pointed outward, perpendicularly to the surface of the array. The target directivity A is a fourth order cardioid without any side or rear lobes [36] defined as follow:

$$A(\vartheta) = [0.5 + 0.5 \cos(\vartheta)]^4 \quad (2)$$

The virtual microphone is obtained by multiplying, in the frequency domain, the response of the array $C_{m,d,k}$ with the beamforming filter $H_{m,v,k}$, as follow:

$$V_{d,k} = C_{m,d,k} \cdot H_{m,v,k} \quad (3)$$

Ideally, i.e., in case of perfect reconstruction, it would result $V = A$ for all k frequencies and d directions. Two-dimensional and three-dimensional directivity of a fourth-order cardioid virtual microphone are shown in Fig. 1.

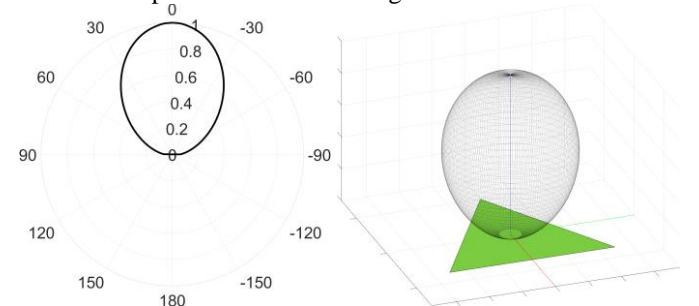


Fig. 1. Ideal polar pattern of a fourth order cardioid, 2D (left) and 3D (right).

The effective directivity is evaluated by employing two parameters as a function of frequency, directivity factor Q and half-power beamwidth BW [37]. The directivity factor Q is given by:

$$Q_{v,k} = \frac{I_{v,k}^{\max}}{I_{v,k}^0} \quad (4)$$

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

where I^{max} is the magnitude of the sound intensity vector in the direction of maximum emission, I^0 is the average of the magnitude of sound intensity over the whole sphere, v is the virtual microphone, and k is the frequency. The parameter BW is equal to twice the angle of the beam between the direction of maximum sensitivity and the direction at -3 dB below the maximum, hence:

$$BW_{v,k} = 2\angle(\vec{S}_{v,k}^{max}, \vec{S}_{v,k}^{max-3dB}) \quad (5)$$

where $\vec{S}_{v,k}^{max}$ is the direction of maximum directivity, $\vec{S}_{v,k}^{max-3dB}$ is the direction where directivity is reduced by 3 dB respect to the maximum, \angle denotes the angle, v is the virtual microphone, and k is the frequency.

III. ARRAY DESIGN AND SIMULATION

Beamforming is physically constrained by the ratio between the distances of the capsules within the array and the wavelength, as:

$$s = \frac{1}{4} \frac{c}{f} \quad (6)$$

where $c = 343$ m/s is the speed of sound and f is frequency. The minimum distance s between the capsules was designed considering the typical frequency range of vocal band, which is usually 300 Hz – 3.4 kHz in telecommunications [38], [39]. Thus, the array would fit voice applications, such as teleconferencing, speech recognition, speakerphone, or ANC systems. The array was designed with a distance between the capsules $s_{min} \approx 25$ mm, obtained by substituting $f_{max} = 3.4$ kHz in (6). In this way, array dimensions fit consumer electronics applications. By exploiting meta-arrays geometries, it is possible to increase the dimension of the array and the number of capsules, hence the maximum distance between the capsules s_{max} , thus improving low frequency performance without affecting high frequency performance.

To investigate the minimum number of capsules satisfying the requirements, four arrays were designed, having shape of equilateral triangle, square, regular pentagon, and regular hexagon, arranging the capsules in the vertexes along a circle, and keeping constant the radius $r = s_{min} = 25$ mm, as shown in Fig. 2. The analyzed layouts feature a central capsule, as suggested in [31].

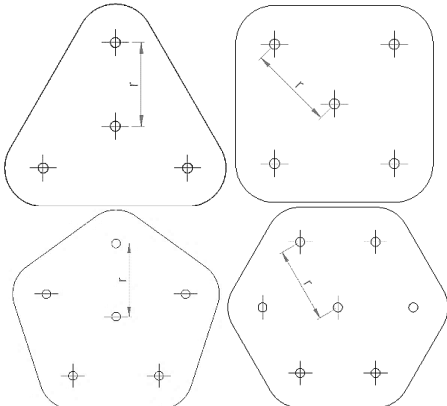


Fig. 2. Design of triangular, square, pentagonal, and hexagonal arrays with capsules arranged with constant radius.

The four geometries were numerically simulated in frequency domain by employing the Finite Elements Method (FEM). The material of the domain is air, while the arrays were modeled as a rigid body. The simulation is performed considering the near field effect: the system is stimulated by a point source radiating spherical waves of 1 Pa at 1 m distance. The simulations were calculated for each direction d of the previously described grid ($D = 240$).

A 3-dimensional modelling was used, thus discretized with a tetrahedral mesh, featuring six elements per wavelength [40]. The simulations were solved within the frequency range 300 Hz – 3.4 kHz, with a frequency resolution of 5 Hz. For each direction d , the solution is evaluated at the M points corresponding to the position of capsules, considering an ideal frequency response. This provided the matrix $C_{m,d,k}$ required to solve (1).

The solutions were processed by combining (2) in (1), for values of M ranging between 4 and 7. Then, the frequency dependent parameters Q (Fig. 3) and BW (Fig. 4) were evaluated for each of the four different cases.

