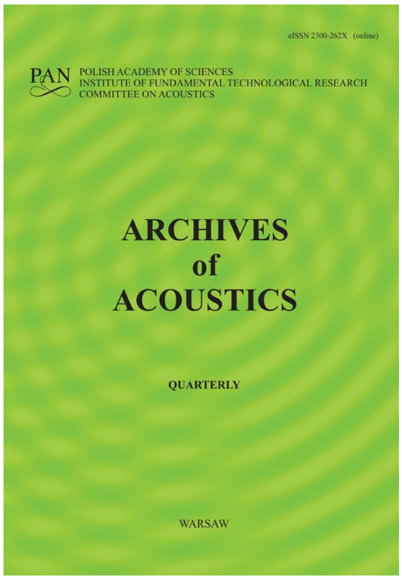
JOURNAL PRE-PROOF



This is an early version of the article, published prior to copyediting, typesetting, and editorial correction. The manuscript has been accepted for publication and is now available online to ensure early dissemination, author visibility, and citation tracking prior to the formal issue publication.

It has not undergone final language verification, formatting, or technical editing by the journal's editorial team. Content is subject to change in the final Version of Record.

To differentiate this version, it is marked as "PRE-PROOF PUBLICATION" and should be cited with the provided DOI. A visible watermark on each page indicates its preliminary status.

The final version will appear in a regular issue of *Archives of Acoustics*, with final metadata, layout, and pagination.

Title: Comparative Perceptual Assessment of Sound Quality: Cone vs. Distributed Mode Loudspeakers

Author(s): Piotr Kleczkowski, Paweł Małecki, Dorota Młynarczyk

DOI: https://doi.org/10.24423/archacoust.2026.4222

Journal: *Archives of Acoustics*

ISSN: 0137-5075, e-ISSN: 2300-262X

Publication status: In press

Received: 2025-03-25

Revised: 2025-11-26

Accepted: 2026-01-12

Published pre-proof: 2026-01-19

Please cite this article as:

Kleczkowski P., Małecki P., Młynarczyk D. (2026), Comparative Perceptual Assessment of Sound Quality: Cone vs. Distributed Mode Loudspeakers, *Archives of Acoustics*, https://doi.org/10.24423/archacoust.2026.4222

Comparative Perceptual Assessment of Sound Quality: Cone vs. Distributed Mode Loudspeakers

Piotr Kleczkowski, Paweł Małecki, Dorota Młynarczyk

AGH University of Krakow, al. A. Mickiewicza 30, 30-059 Krakow, Poland

Abstract

A conventional cone loudspeaker has a limited capacity for creating the impression of spatiality, while a distributed mode loudspeaker (DML) has an inherent ability to evoke it. DMLs have their specific drawbacks, but some of these can be compensated for. A key question arises – is it a cone loudspeaker or a compensated DML that is preferred by listeners? A listening experiment with carefully controlled conditions was carried out to answer this question; 30 subjects participated. The participants evaluated three stereo systems based on a DML speaker (with its power response equalized) and two conventional two-way active systems. Two perceptual attributes were evaluated: “overall preference,” and “spatial impression.” A graded pairwise comparison was used as an experimental paradigm; the results were analyzed according to the law of comparative judgment. The findings indicated that, even though the DMLs achieved slightly lower ratings than the conventional systems on average, the perceptual differences were very small. This was confirmed by the hypothesis testing that was performed on the raw results of the pairwise comparisons.

Keywords: distributed mode loudspeakers, loudspeaker evaluation, spatial sound, pairwise comparison, listening experiment

Comparative Perceptual Assessment of Sound Quality: Cone vs. Distributed Mode Loudspeakers

Introduction

The relationships among the directivity of loudspeakers, the acoustics of home listening rooms, and perception have been investigated by a number of researchers: Evans, Dyreby, Bech, Zielinski, and Rumsey (2009), Toole (1986a), Toole (1986b), Toole (2018), Bertland (1985), Olive (2004c), Olive (2004a) and Zacharov (1998). Key findings can be summarized as follows: a joint indicator that combines both the directivity of a loudspeaker and the acoustic properties of a room is the ratio of the direct to the early-reflected sounds. This indicator is closely related to the perception of reproduced sounds, with its higher values (more-direct sound) favoring the accurate localizations of sound sources and its lower values favoring the perception of space. Widening the radiation of the loudspeaker and increasing the reflectivity of the room boundaries both reduce the ratio of the direct to the early-reflected sounds. Reducing the lateral reflections in a listening room tends to have the same effect as narrowing the loudspeaker dispersion (Moulton, 1986). There are several works that support the opinion that the wide directivity of loudspeakers is preferable when compared to the directivity of conventional loudspeakers (Allison, 1995; Bertland, 1985; Ferralli & Moulton, 1995; Flindell, McKenzie, Negishi, Jewitt, & Ward, 1991; Linkwitz, 2007; Moulton, 1986). Extensive work on the subject was carried out by Toole (1986a, 1986b, 2018), who noticed that wide dispersion loudspeakers were preferred by listeners—especially for recreational listening (but not exclusively). The preference for more- or less-dry listening conditions depends on the purpose of the listening; audio engineers favor drier spaces for their work, while more reverberance is preferred for recreational listening.

Loudspeaker systems that are based on electrodynamic units with cone-shaped diaphragms are incapable of wide radiation—even when they employ dome-shaped tweeter units. There is an unconventional type of loudspeaker that offers very wide radiation as its inherent property: the distributed mode loudspeaker (DML), which belongs to a wider class of flat panel loudspeakers.

A DML employs a different sound-radiating element than a conventional dynamic one does. In the DML, the radiator is a stiff flat panel with a rectangular shape and considerable mass. An electrodynamic or piezoelectric exciter is attached to the panel and induces uniformly distributed bending wave vibrations. This is entirely different than with a dynamic speaker, which was designed to vibrate like a rigid piston. In what follows, a loudspeaker system that is based on dynamic units will be referred to as a piston loudspeaker system (PLS).

DML technology began to attract the interest of researchers at the turn of this century, and numerous works on the subject have been published. An introduction to the technology can be found in (Angus, 2000; G. Bank & Harris, 1998; N. J. Harris & Hawksword, 2000; Newell & Holland, 2019), and a review of its history was written by M. C. Heilemann, Anderson, Roessner, and Bocko (2021).

DMLs are normally mounted in walls or ceilings (i.e., they become architectural loudspeakers [this mounting is also referred to as flush mounting]), which is one of their advantages. With this mounting, their directivity can be described as quasi-omnidirectional in the hemisphere. Comprehensive anechoic measurements were presented in (Bai & Huang, 2001; Czesak & Kleczkowski, 2023); the directivity characteristics were irregular, but the property of omnidirectionality was maintained.

DMLs have another unique property besides quasi omnidirectionality: they behave like many sound sources; thus, their radiation is incoherent (Azima & Harris, 1997; Gontcharov & Hill, 2000; N. Harris, Gontcharov, & Hawksford, 2000). Its advantage is that the interference of the direct sound with the first reflections is largely suppressed; thus, the comb-filtering effect is reduced, and the advantage of the reflections (i.e., the perception of space) is maintained. The effect of using an incoherent sound source is similar to applying acoustic diffusors. Wendt and Höldrich (2021) analyzed the consequences of specular and diffuse reflections on the precedence effect.