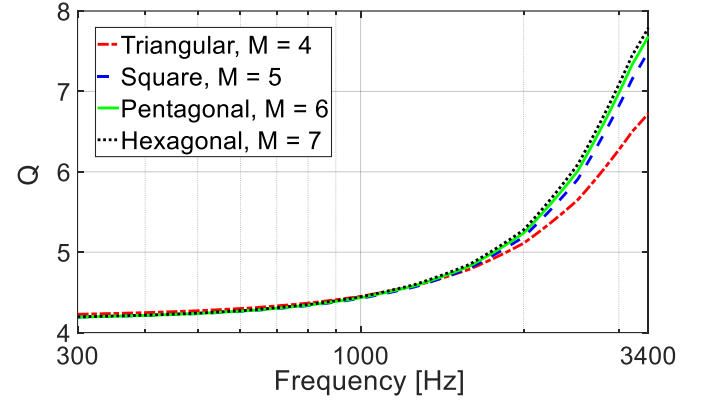


Fig. 3. Directivity factor Q for triangular (red), square (blue), pentagonal (green), and hexagonal (black) arrays, with number of capsules ranging between $M = 4$ and $M = 7$. Higher Q is better.

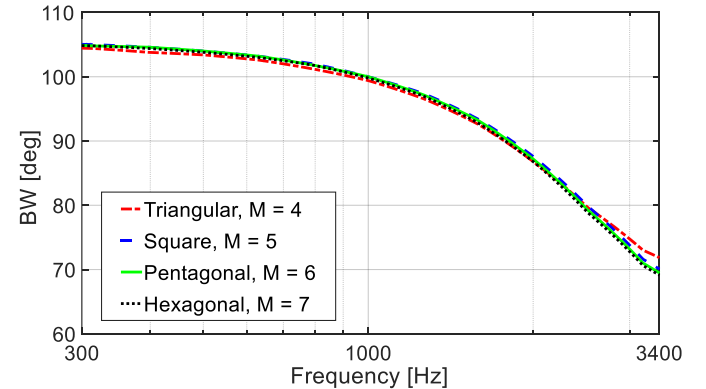


Fig. 4. Half-power beamwidth BW for triangular (red), square (blue), pentagonal (green), and hexagonal (black) arrays, with number of capsules ranging between $M = 4$ and $M = 7$. Lower BW is better.

The values of Q are in the range 4.2 – 7.8, while BW values are comprised between 105° and 69°. It is possible to note that the parameter BW as a function of M has slight variations in the

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

frequency range of interest. The same applies to parameter Q in the range 300 Hz – 2 kHz. Instead, the value of Q increases from 6.7 to 7.8 at 3.4 kHz. This means an increment of 16.4% of Q when using 7 capsules instead of 4, that is a 75% increase of M . The ratio Q/M , which is the directivity factor normalized on the number of capsules, represents the effectiveness of a single capsule and it is maximum when $M = 4$, as shown in Table I.

M	Q/M at 3.4 kHz
4	1.68
5	1.5
6	1.28
7	1.11

In conclusion, the triangular array with four capsules allows maximizing the directivity and minimizing the half-power beamwidth with the minimum number of capsules.

IV. A²B ARCHITECTURE AND CLOCK PROPAGATION MODEL

The block diagram of the architecture of an A²B network is shown in Fig. 5. The network is composed by a single main node and multiple subordinate nodes. Each subordinate node has I/O ports to communicate with local devices, e.g., transducers and codecs. The main node provides the clock to the network, and it is connected to an interface board, which converts A²B signals into protocols commonly used for digital audio distribution, such as USB, AES67 or MADI. A²B allows to transport up to 32 audio channels, as well as control data, on a single network composed of up to 10 subordinate nodes. In addition, subordinate nodes can be power supplied through the same data bus (up to 2.7 W), thus keeping the wiring as simple as possible. Access to the bus and audio flow are managed by dedicated A²B transceivers, removing the need of additional devices that would increase the system cost and design complexity. A simple microcontroller (e.g., an 8-bit microcontroller) is required on the main node to configure the network at the start-up, but it is not needed anymore during normal operation.

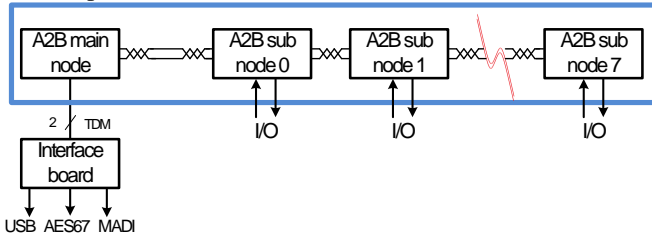


Fig. 5. Architecture of an A²B network.

Since A²B subordinate nodes reconstruct the clock from the network, in this chapter, the effects of clock latency and jitter on the beamforming performance are evaluated by means of an equivalent additive noise [41]–[43]. Fractional delays were applied in frequency domain, overcoming the limited resolution of the discrete time domain, namely ± 1 sample (e.g., 20.83 μ s at 48 kHz). Clock propagation was introduced in the post-processing of numerical simulations by means of (7):

$$C'_{m,d,k} = C_{m,d,k} \cdot e^{-j\omega t_{delay}} \quad (7)$$

where C is the numerical response of the array and $e^{-j\omega t_{delay}}$ accounts for the clock skew contribution. Finally, t_{delay} [s] is:

$$t_{delay} = \mathcal{N}(\mu, \sigma^2, d, m, k) \quad (8)$$

where \mathcal{N} is the standard normal distribution, μ is the mean value of \mathcal{N} , σ^2 is the variance of \mathcal{N} , $d = [1, \dots, D]$ is the DoA index of the sound waves, $m = [1, \dots, M]$ is the capsule index, k is the frequency index. Hence, the clock propagation was modelled with uncorrelated normal distributions, whose mean value (μ) represent a deterministic latency and the standard deviation (σ) represent the effect of a random jitter.