As a consequence of their principle of operation (which consists of the excitation of modal frequencies), DMLs have irregular frequency responses that are far from being flat and bring coloration. This drawback could be the main reason why DMLs have never been considered to be high-end devices. Nevertheless, their development is ongoing (Bai & Huang, 2001; M. C. Heilemann et al., 2021; Jeon, Ryu, Kim, & Wang, 2020; Jung, Jensen, Jeong, Jeon, & Wang, 2021; Lu & Shen, 2009; Lu, Shen, & Liu, 2012; Yu, Zhu, Wu, & Yang, 2023; Zenker, Schurmann, Merchel, & Altinsoy, 2020).

It is likely that improvements to DML technology, new areas of application (like the screens in OLED TVs), and the current trend in the home-entertainment market toward the use of architectural loudspeakers with multi-speaker and multi-room installations will bring another wave of interest in DMLs. Therefore, a comparison of the perceptual qualities of DMLs and PLSs is a timely topic.

There have been a few works that have been published on the perceptual properties of DMLs – especially as compared to PLSs. N. Harris, Flanagan, and Hawksford (1998) found that DMLs improved the stereo localization of pink noise stimuli when compared to PLSs in an untreated room. Flanagan and Moore (2000) showed that detecting a spectral ripple was easier from a DML than it was from a PLS, but the accuracy of the vowel identification was similar for the two loudspeakers (Flanagan & Moore, 2001). Flanagan and Harris (1999) proposed a hypothesis that the loudness attenuation with distance in a given space was reduced by the use of a DML.

M. Heilemann, Anderson, Roessner, and Bocko (2018) performed anechoic measurements of three different types of DMLs (one- and multi-exciter) and a two-way PLS. They used a prediction model of loudspeaker preferences (Olive, 2004b) and obtained an objective evaluation that strongly favored the PLS. Roessner, Heilemann, and Bocko (2019) performed a listening comparison among two multi-exciter DML prototypes, one one-exciter commercial DML, and two two-way passive PLSs, in monophonic reproduction. The two PLSs obtained the highest scores, while the two prototype DMLs scored about 10% lower. The one-exciter DML scored distinctively lower than the others.

Newell and Holland (2019) reported an experiment where conventional loudspeakers were used as a stereo pair and four DMLs were used to reproduce separately recorded ambience signals (with a remarkably realistic effect).

The purpose of this work was to evaluate the perceptual properties of an example DML by a subjective perceptual comparison with two examples of professional quality active PLSs in a stereo listening format. A key assumption for the experiment was to reduce the main disadvantage of DMLs (i.e., an irregular frequency response), as it was likely that it could override the acoustic advantages of DML technology. The smoothing of the frequency response was achieved by equalization (which is an easily accessible technological option). Although the DML technology has found wide range of applications, no thorough perceptual experiment on DMLs with carefully controlled conditions satisfying the requirements for an objective comparison, participated by a relatively large panel of listeners, and other than a monophonic format was known to the authors at the time of writing this article.

Method

Assumptions for Experiment

1. It is widely agreed that the flatness of the amplitude response is the key factor in the preference ratings of loudspeakers (Gabrielsson, Lindström, & Till, 1991; Olive, 2004c). Therefore, it was decided to implement equalization in order to avoid the overriding of the results by the inherent irregularity of the frequency responses of DMLs. The possible generalization of results required a simple and robust method of equalization. Magnitude-only equalization was chosen.

2. The most widespread listening format was assumed, i.e., stereophonic.

3. It was decided to use pairwise comparison as an experimental method due to its sensitivity to small perceptual differences (ITU-R, 2015). To keep the durations of the listening sessions within recommended limits (Bech & Zacharov, 2006), only three loudspeaker systems could be used; this resulted in the decision to compare three pairs.

4. The pilot listening evaluation of an equalized DML speaker indicated that the overall sound quality was comparable to a PLS of good quality. Therefore, three systems were elected for comparison: one DML, and two near-field two-way active monitors

as examples of high-end PLSs. As an example of DML, a commercial unit from the line of the most widely available full DML-type panels from Amina Technologies Ltd. was selected – a one-exciter Edge 5 model. Two PLSs were chosen: a pair of Dynaudio BM15s, and a pair of Genelec 8030s. To minimize any commercialism, the DML will be referred to as "DML," the Dynaudio PLS as "PLS_D," and the Genelec PLS as "PLS_G" throughout the rest of this paper.

5. To implement the pairwise comparison, fast switching between sound sources was necessary. Therefore, all units had to be permanently installed during each experimental session.

Figure 1. Exemplary arrangement of loudspeakers. During the experiment, loudspeakers were hidden behind an acoustically transparent curtain.

Positions of Loudspeakers

An important advantage of DMLs is the ease of their flush mounting, which provides a number of acoustic advantages (Newell & Holland, 2019). However, it was decided not to use this option since it would limit the objectivity of the comparison. The flush mounting of all of the compared loudspeakers was not feasible.

The loudspeakers of both stereo channels were positioned as close to each other as possible (as can be seen in Fig. 1), but this still introduced experimental biases – from the different spacings between the loudspeaker units in the stereo pairs and the slightly different positions of the loudspeakers in the room (thus, exciting different room modes and reflections). In order to reduce these biases, the spatial arrangements of all of the loudspeakers were changed from session to session so that the participants listened to different arrangements.

The geometric centers of the speaker triplets were positioned at a height of 120 cm (corresponding to the average ear level of the seated participants). These centers formed a stereo triangle with the listener's head. The distance from the listener to the center of the triplet was precisely measured, and the distances of the loudspeakers from the walls were also controlled. The stereo base was 2.5 meters (measured from the center of the arrangement). The distance from the walls was no less than one meter.

There were $n = 3! = 6$ possible speaker arrangements, and the arrangements of both channels were made to be mirror images in order to maintain symmetry (cf. Fig. 5). With 12 experimental sessions, each arrangement was repeated twice in random order. A carefully designed protocol for exchanging the positions of the speakers between sessions was implemented (as shown in Fig. 2).

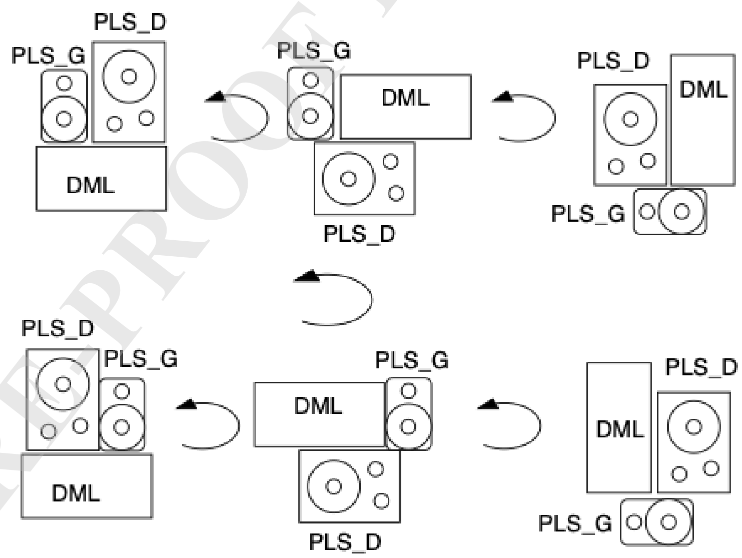


Figure 2. Changing positions of speakers between experiment sessions.

Equalization of Frequency Responses

Equalization of Frequency Responses of DML Loudspeakers Efforts toward loudspeaker equalization may be grouped into two types: equalization of the loudspeaker itself (as measured under anechoic conditions), and loudspeaker + room equalization (based on the loudspeaker's in-room frequency response). In principle, loudspeaker + room equalization provides more of a flat-amplitude response at the point of listening; however, room equalization is sometimes questioned from the point of view of psychoacoustics. This also introduces an experiment-specific factor, thus limiting the external validity of an experiment. Therefore, we chose loudspeaker-only equalization.

The frequency responses of a DML measured from different directions largely differ (Czesak & Kleczkowski, 2023). Therefore, a routine measurement at the axis that is perpendicular to the loudspeaker surface is inappropriate for DMLs; multipoint measurements with averaging should be used instead.

There have been numerous works on loudspeaker equalization (e.g. Karjalainen, Piirila, Järvinen, and Huopaniemi (1999); Norcross, Soulodre, and Lavoie (2004) and B. Bank (2013)), but fewer attempts have been published on equalizing DMLs. Pueo, López, Ramos, and Escolano (2009) and M. C. Heilemann, Anderson, and Bocko (2017) studied the equalization of multiactuator DMLs, and their measurements were limited to one point (as is typical for piston loudspeakers). Ho and Berkhoff (2015) investigated a new honeycomb structure of a DML panel with multiple actuators and applied velocity feedback controllers for each actuator. Hörchens and de Vries (2011) compared measurement methods for equalizing DMLs and concluded that this could not be based on measurements taken at a single position (or only a few positions) in front of the panel but rather on its average radiation spectrum.

Equalization Above 100 Hz In this work, it was chosen to implement the magnitude equalization of the frequency response. The procedure presented below was limited to the range above 100 Hz, since the efficiency of the DML used drops off rapidly below this frequency and the equalization would require excessive power.

A precise multipoint measurement was performed in order to obtain a reliable average magnitude of the frequency response (Czesak & Kleczkowski, 2023). Each of the DMLs used in the left and right stereo channels was analyzed. The frontal hemisphere radiation was investigated, as the units were supplied in an enclosure that reduced backward radiation. The measurements were performed on a dense grid of 325 points over the hemisphere, with an angular resolution of 10° (both in azimuth and elevation).The grid of measurement points is presented in Figure 3. There was one measurement point at the axis perpendicular to the loudspeaker surface, and 9 × 36 points were distributed along nine circles (representing the parallels of the hemisphere).

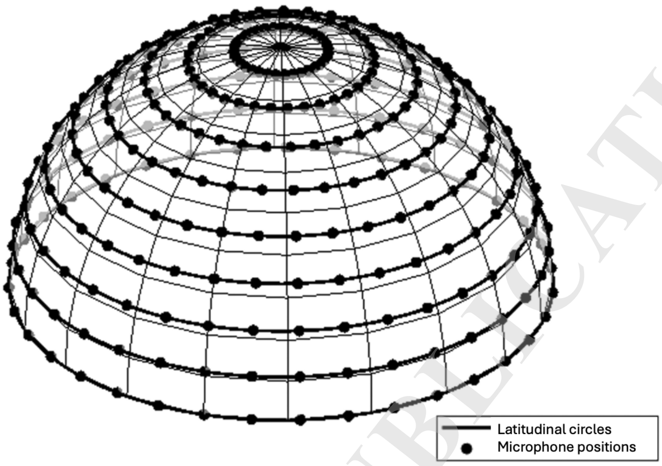


Figure 3. The grid of measurement points of the DML on the hemisphere (from (Czesak, 2025))

The measurements were carried out in a 1000 m^3 anechoic chamber at the Department of Mechanics and Vibroacoustics, AGH University of Krakow, using a custom-made automated system for positioning a measurement microphone. To minimize near-field effects, the maximum available measurement radius was used (2.5 m).

Due to constant angular resolution, geometric correction was required: points at higher elevation angles (e.g., 80°) are denser than those at 0° and represent smaller areas. Correction coefficients derived in Czesak, Kleczkowski, and Król-Nowak (2022) are presented in Table 1.

Table 1

Relative correction factors for parallel circles in measurement hemisphere

Elevation angle [°]	0	10	20	30	40	50	60	70	80	90
Correction factor	1.000	0.985	0.940	0.866	0.766	0.643	0.500	0.342	0.174	0.786

Averaging acoustic pressure over the 325 points of the measurement hemisphere was performed according to Equation 1:

$$\bar{p}_h(f) = \sqrt{\sum_{i=0}^{8} k_i \sum_{j=0}^{35} |p_{i,j}(f)|^2 + k_9|p_{9,0}(f)|^2} \quad (1)$$

where:

- $\bar{p}_h(f)$ – average acoustic pressure over the hemisphere,
- k_i – correction factors for elevations from Table 1,
- $p_{i,j}(f)$ – acoustic pressure at point (i, j),
- f – frequency.

At each measurement point, amplitude-frequency responses were obtained using narrow-band noise excitation in 1/20th-octave bands (193 bands, center frequencies: 63–16,200 Hz). Custom software developed at the anechoic chamber (Pilch & Kamisiński, 2011) provided sound pressure values in dB for each band, offering constant-Q resolution across the frequency range.

An equalizing filter was implemented using a 220-point FFT filter. The FFT filter, with linear-phase property and numerical stability, allowed accurate equalization and easy generalization. Since offline filtering was used, causality was not a requirement, and the zero-phase property was advantageous for perceptual comparisons. Informal listening revealed no pre-ringing effects.

The equalization procedure was as follows:

1. Conversion of 1/20-octave frequency scale (63–16,200 Hz) to a linear scale of 220 points;
2. Linear interpolation to obtain $A(f)$ – the linear-frequency amplitude response;
3. Computation of $A_c(f) = 1/A(f)$ – the equalizing filter;
4. Offline filtering of each musical excerpt by $A_c(f)$ using a 220-point zero-phase FFT filter;

5. Conversion back to time domain by IFFT and storage as 24-bit/44.1 kHz WAV files. It was decided to requantize the floating-point results of filtering not to 16 bit but to 24 bit with appropriate dithering, as D/A conversion during reproduction was of 24 bit resolution.

Figure 4 shows the equalizing filters for both DMLs. The right channel was equalized from 100 Hz, while the left channel from 85 Hz due to better low-end response. The filters are highly similar, indicating consistency between the DML units. Flat sections below 100 Hz and above 16 kHz were manually limited to avoid overload.

To preserve objectivity, musical excerpts used for comparing DML with the two PLS systems were also passed through the same 220-point FFT/IFFT pipeline. Above 100 Hz, this procedure was transparent.

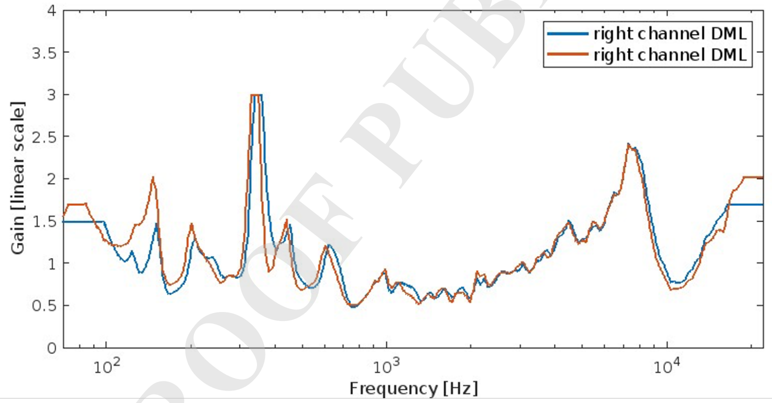


Figure 4. Frequency responses of equalizing filters for DML loudspeakers.