The model was tuned accordingly to experimental measurements specifically performed, whose results are summarized in Table II and represented in Fig. 6. Each A²B chip reconstructs the clock on each subordinate node, therefore all the capsules of an array connected to the same A²B transceiver share the same clock source and are affected by the same amount of jitter and latency. Hence, assigning different latency and jitter values at each capsule of a single node models the worst-case condition. Such condition reflects what happens when building meta-arrays, where several A²B transceivers are employed.

Node	Mic	μ [ns]	σ [ns]
Sub 1	1	0	1.6
Sub 2	2	+10	1.8
Sub 3	3	-10	2.0
Sub 4	4	+5	2.1

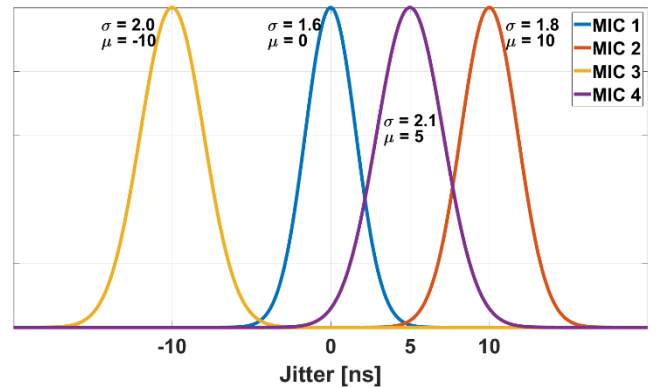


Fig. 6. Clock propagation applied to four capsules connected to four different A²B chips in daisy chain.

The central capsule, assumed as relative reference of the time scale, has $\mu = 0$. The latency of the other capsules is modelled with the worst possible combination. The standard deviation values are measured, and they reflect the tendency of the jitter to increase at each reconstruction of the clock on subsequent subordinate nodes. By processing numerical simulations with and without jitter and latency, it was possible to assess that any effect is completely absent. This result confirms that A²B technology is optimal for building

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

microphone arrays and particularly advantageous when several A²B chips are employed at the same time, which is the case of meta-arrays.

V. EXPERIMENTAL RESULTS

A prototype of the triangular array was built (Fig. 7, microphone capsules are highlighted with red circles). One can note that capsules face on a perfectly flat and smooth PCB surface to minimize diffractions. Therefore, the electronics was designed to have all components and connectors on the rear side. The employed digital MEMS capsule is characterized by an Acoustic Overload Point (AOP) of 130 dB (SPL), SNR of 69 dB(A), dynamic range of 105 dB and an operating voltage range 1.62 – 3.6 V [44].

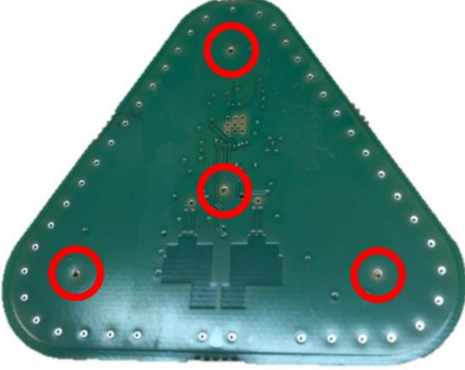


Fig. 7. Prototype of the triangular microphone array with four capsules, three at vertices and one in the center.

The block diagram of the designed triangular array is shown in Fig. 8. The triangular array is seen as a single A²B subordinate node. The UTP cable length connecting the main node with the subordinate node can be up to 15 m. This allows placing the microphone array in the desired position, keeping the acquisition interface away. The acquisition interface comprises the A²B main node and the interface board. In this work, it has been adopted an USB connection between the interface board and the PC.

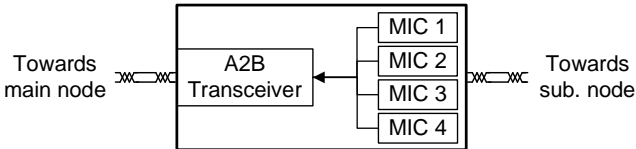


Fig. 8. Block diagram of the designed triangular microphone array.

Such design allows connecting in daisy-chain many arrays to form a larger A²B network, realizing both distributed arrays or the previously mentioned meta-array, made by positioning side-by-side, in two or three dimensions, several array boards.

The prototype was measured in an anechoic chamber with a two-axis turntable [23] and a loudspeaker positioned at 1 m. The schematic of the experimental setup is shown in Fig. 9. The turntable is controlled by the PC via ethernet link. The test signal is sent from the PC to the interface board through USB. This system allows playing the test signal and synchronously recording all the channels of the Device Under Test (DUT), in this case the microphone array. The sound source is a studio

monitor, which is connected to the interface board through S/PDIF. In this way, the whole measurement system is full-digital.

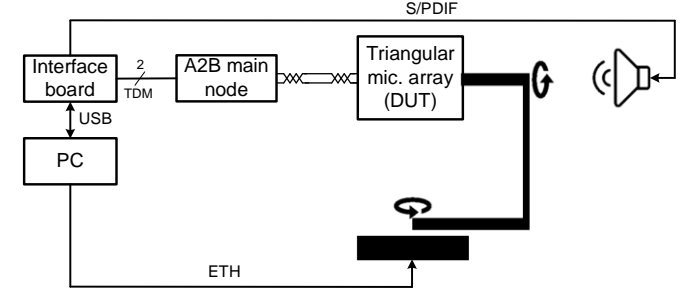


Fig. 9. Experimental measurement setup.