Correction at Frequencies Below 100 Hz As the frequency responses of the DMLs could not be corrected below 100 Hz, an alternative approach to equalization in this range was adopted. The low-frequency responses were essentially flat down to 40 Hz in the case of PLS_D and to 54 Hz in the case of PLS_G (both at –2 dB). The responses of the PLSs were corrected to match those of the DMLs using high-pass filtering of the test material. This filtering was implemented within the FFT/IFFT procedure described earlier.

Analysis of the spatially averaged frequency responses of the DMLs revealed that their downward slope below 100 Hz (right speaker) and 85 Hz (left speaker) was approximately

–18 dB/oct. Therefore, this slope was applied in the high-pass filtering of material reproduced by both PLS systems.

Listeners

The experiment involved a total of 30 participants, primarily fourth-year students from the Acoustical Engineering program, with an average age of 23 years. The distribution of participants across experimental sessions was as follows: 7 attended 1 session, 8 attended 2 sessions, 7 attended 3 sessions, 6 attended 4 sessions, and 2 attended 5 sessions. We decided to divide the entire panel of listeners into two groups. The first consisted of those who participated in one session, as well as those who participated in two sessions, but in the latter case only the results from the first session were included – these participants were classified as one-time listeners (1-TL). The second group consisted of those who participated in at least three sessions. For those who took part in more than three sessions, only the first three sessions were included – these were classified as three-times listeners (3-TL). The advantage of this division was that the two groups were mutually exclusive, with the 3-TL group containing, on average, somewhat more experienced listeners. Both groups included an equal number of participants (15 each).Most of the participants had similar moderate levels of experience due to their completions of courses in ear training and sound engineering (although some individuals had additional experience working in the industry or through other relevant experiences). However, they had little experience with listening tests. About half of the members of the 1-TL group were not students of acoustical engineering and had no experience at all. Because participation was voluntary, it may be hypothesized that more experienced listeners were more interested in taking part and thus were more likely to appear in the 3-TL group. None of the participants reported any hearing problems.

Setup

During the experiments, the loudspeakers were concealed behind an acoustically transparent curtain, rendering them invisible to the listeners. The experimental setup is illustrated in Fig. 5. The experiment was conducted in a room with a floor area of 44 m^2

and a ceiling height of 2.8 m, yielding a volume of 123 m^3. The room was acoustically treated with appropriate panels and materials, resulting in a reverberation time of approximately 0.3 seconds. Based on its volume and reverberation characteristics, the room conformed to the standards for listening rooms as specified in ITU-R BS.1116-2 (ITU-R, 2015) and EBU Tech. 3276 (EBU Tech, 1998). However, the background noise level in the room, influenced by a neighboring street, was higher than recommended by these standards—approximately 35 dB SPL(A), comparable to that of an empty office space. Nonetheless, the test signal level was approximately 80 dB SPL(A), ensuring a sufficient signal-to-noise ratio for perceptual testing. While the room did not fully comply with the noise level specification, the elevated test level mitigated any potential perceptual masking effects.

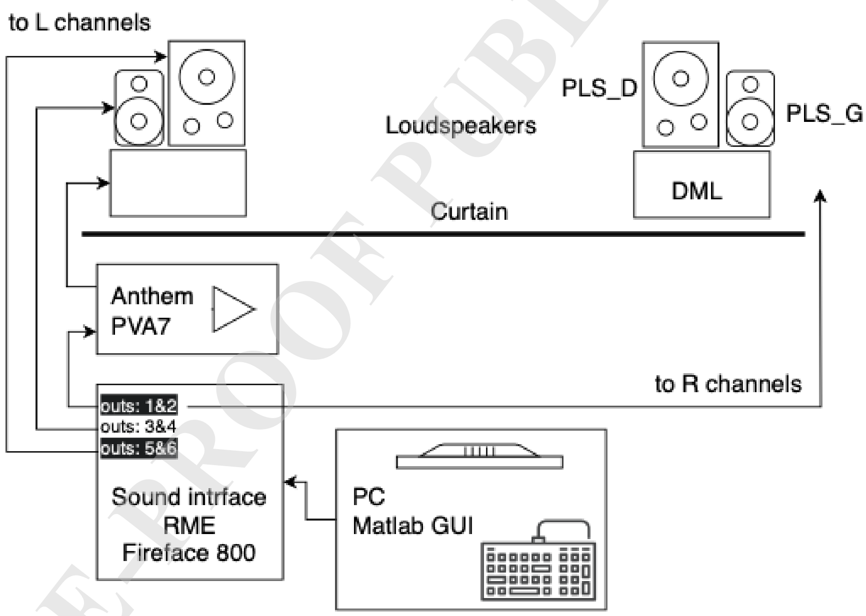


Figure 5. Experimental setup.

Evaluation Attributes

The evaluation process was facilitated through a graphical user interface (GUI) developed in MATLAB (see Fig. 6). The interface enabled participants to assess two perceptual attributes: *overall preference* and *spatial impression*. The test was conducted in Polish, and the respective original terms were: *preferencja ogólna* and *wrażenie przestrzenności.*

Both attributes were evaluated in the same session. The position (upper/lower) of each attribute in the graphical interface (Fig. 6) was not randomized, but users were free to evaluate attributes in the order of their choice. These attributes were rated using a continuous slider scale ranging from *A much better* through *Very similar quality* to *B much better*. This design allowed for nuanced judgments and captured subtle perceptual differences between loudspeaker systems.

The experiment followed a double-blind protocol in which neither participants nor administrators knew which speaker system was assigned to version A or B during any given trial. The GUI assigned control numbers to each evaluation, tracked trial progression, and maintained systematic consistency across sessions.

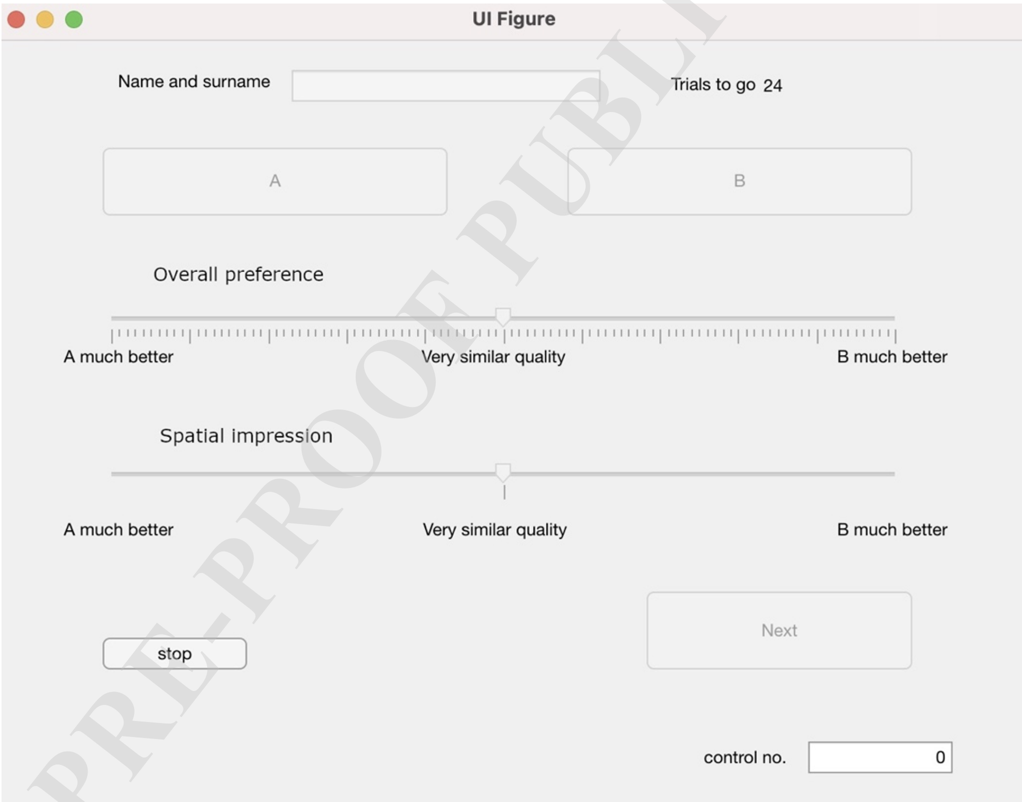


Figure 6. Graphical user interface used in the experiment.