The test signal is an Exponential Sine Sweep (ESS) [45], pre-equalized for flattening the spectrum of the sound source in the range 50 Hz – 18 kHz within ± 0.5 dB. The same test grid employed for simulations was used ($D = 240$ directions). After each measurement, the PC sends to the turntable the new measurement direction and then the ESS is played and recorded. Impulse Responses (IRs) were calculated by convolving the recorded signals with the inverse filter associated to the test signal, namely the inverse sweep [45]. Finally, the $C_{m,d,k}$ matrix of (1) is obtained by applying a Fast Fourier Transform (FFT) to the IRs. Subsequent processing is unchanged with respect to the simulations.

The triangular array was processed and compared to the simulation, by superimposing the directivity polar patterns (Fig. 10).

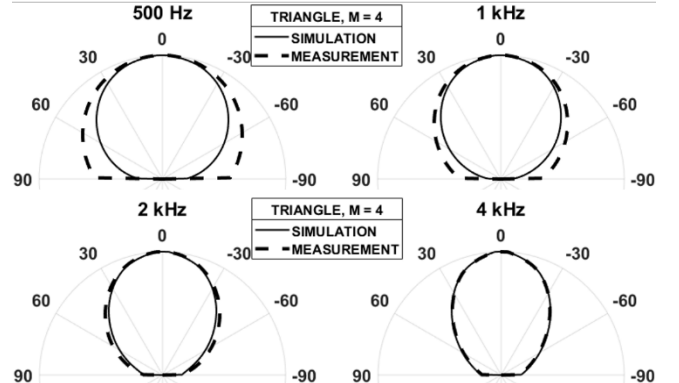


Fig. 10. Polar patterns of a triangular microphone array with four capsules, comparison of numerical (solid line) and experimental (dashed line) solutions.

One can note a very good agreement between the two methods, and particularly at the frequency bands centered at 2 kHz and 4 kHz. Instead, the numerical solution provided slightly narrower polar patterns at 500 Hz and 1 kHz. This behavior is explained by the non-idealities affecting the experimental approach. These non-idealities are geometrical, related to the measurement system, and constructive, related to the MEMS capsules, which are not identical in terms of magnitude and phase response.

VI. META ARRAY SIMULATION

The planar meta-array of Fig. 11 was designed and analyzed accordingly to the outcomes and methodology previously described in Section III. The simulation was performed up to 3.4 kHz but extended down to 20 Hz, to analyze the performance improvement in the low frequency range provided by the increased value of s_{max} . Seven triangular arrays were employed, thus the total number of capsules of the meta-array is $M = 28$ (4 capsules for each triangular array).

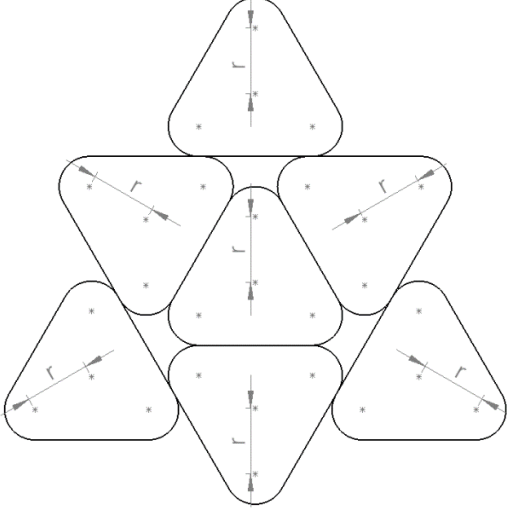


Fig. 11. Design of a meta-array with seven triangular arrays and 28 capsules.

For exploiting the increased beamforming capabilities of such meta-array, the target function required in (1) was modified for producing three virtual microphones of the same type described by (2), uniformly distributed along the azimuth at angles 0° , 120° , and -120° , with an elevation of 45° , as shown in Fig. 12.

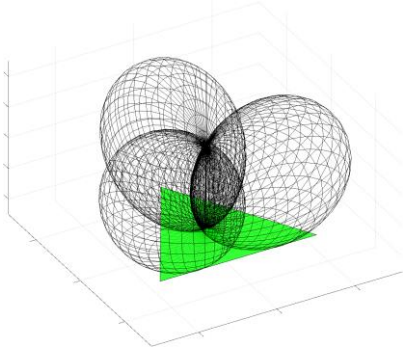


Fig. 12. 3-dimensional ideal polar patterns of three cardioids of fourth order.

Then, the parameters Q (Fig. 13) and BW (Fig. 14) were calculated for the following three cases: only central triangular array ($M = 4$), central triangular plus the three triangular arrays around it ($M = 16$), and all the seven triangular arrays ($M = 28$). Each parameter was averaged among the three virtual microphones, thus reducing the number of curves, and improving the reliability of the result.

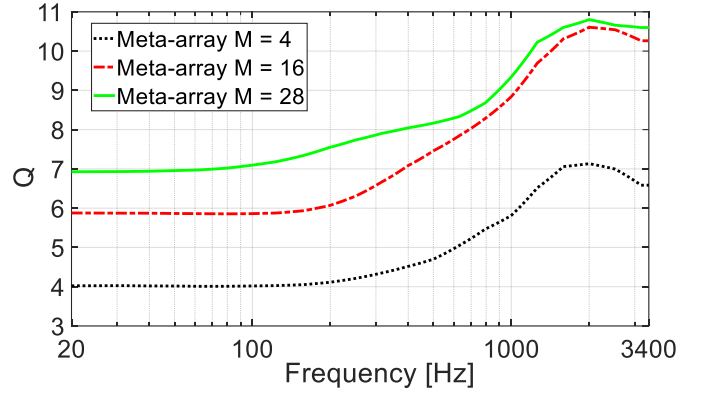


Fig. 13. Directivity factor Q for meta-arrays with one triangular array ($M = 4$, black), four triangular arrays ($M = 16$, red), and seven triangular arrays ($M = 28$, green).