The *overall preference* attribute referred to the overall auditory experience delivered by the loudspeaker system. The *spatial impression* attribute addressed the spatial impression perceived by the listener.

The ITU-R (2019) standard offers a general framework for sound quality evaluation, including aspects such as clarity, spatiality, and timbre. It extends the methodologies outlined in ITU-R (2015), which is focused on the subjective assessment of small impairments in audio systems, including multichannel configurations. Additionally, the EBU Tech (1998) recommendation provides detailed guidance on listening conditions and the assessment of sound program material for mono and stereo reproduction, with emphasis on spatial attributes and perceptual clarity.

In our experiment, the focus was narrowed to two key evaluation dimensions: *overall preference* and *spatial impression*. This decision was informed by informal pretests, in which listeners struggled to distinguish timbral and clarity differences after equalization of the DMLs. Therefore, such aspects were assumed to be implicitly accounted for in the "general quality" rating. This also helped reduce the cognitive load for participants with limited critical listening experience.

The most pronounced audible differences stemmed from the contrasting directivity patterns of the loudspeakers. These were considered to be captured in the *spatial impression* attribute. The incoherent radiation of the DMLs, their longer impulse responses (Anderson & Bocko, 2015), and the use of non-coaxial configurations likely contributed to broader perceptual differences, all of which were assumed to be reflected within the *overall preference* ratings.

Method for graded pairwise sound-quality comparison

Two most often used experimental paradigms in research on perception are independent rating (referred to as multiple comparisons in audio evaluation) and pairwise comparison. Multiple comparisons are more prevalent in audio evaluation, as can be inferred from the published literature. According to (Perez-Ortiz et al., 2019), pairwise comparison eliminates observer bias. We chose the latter paradigm, with an extension referred to as graded pairwise comparison (or scaled pairwise comparison) (Koczkodaj, 2016; Perez-Ortiz et al., 2019). Below we outline the basics of pairwise comparison, while information on how we used the graded version is included at the end of this section.

The raw results of the pairwise comparisons had to be transformed into a set of scalar parameters assigned to each of the three compared stimuli, allowing for direct ranking. This is typically achieved using statistical models, most notably Thurstone's law of comparative judgment (Thurstone, 1927a, 1927b; Tsukida & Gupta, 2011) or the Bradley-Terry model (later extended to the Bradley-Terry-Luce model) (Tsukida & Gupta, 2011).

In this study, Thurstone's model was employed. According to Thurstone, each comparison between two stimuli Ψ_1 and Ψ_2 evokes a "discriminal process" in the listener, assumed to be normally distributed. Consequently, the perceived difference between two stimuli is also normally distributed, and its mean reflects the most frequent judgments made by the listeners. If Ψ_1 is perceived as stronger or better than Ψ_2, the probability area under the $\Psi_1 - \Psi_2$ distribution is greater than that under the reverse comparison.

Thus, the perceptual magnitude of difference between two stimuli can be described by the following expression (Thurstone, 1927b):

$$\bar{\Psi}_1 - \bar{\Psi}_2 = z_{12} \cdot \sigma_{\Psi_1 - \Psi_2} \tag{2}$$

where:

- $\bar{\Psi}_1, \bar{\Psi}_2$ – means of the normally distributed discriminal processes,
- z_{12} – z-score corresponding to the observed probability $p_{\Psi_1 > \Psi_2}$,
- $\sigma_{\Psi_1 - \Psi_2}$ – standard deviation of the distribution of $\Psi_1 - \Psi_2$.

Thurstone's original formulation was based on dichotomous judgments (i.e., A > B or B > A), which yielded raw scores in the form of probability values (p), later transformed into z-scores. As the model does not provide values for $\sigma_{\Psi_1 - \Psi_2}$, several assumptions were required. Thurstone proposed five model cases, depending on the assumptions concerning standard deviations. The model used in this study is a hybrid of Case I (single observer with repeated judgments) and Case II (multiple observers, single judgment per pair), because the design involved multiple observers each providing repeated judgments.

In graded pairwise comparison listeners express the magnitudes of their preferences toward A or B with some numeric scale. With a graded comparison, the distribution of the values of $\Psi_1 - \Psi_2$ is readily available, and the p values are calculated from the sums of the respective scores (unlike in a dichotomous judgement). We assumed that the distribution of $\delta_{\Psi_1-\Psi_2}$ calculated this way provided an estimate of $\delta_{\Psi_1-\Psi_2}$ from Thurstone's dichotomous model. This assumption allowed direct computation of $\delta_{\Psi_1-\Psi_2}$ without making any assumptions about the underlying distributions. Thus, direct estimation of $\overline{\Psi_1} - \overline{\Psi_2}$ was possible. Formally, the model that was used in this work was a combination of Thurstone Case I (no simplifying assumptions about the distributions of the data, a single observer, and repeated judgements) and Case II (no simplifying assumptions, and many observers making single judgements).

It seems that graded pairwise comparison have seldom been used in audio evaluation. The authors of (Schuck et al., 1993) collated practical evaluations of loudspeakers using multiple comparisons and pairwise comparison, as part of a study on the interaction between the loudspeaker and listening room. Pairwise comparison had primarily been used as the dichotomous choice, but the authors also asked subjects to give each loudspeaker a rating. They found that similar results were obtained using the multiple-comparisons paradigm, the graded paired comparison paradigm (analyzed with MANOVA), and dichotomous pairwise comparison analyzed with the Bradley-Terry method. They noticed that multiple comparisons resulted in a wider spread of ratings than pairwise comparison, but that the power to detect differences between loudspeakers was greater in the multiple-comparisons paradigm.

Francombe, Brookes, Mason, Woodcock, et al. (2017) used a continuous scale rating during pairwise comparisons. The authors admitted that this was more demanding for participants than a forced-choice task. No further details were provided in (Francombe et al., 2017) regarding the use of the rating scale in the context of Thurstone's Case V. In (Lee & Rumsey, 2004) multiple comparisons were used, but technically the rating involved a pairwise comparison of each stimulus with a reference stimulus. However, multivariate ANOVA — not the Thurstone's probabilistic model — was used to quantify the effect under investigation.

All computations and model implementations were carried out in MATLAB.

Design and course of experiment

Each subject participated on different days, i.e. in different sessions. Sessions differed only in the loudspeaker setup (see Fig. 2), so that each subject experienced a different spatial arrangement of loudspeakers in each session. Listeners were free to choose the dates of their participation, but sessions with repeated arrangements were excluded. Subjects were unaware of the actual loudspeaker arrangement on the days they selected. Thus, the arrangement factor was randomized. The experiment employed a randomized pairwise AB test methodology without repetitions, focused on the perceptual evaluation of audio reproduced by different loudspeaker systems. Prior to the test phase, each participant completed a training session comprising several examples, which served to familiarize them with the experimental interface and procedure. Participants were presented with eight distinct audio samples. Each sample was played in two versions, labeled "A" and "B", corresponding to a pairwise comparison between two of the three systems (DML, PLS_D, and PLS_G). All possible system pairings were included: DML/PLS_D, DML/PLS_G, and PLS_D/PLS_G. To avoid bias, each pairing was tested in both assignment orders (e.g., A = DML, B = PLS_D; and A = PLS_D, B = DML). However, each participant only experienced one fixed order per pairing during their session to limit session length. The order of sample presentations was randomized individually for each participant, ensuring unbiased and diverse evaluation sequences. Participants could listen to each sample as many times as needed and switch freely between versions A and B, in accordance with ITU-R (2015) recommendations. After making a selection, they could not return to that trial, preserving the integrity of each evaluation. The GUI was designed to avoid audible artifacts during switching. Generally, participants operated the system with ease. Post-session remarks often indicated that the differences between versions were subtle but perceptible, validating the sensitivity and clarity of the test environment. Data from each session contributed to the global analysis across all 29 participants. Each audio sample was limited to a duration of 48 seconds and shaped with fade-in and fade-out envelopes. This design follows findings by Koehl and Paquier (2013), who demonstrated that excerpts longer than five seconds enhanced discrimination sensitivity between loudspeakers. The musical material represented a variety of genres: two classical music excerpts (symphonic and chamber), choral music, instrumental and vocal jazz, stage music (Latin and guitar), blues rock, and fado. All samples were in CD-quality WAV format (16-bit, 44.1 kHz) and normalized based on their RMS levels after equalization. Differences between samples were reduced to below 1 dB, and all samples were presented at the same calibrated level.

Detailed information regarding the musical excerpts that were used is provided in Table 2.

Table 2

Music excerpts used in the experiment

Excerpt	Title	Artist	Music Genre
1	*Piano Concerto No. 1 in E Minor, Op. 11: I*	Sinfonietta Cracovia - dir. Jerzy Dybał, sol. Szymon Nehring	Classical – symphonic music
2	*Nyne otpushchayeshi*	Church Slavonic School Choir	Classical – choral music
3	*Uma Casa Portuguesa by Artur Fonseca*	Amalia Rodrigues	Fado
4	*Machine Gun*	Jimi Hendrix	Rock / Funk fusion
5	*Get It While You Can by J. Ragovoy and C. Taylor*	Janis Joplin	Blues rock and soul
6	*Leaving*	Mateusz Pałka Trio	Contemporary jazz
7	*Candeeiro de Saudade by Roque Ferreira*	Thais Macedo	Samba / MPB
8	*Piano Trio No. 2 in E minor*	AMKP Piano Trio	Classical – chamber music

Most individual sessions lasted between 20 and 30 minutes. While the length of samples and sessions slightly exceeded the recommendations from ITU-R (2019), participants were permitted to pause at any time. The procedure closely followed the "paired comparisons" method suggested therein.

Results

The structure of the data that was obtained in this experiment is presented in Fig. 7. This was organized into two $75 \times 8 \times 3$ matrices (indexed as $i \times j \times k$), with the individual participants' sessions as rows (the total number of evaluations was 75 – see Section Listeners), eight musical excerpts as columns, and the three pairs of compared systems as layers (the third dimension).

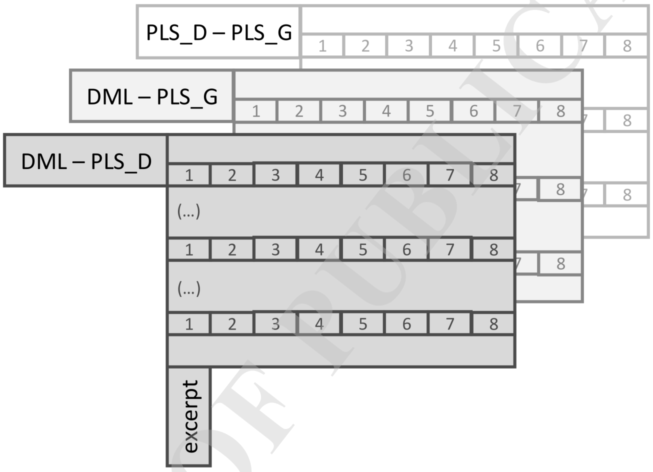


Figure 7. Structure of data obtained in the experiment, for both matrices (GQ and FS).

One matrix represented assessments of *overall preference* (OP), the other *Spatial Impression* (SI). The matrixes contained all participants' results from the sessions they participated in. Analyzing data in the 1-TL and 3-TL groups (see Section Listeners) required dividing OP and SI matrixes into appropriate submatrixes. In this work, experimental variables z_{12} and $\delta_{\Psi_1-\Psi_2}$ (2) were determined individually from the data that was contained in each of the columns of the **OP** and **SI** matrixes.

When the pairs of perceptual differences were determined according to (2), the final set of quality scores for all three systems was calculated with the least squares solution (Tsukida & Gupta, 2011). These results are presented in Figures 8-11 separately for each of the excerpts. The mean values of the quality scores for each loudspeaker pair are also shown. The values in reffig:OP1-11 are perceptual units. According to Thurstone's model, the placement of the zero point of a perceptual scale is completely arbitrary, so the scale is an intervallic one. It is convenient to perform calculations so that the mean of each of the three quality scores is zero for each excerpt. This is the way that the scores are presented.

The value of one (z score = 1) in Thurstone's model is interpreted as one perceptual unit; therefore, most of the scores should be interpreted as being low; consequently, the perceptual differences that could be found in this experiment were low. The scores in the 1-TL group exhibited a substantially higher spread than those in the 3-TL group. This can be attributed to lower experience of participants in the 1-TL group and to intra-subject variability in assessments, which is reduced in the 3-TL group by averaging three independent evaluations. In the 3-TL group, only five cases were the perceptual differences $\overline{\Psi_1} - \overline{\Psi_2}$ between loudspeaker pairs close to 1.5. For overall preference, these cases were DML and PLS_G (in Excerpts 4 and 5), PLS_D and PLS_G (in Excerpt 7); for spatial impression: DML and PLS_G (in Excerpts 4 and 5). Excerpts 4, 5 and 7 appeared to be the most revealing in the listening evaluations according to both attributes. It is noteworthy that for both attributes, DML scored substantially lower than the other systems in Excerpt 4, while it scored substantially higher than the others in Excerpt 5. This demonstrates considerable effect of the excerpt on evaluation.

Another observation is that when scores were averaged over excerpts, for both attributes, the 1-TL group evaluated DML lower than both cone speaker systems (although the difference, expressed in perceptual units, was low), whereas the 3-TL group ranked it between the two cone systems.

With a graded pairwise comparison, it is possible to perform hypothesis testing based on the raw perceptual comparisons using parametric tools. The same procedure was performed on the data from **OP** and **SI** matrixes, containing results for the 1-TL and 3-TL groups. The columns of each matrix (see Fig. 7) were assumed to be samples, and μ_{jk} sample averages were found. For each column j, the average of means $\mu_j = (\mu_{j1} + \mu_{j2} + \mu_{j3})/3$ was calculated. Then, individual divergencies from μ_j were calculated according to $\Delta_{j1} = \mu_{j1} - \mu_j$, $\Delta_{j2} = \mu_{j2} - \mu_j$, $\Delta_{j3} = \mu_{j3} - \mu_j$. The Δ_{jk} values were effects of interest; they were used in the standard procedure for evaluating the statistical significance at a $p = 0.05$ level, with the two-tailed t-test ($df = 14$ for the 1-TL group and $df = 44$ for the 3-TL group), and with Bonferroni correction ($m = 3$). All of the Δ_{jk} values for both matrixes turned out to be insignificant; this confirmed that perceptual differences that are presented in Figures 8–11 were low.