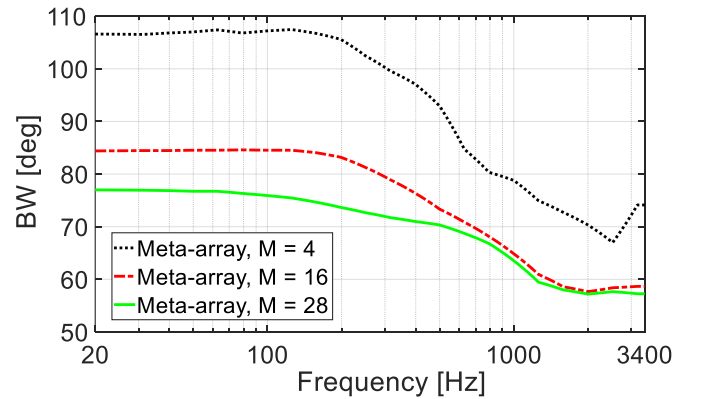


Fig. 14. Half-power beamwidth BW for meta-arrays with one triangular array ($M = 4$, black), four triangular arrays ($M = 16$, red), and seven triangular arrays ($M = 28$, green).

One can note that the directivity factor Q increases in the whole frequency range from $M = 4$ (black line) to $M = 16$ (red line) and a further increase is observed below 700 Hz with $M = 28$ (green line). Conversely, the parameter BW decreases in the whole frequency range between the cases $M = 4$ and $M = 16$ and a further decrement below 700 Hz is observed with $M = 28$. At high frequency the improvement is provided by the increased number of capsules, which in turn gives a better conditioning of the Kirkeby inversion in (1), while at low frequency by the increment of the maximum distance s_{max} between the capsules of the meta-array.

Clock propagation effects were introduced in meta-array simulations, to consider the non-idealities introduced by the data acquisition and transmission architecture. Clock propagation measurements obtained in [20] were employed, as discussed in Section IV. A latency value μ between -10 ns and 10 ns was randomly assigned to each node, with a uniform probability distribution. The standard deviation of random jitter σ was linearly increased by 0.2 ns for each node, starting from the minimum measured value of 1.6 ns. Hence, the four capsules of each node are affected by the same jitter amount, as it happens effectively. It was concluded that clock propagation has no effect on both Q and BW even in large meta-arrays. Fig. 15 shows the results for Q parameter of the meta-array with

7 triangular boards and $M = 28$. It can be seen that the ideal case without jitter (green solid line) is overlapped with the case considering the measured values of jitter (black dashed line). The clock non-idealities must be increased by three orders of magnitude to make deviations observable (red dashed line).

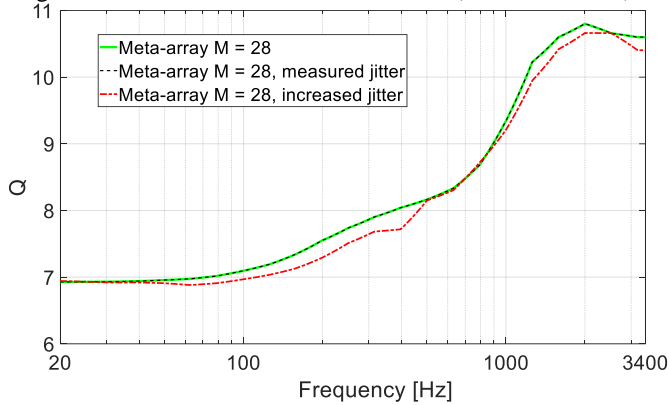


Fig. 15. Directivity factor Q for meta-array with 7 triangular arrays, no jitter (green), jitter measured values (black), and jitter increased by a factor of 1000 (red).

VII. CONCLUSION

An A²B based, full-digital, and modular microphone array was proposed. A²B offers optimal characteristics for automotive and consumer electronics applications such as robust data transmission (up to 40 m), synchronized acquisition, cheap cabling, and expandability thanks to daisy-chain connections. In addition, the design complexity of the proposed microphone array is minimal since data acquisition of digital MEMS microphones and data transmission are carried out by dedicated A²B transceivers. Hence, neither A/D converters or programmable devices (e.g., microcontroller, DSP, or FPGA) are required in case of analog or digital capsules respectively. The system architecture allows creating both distributed microphone arrays and meta-arrays, by positioning side-by-side several array boards in two or three dimensions. In such applications the adoption of A²B, which communicates by means of UTP cables, allows avoiding bulky wiring between capsules, arrays, and acquisition boards.

The effect of clock propagation in A²B bus was studied by modelling latency and jitter. The model is implemented with normal distributions, where mean and standard deviation representing latency and jitter respectively, tuned against experimental measurements. In conclusion, it was possible to assess that the effect of clock propagation can be neglected, confirming that the A²B bus is an optimal solution for building arrays of transducers, even when capsules are connected to different A²B nodes, as it happens for meta-arrays.

Numerical simulations at finite elements were employed for analyzing several planar geometries, with the aim of optimizing the design of the array in terms of position and number of capsules. Only regular polygon geometries were considered, as they are the most suitable to be combined in two or three dimensions to obtain meta-arrays. Triangular, square, regular pentagon, and hexagon arrays with microphones located on the vertices and in the center of the polygons were studied.

Beamforming performance was analyzed by means of two metrics, Q and BW . The triangular array with three capsules in the vertices and one in the center allowed optimizing the beamforming with the minimum number of capsules. As suggested from previous theoretical results, the presence of a central capsule provided an improvement in terms of spatial accuracy, allowing to minimize the deviation between the directivity of the virtual microphone (beam) and the target function (4th order cardioid).

A prototype of the triangular array was built and measured, employing a full-digital test bench. Good match between numerical and experimental results was obtained, as demonstrated by the superimposition of the directivity polar patterns. The proposed architecture is particularly suitable for different consumer electronics applications. In hand-free phones for meeting rooms, depending on room size, delocalized microphones can be easily connected to the main unit thanks to flexible wiring, thus creating meta-arrays. This can be exploited also in smart home or voice assistant devices, where the flexible positioning of the microphones, with respect to the main unit, can enhance acoustic performance. The proposed architecture guarantees positioning flexibility, ease of connection using a simple UTP wiring, and modularity.

Finally, a concept of planar meta-array was presented, obtained by positioning side by side seven triangular arrays. Numerical simulations were performed and processed by including a growing number of units, and therefore by increasing the number of capsules and the maximum distance between them. The improvement observed in the metrics adopted for evaluating the accuracy of beamforming, Q and BW , demonstrated the potential of meta-arrays, whose implementation is made particularly advantageous by the adoption of the A²B architecture. The clock propagation model was applied to meta-array simulations, showing that jitter effects do not worsen beamforming performance.

REFERENCES

- [1] M. Fukui, K. Kobayashi, S. Shimauchi, Y. Hioka, and H. Ohmuro, "Hands-free audio conferencing unit with low-complexity dereverberation," in *2015 IEEE International Conference on Consumer Electronics (ICCE)*, Jan. 2015, pp. 128–129. doi: 10.1109/ICCE.2015.7066349.
- [2] Y. Hioka, M. Okamoto, K. Kobayashi, Y. Haneda, and A. Kataoka, "A display-mounted high-quality stereo microphone array for high-definition videophone system," in *Digest of Technical Papers - IEEE International Conference on Consumer Electronics*, 2008, no. 6, pp. 6–7. doi: 10.1109/ICCE.2008.4587858.
- [3] J. Hong, "Stereophonic Acoustic Echo Suppression for Speech Interfaces for Intelligent TV Applications," *IEEE Transactions on Consumer Electronics*, vol. 64, no. 2, pp. 153–161, May 2018, doi: 10.1109/TCE.2018.2843293.
- [4] J. Hong, S. Jeong, and M. Hahn, "Wiener filter-based echo suppression and beamforming for intelligent TV interface," *IEEE Transactions on Consumer Electronics*, vol. 59, no. 4, pp. 825–832, Nov. 2013, doi: 10.1109/TCE.2013.6689695.
- [5] T. Kawase, M. Okamoto, T. Fukutomi, and Y. Takahashi, "Speech Enhancement Parameter Adjustment to Maximize Accuracy of Automatic Speech Recognition," *IEEE Transactions on Consumer Electronics*, vol. 66, no. 2, pp. 125–133, May 2020, doi: 10.1109/TCE.2020.2986003.
- [6] C.-Y. Chang, A. Siswanto, C.-Y. Ho, T.-K. Yeh, Y.-R. Chen, and S. M. Kuo, "Listening in a Noisy Environment: Integration of active noise