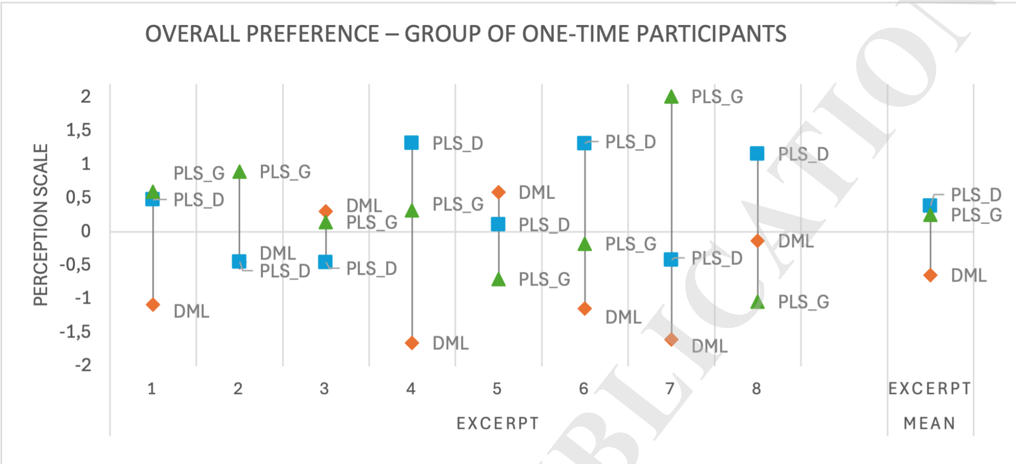


Figure 8. Quality scores for all three systems: overall preference, 1-TL group.

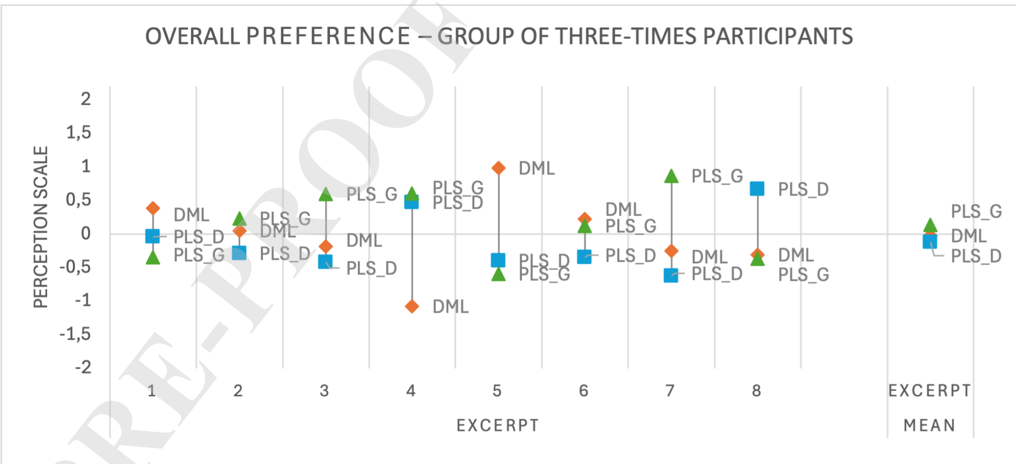


Figure 9. Quality scores for all three systems: overall preference, 3-TL group.

Assessments of the perceptual differences among the loudspeaker systems were dispersed among the individual listeners, with fairly consistent values of standard deviation (which were calculated from the columns of the **OP** and **SI** matrixes). The average values of the standard deviation were for **OP** – 1-TL group: 2.06 and 3-TL group: 2.23; for **SI** – 1-TL group: 1.93 and 3-TL group: 2.09.

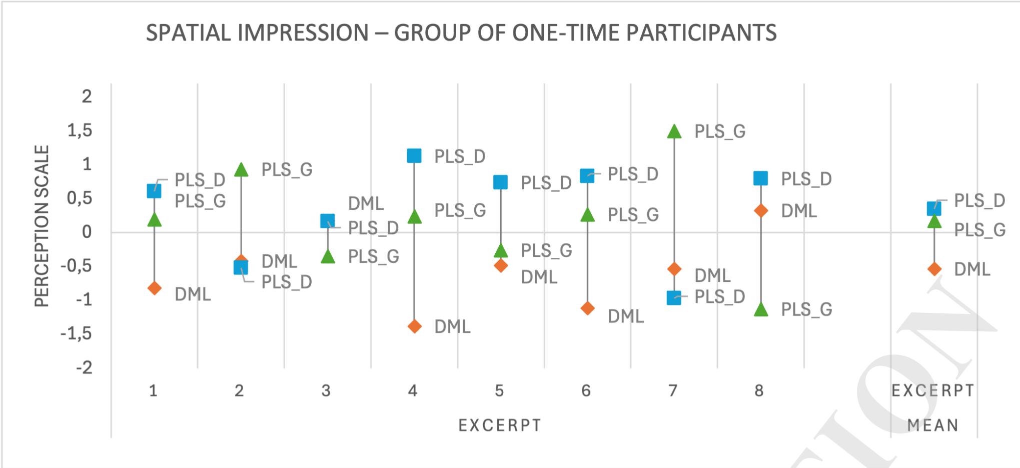


Figure 10. FQuality scores for all three systems: overall preference, 1-TL group.

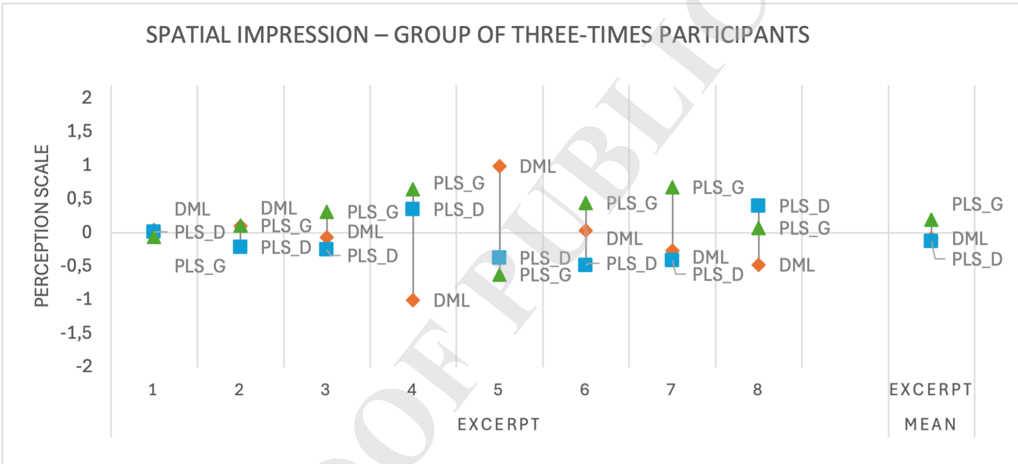


Figure 11. Quality scores for all three systems: overall preference, 3-TL group.

Although eliminating less-reliable listeners from the results of listening tests is not recommended in general (Bech & Zacharov, 2006), such a possibility was analyzed in this work. At first, intra-listener consistency was evaluated by calculating the pairwise correlations between the rows of the **OP** and **SI** matrixes that represented the sessions of the same listener. Correlations were not expected, as the sessions of an individual listener were carried out with different arrangements of loudspeakers (see Section - Positions of Loudspeakers). All of the correlations were low, so care should be taken when drawing conclusions from them.