> REPLACE THIS LINE WITH YOUR MANUSCRIPT ID NUMBER (DOUBLE-CLICK HERE TO EDIT) <

- control in audio products,” *IEEE Consumer Electronics Magazine*, vol. 5, no. 4, pp. 34–43, Oct. 2016, doi: 10.1109/MCE.2016.2590159.
- [7] V. Patel, J. Cheer, and S. Fontana, “Design and Implementation of an Active Noise Control Headphone With Directional Hear-Through Capability,” *IEEE Transactions on Consumer Electronics*, vol. 66, no. 1, pp. 32–40, Feb. 2020, doi: 10.1109/TCE.2019.2956634.
- [8] S. M. Kim, C. J. Chun, and H. K. Kim, “Multi-channel audio recording based on superdirective beamforming for portable multimedia recording devices,” *IEEE Transactions on Consumer Electronics*, vol. 60, no. 3, pp. 429–435, Aug. 2014, doi: 10.1109/TCE.2014.6937327.
- [9] N. D. Gaubitch, J. Martinez, W. B. Kleijn, and R. Heusdens, “On near-field beamforming with smartphone-based ad-hoc microphone arrays,” in *2014 14th International Workshop on Acoustic Signal Enhancement (IWAENC)*, Sep. 2014, pp. 94–98. doi: 10.1109/IWAENC.2014.6953345.
- [10] R. Kerstens, D. Laurijssen, and J. Steckel, “Low-cost one-bit MEMS microphone arrays for in-air acoustic imaging using FPGA’s,” in *2017 IEEE SENSORS*, Oct. 2017, pp. 1–3. doi: 10.1109/ICASSP.2017.8234087.
- [11] M. Turqueti, J. Saniie, and E. Oruklu, “Scalable acoustic imaging platform using MEMS array,” in *2010 IEEE International Conference on Electro/Information Technology*, May 2010, pp. 1–4. doi: 10.1109/EIT.2010.5612131.
- [12] H. A. Sanchez-Hevia, R. Gil-Pita, and M. Rosa-Zurera, “FPGA-based real-time acoustic camera using pdm mems microphones with a custom demodulation filter,” in *2014 IEEE 8th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Jun. 2014, pp. 181–184. doi: 10.1109/SAM.2014.6882370.
- [13] A. Farina and S. Fontana, “Microphone Assembly Having a Reconfigurable Geometry,” US Patent 0379969 A1, 2019
- [14] B. Li and T. Ju, “An Improved Sound Source Location Method for MEMS Microphone Array,” in *2019 IEEE 19th International Conference on Communication Technology (ICCT)*, Oct. 2019, pp. 469–472. doi: 10.1109/ICCT46805.2019.8947027.
- [15] S. Chowdhury, M. Ahmadi, and W. C. Miller, “Design of a MEMS acoustical beamforming sensor microarray,” *IEEE Sensors Journal*, vol. 2, no. 6, pp. 617–627, Dec. 2002, doi: 10.1109/JSEN.2002.807773.
- [16] AnalogDevices, “A2B Audio Bus: an Easier, Simpler Solution for Audio Design.” <https://www.analog.com/en/applications/technology/a2b-audio-bus.html>
- [17] M. Kessler, “Introducing the Automotive Audio Bus (A2B),” in *2017 AES International Conference on Automotive Audio*, Sep. 2017, pp. 1–5.
- [18] L. Li, K. Lian, J. Fu, P. Zhu, Z. Hu, and C. Guo, “Acoustic Enhanced Camera Tracking System Based on Small-Aperture MEMS Microphone Array,” *IEEE Access*, vol. 8, pp. 215827–215839, 2020, doi: 10.1109/ACCESS.2020.3041445.
- [19] E. Zwysig, F. Faubel, S. Renals, and M. Lincoln, “Recognition of overlapping speech using digital MEMS microphone arrays,” in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 7068–7072. doi: 10.1109/ICASSP.2013.6639033.
- [20] N. Rocchi, A. Toscani, G. Chiorboli, D. Pinardi, M. Binelli, and A. Farina, “Transducer Arrays Over A²B Networks in Industrial and Automotive Applications: Clock Propagation Measurements,” *IEEE Access*, vol. 9, pp. 118232–118241, 2021, doi: 10.1109/ACCESS.2021.3106710.
- [21] J. S. Abel *et al.*, “A configurable microphone array with acoustically transparent omnidirectional elements,” in *127th Audio Engineering Society Convention 2009*, 2009, pp. 3–12.
- [22] D. Ayllón, V. Benito-Olivares, C. Llerena-Aguilar, R. Gil-Pita, and M. Rosa-Zurera, “Three-dimensional microphone array for speech enhancement in hands-free systems for cars,” in *Proceedings of the AES International Conference*, 2012, pp. 1–7.
- [23] A. Farina, S. Campanini, L. Chiesi, A. Amendola, and L. Ebri, “Spatial sound recording with dense microphone arrays,” in *AES 55th International Conference*, 2014, pp. 1–8.
- [24] R. González, J. Pearce, and T. Lokki, “Modular design for spherical microphone arrays,” in *Proceedings of the AES International Conference*, 2018, pp. 1–7.
- [25] J. Meyer and G. Elko, “A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield,” in *IEEE International Conference on Acoustics Speech and Signal Processing*, May 2002, vol. 2, pp. II-1781–II-1784. doi: 10.1109/ICASSP.2002.5744968.
- [26] S. Sakamoto, S. Hongo, T. Okamoto, Y. Iwaya, and Y. Suzuki, “Sound-space recording and binaural presentation system based on a 252-channel microphone array,” *Acoustical Science and Technology*, vol. 36, no. 6, pp. 516–526, 2015, doi: 10.1250/ast.36.516.
- [27] R. Schultz-Amling, F. Kuech, M. Kallinger, G. del Galdo, J. Ahonen, and V. Pulkki, “Planar microphone array processing for the analysis and reproduction of spatial audio using Directional Audio Coding,” in *124th Audio Engineering Society Convention 2008*, 2008, pp. 1–10.
- [28] J. Lopez-Ballester, M. Cobos, J. J. Perez-Solano, G. Moreno, and J. Segura, “General purpose modular microphone array for spatial audio acquisition,” in *138th Audio Engineering Society Convention*, 2015, pp. 1–6.
- [29] I. Andras, P. Dolinsky, L. Michaeli, and J. Saliga, “Beamforming with small diameter microphone array,” in *2018 28th International Conference Radioelektronika (RADIOELEKTRONIKA)*, Apr. 2018, pp. 1–5. doi: 10.1109/RADIOELEK.2018.8376368.
- [30] Z. I. Skordilis, A. Tsiami, P. Maragos, G. Potamianos, L. Spelgatti, and R. Sannino, “Multichannel speech enhancement using MEMS microphones,” in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2015, pp. 2729–2733. doi: 10.1109/ICASSP.2015.7178467.
- [31] B. Rafaely, “The Spherical-Shell Microphone Array,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 4, pp. 740–747, May 2008, doi: 10.1109/TASL.2008.920059.
- [32] O. Kirkeby, F. Orduna, P. A. Nelson, and H. Hamada, “Inverse Filtering in Sound Reproduction,” *Measurement and Control*, vol. 26, no. 9, pp. 261–266, 1993.
- [33] H. Tokuno, O. Kirkeby, P. A. Nelson, and H. Hamada, “Inverse Filter of Sound Reproduction Systems Using Regularization,” *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E80-A, no. 5, pp. 809–820, 1997.
- [34] R. H. Hardin and N. J. A. Sloane, “McLaren’s improved snub cube and other new spherical designs in three dimensions,” *Discrete And Computational Geometry*, vol. 15, no. 4, pp. 429–441, Apr. 1996, doi: 10.1007/BF02711518.
- [35] D. Pinardi, “Spherical t-Design for Characterizing the Spatial Response of Microphone Arrays,” in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, Sep. 2021, pp. 1–8. doi: 10.1109/I3DA48870.2021.9610850.
- [36] A. Farina, A. Amendola, L. Chiesi, A. Capra, and S. Campanini, “Spatial Pem Sampling ;,” in *52nd international conference of Audio Engineering Society*, 2013, pp. 1–12.
- [37] D. Pinardi and A. Farina, “Metrics for Evaluating the Spatial Accuracy of Microphone Arrays,” in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, Sep. 2021, pp. 1–9. doi: 10.1109/I3DA48870.2021.9610887.
- [38] J. Ahonen, V. Pulkki, and T. Lokki, “Teleconference Application and B-format Microphone Array for Directional Audio Coding,” in *Proceedings of the AES 30th International Conference*, Mar. 2007, pp. 1–10.
- [39] Qiyue Zou, Xiaoxin Zou, Ming Zhang, and Zhiping Lin, “A Robust Speech Detection Algorithm in a Microphone Array Teleconferencing System,” in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*, 2001, vol. 5, pp. 3025–3028. doi: 10.1109/ICASSP.2001.940295.
- [40] “COMSOL, Acoustics Module User’s Guide.”
- [41] C. Travis and P. Lesso, “Convention Paper 6293 Specifying the Jitter Performance,” in *117th AES Convention*, 2004, pp. 1–15.
- [42] “IEEE Standard for Terminology and Test Methods for Analog-to-Digital Converters,” *IEEE Std 1241-2010 (Revision of IEEE Std 1241-2000)*, no. January, pp. 1–139, 2011. [Online]. Available: <http://ieeexplore.ieee.org/servlet/opac?punumber=5692954>
- [43] “IEEE Standard for Terminology and Test Methods of Digital-to-Analog Converter Devices,” *IEEE Std 1658-2021*, vol. 2012, pp. 1–126, Feb. 2011. doi: 10.1109/IEEESTD.2012.6152113.
- [44] InfineonTechnologies, “Im69D130,” 2017. https://www.infineon.com/dgdl/Infineon-IM69D130-DS-v01_00-EN.pdf?fileId=5546d462602a9dc801607a0e46511a2e
- [45] A. Farina, “Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique,” in *108th Convention*, 2000, pp. 1–24.



D. Pinardi received the M.S. degree (cum laude) in mechanical engineering from the University of Parma, Italy, in July 2016, with a thesis on loudspeaker modeling, and the Ph.D. degree in industrial engineering from the University of Parma, in March 2020, with a thesis on the design of microphone, hydrophone, and camera arrays for spatial audio recording.

He is a Research Assistant of Prof. Angelo Farina at University of Parma since 2016, mainly specialized in spatial audio, design of transducer arrays, acoustics simulations and 3D auralization, applied to automotive field, and underwater acoustics.

Dr. Pinardi is member of the *Audio Engineering Society* from 2017.



N. Rocchi received the B.S. and M.S. degrees (cum laude) in electronic engineering from the University of Parma, Italy, in 2017 and 2019, respectively.

He is currently pursuing the Ph.D. degree in information technology at University of Parma. His research interests include multichannel audio distribution, acquisition systems, and audio amplifiers.



A. Toscani received the M.S. degree (cum laude) in electronic engineering and the Ph.D. degree in information technology from the University of Parma, Italy, in 2004 and 2008, respectively. His research activity is mainly focused on power electronics, high-performance electric drives, diagnostic techniques for industrial electric systems, and power converter for audio application.

Since 2004, he has been working with the Department of Information Engineering (now, the Department of Engineering and Architecture), University of Parma, where he is currently a Research Fellow.

Dr. Toscani is author of two patents.



M. Binelli received the B.S. and M.S. degrees in electronic engineering from the University of Parma, Italy, in 2003 and 2006, respectively, and the Ph.D. degree from the University of Parma, in 2010, with a thesis on analysis and equalization techniques in automotive acoustics. His research interests include equalization,

microphones and loudspeakers arrays, spatial audio psychoacoustics and active noise control.

Since 2010, he has been working as Research Fellow at University of Parma.



G. Chiorboli received the M.S. degree (cum laude) in electronic engineering from the University of Bologna, Bologna, Italy, in 1987. His current research interests include electronic instruments and sensors, analog-to-digital and digital-to-analog modelling and testing, and electrical characterization of semiconductor devices.

He is currently an Associate Professor of electronic measurements with the University of Parma, Parma, Italy.



A. Farina received the M.S. degree in civil engineering from the University of Bologna, Italy, in December 1982, with a thesis on the acoustics and vibrations inside a tractor cab, and the Ph.D. degree in technical physics from the University of Bologna, in 1987, with a thesis on experimental assessment of concert hall acoustics. His research activity includes

several fields of applied acoustics, including noise and vibration, concert hall acoustics, simulation software, and advanced measurement systems. In the last ten years, he focused mostly on applications involving massive microphone and loudspeaker arrays.

He has been a full-time Researcher with the University of Bologna, since November 1986, and the University of Parma since March 1992. In November 1998, he became an Associate Professor with the University of Parma, where he has been a Full Professor of environmental applied physics since May 2005. He has the Chair of applied acoustics and technical physics with the University of Parma. He is the author of more than 300 scientific articles and three widely employed software packages: Ramsete, Aurora Plugins, and DISIA.

Prof. Farina was awarded with the AES Fellowship for his pioneering work on electroacoustic measurements based on exponential sine sweeps.



L. Cattani received the Ph.D. degree in Information Technologies from the University of Parma (Italy) in 2016. His research interests include Advanced Signal Processing (Audio and Radio Frequency) for automotive applications.

He is currently a Project Leader in the R&D department of Ask Industries SpA.