In principle, the credibility results of the triple pairwise comparisons can be verified by the transitivity test Birnbaum (2023). Preferences are said to be transitive when if $X > Y$

and $Y > Z$ for all X, Y, and Z (then, $X > Z$). However, this is not a reliable measure of the consistency of the preferences, as individual preferences contain random errors. Moreover, sets of X, Y, and Z may be transitive by chance, so listeners who did not hear any differences between the stimuli might pass the test.

When applied to a particular problem, the quality of the Thurstone model can be verified by reversing the procedure. The final values in the perceptual scale $\overline{\Psi_1}, \overline{\Psi_2}$ can be used to determine z_{12} from (2), and the latter value can be converted to estimated value $p_e(\Psi_1 > \Psi_2)$. Finally, the two values of p that were obtained experimentally ($p_{\Psi_1>\Psi_2}$) and estimated ($p_e(\Psi_1 > \Psi_2)$) may be used to evaluate the quality of the model. The close correspondence between the two values indicated the applicability of the model.

The reverse procedure was performed on the entire **OP** and **SI** matrixes, with the 1-TL and 3-TL groups combined, and the pairs of the results were tested for statistical significance by a χ^2 test. Forty-eight pairs were tested; these results are given in Table 3. The analysis produced all insignificant values except for one instance, which indicated that the used model was applicable.

The only significant value (indicating the inapplicability of the model that was used) occurred in just one comparison – in the evaluation of overall preference regarding Excerpt 7. The other two comparisons in this group also produced considerably higher χ^2 values than could be found in all of the other groups. Excerpt 7 is a samba piece with numerous percussion instruments, featuring a high amount of high-frequency content. This genre may benefit from the narrow directional characteristics of PLS monitors. It can be noticed that the lowest values of χ^2 (which indicated the very good performance of the model in both of the perceptual attributes) were obtained with Excerpts 1, 2, 3, and 5.

Discussion and Conclusions

Two key findings can be observed in the results. The primary finding is that the perceived differences between loudspeaker systems were small. This outcome is somewhat unexpected, considering the simple construction of DML speakers and the level of technological advancement in professional-grade two-way active monitors. The second finding is

Table 3

χ^2 values from tests of the fit between experimental and predicted preference proportions. The only significant value (Ex. 7, GQ, DML vs. PLS_G) is shown in bold.

	Comparison	1	2	3	4	5	6	7	8
Overall preference	DML/PLS_D	0.28	0.06	0.06	0.98	0.10	1.08	3.36	1.07
	DML/PLS_G	0.27	0.06	0.06	1.03	0.09	1.04	**4.21**	0.85
	PLS_D/PLS_G	0.30	0.06	0.06	0.98	0.10	0.94	3.23	0.66
Spatial impression	DML/PLS_D	0.04	0.00	0.02	1.53	0.24	0.12	0.63	0.85
	DML/PLS_G	0.00	0.00	0.02	0.85	0.42	0.01	0.58	0.64
	PLS_D/PLS_G	0.00	0.01	0.05	1.53	0.26	0.21	0.99	0.59

that the 3-TL group evaluated the DML higher than the 1-TL group for both attributes, rating it at a level comparable to that of the cone systems. Whether this difference is due to the greater experience of the 3-TL participants or to more consistent results within this group remains unknown.

After the completion of our study, a follow-up investigation designed to verify and extend our findings was conducted by our research team, with a partly different set of authors (Kleczkowski, Makuch, Król-Nowak, & Czesak, 2025). It was carried out under completely different experimental conditions. Important conclusions can be drawn by referencing our results to those presented in (Kleczkowski et al., 2025). The overall findings of both works are similar: equalized DMLs attain sound quality close to that of high-quality loudspeaker systems based on cone drivers. This similarity was observed despite nearly all possible differences in the applied methods. In (Kleczkowski et al., 2025) the loudspeaker arrangement was fixed, the DMLs (two were examined) were flush-mounted, both the DMLs and cone systems were equalized for their position in the listening room, the experimental paradigm was multiple comparisons with parametric and non-parametric statistical evaluation, six evaluation attributes were used, the musical excerpts differed, five out of six loudspeakers used in both experiments were different, the listening room was quite different, and finally all listeners were distinct.

A specific finding of Kleczkowski, Król, and Małecki (2015) was that for two out of three spatial attributes ("envelopment" and "stage width") the DMLs significantly outperformed the cone system, whereas the cone systems were significantly favored in the "localization" attribute. For two timbral attributes and the global attribute "pleasantness," the loudspeakers of both types were rated close.

The comparison of this study with (Kleczkowski et al., 2025) leads to an important conclusion: loudspeaker and room correction seems to improve the perceived quality of DMLs compared to loudspeaker-only (anechoic) correction. The overall relative evaluation of DMLs was moderately higher in (Kleczkowski et al., 2025) than in the current study. The type of equalization is likely to affect perceptual evaluation more than any of the following factors: flush mounting, listening room and competing cone loudspeakers. Another conclusion concerns the use of pairwise comparison: the combination of findings from both studies confirms the observation reported in (Schuck et al., 1993) that the power to detect loudspeaker differences is greater for multiple comparisons than for pairwise comparison.

Several works that were cited in the Introduction indicated that a wide directivity of loudspeakers is preferable. This was not confirmed in our experiment, as the listeners did not evaluate the spatial impression of the DMLs higher than its overall preference.

Another factor that should be taken into consideration when interpreting our results is that the conditions eliminated a considerable shortcoming of DMLs; i.e., their insufficient low-end extension (see Section - Correction at Frequencies Below 100 Hz). This shortcoming can be overcome by using a subwoofer, as implemented in (Kleczkowski et al., 2025).

The χ^2 test results showed almost all comparisons to be insignificant, affirming the model's applicability in this context (except for one instance in the overall preference evaluation).

The main conclusion from the experiment was that the widely applied technology of DML loudspeakers, after a relatively simple improvement, achieves sound quality that is evaluated as very close to that of professional quality two-way active loudspeaker systems based on traditional cone technology. With its significantly more-versatile mounting options, the DML transducer technology offers a unique potential – particularly in home and commercial applications. This technology allows the loudspeakers to blend seamlessly into the decor of a room, serving as a piece of art or a graphic rather than a freestanding piece of furniture (like traditional speakers). This suggests that DML technology could be particularly advantageous in settings where space and aesthetic integration are critical – especially as part of a multichannel sound system.

Funding

This work was supported by The Faculty of Mechanical Engineering and Robotics, AGH University of Krakow, Research Subsidy No. 16.16.130.942.

Data availability

Data will be made available on request.

References

Allison, R. (1995). Imaging and loudspeaker directivity: To beam or not to beam. In *Proceedings of the 99th audio engineering society convention.*

Anderson, D. A., & Bocko, M. F. (2015). A model for the impulse response of distributed-mode loudspeakers and multi-actuator panels. In *Proceedings of the 139th audio engineering society convention.* New York, USA. (October 29–November 1)

Angus, J. A. (2000). Distributed mode loudspeaker radiation mechanisms. In *Proceedings of the 108th audio engineering society convention.*

Azima, H., & Harris, N. J. (1997). Boundary interaction of diffuse field distributed-mode radiators. In *Proceedings of the 103rd audio engineering society convention.* New York.

Bai, M. R., & Huang, T. (2001). Development of panel loudspeaker system: Design, evaluation and enhancement. *The Journal of the Acoustical Society of America*, *109*, 2751–2761. doi: 10.1121/1.1371544

Bank, B. (2013). Audio equalization with fixed-pole parallel filters: An efficient alternative to complex smoothing. *Journal of the Audio Engineering Society*, *61*(1/2), 28–37.

Bank, G., & Harris, N. (1998, March 1). The distributed mode loudspeaker—theory and practice. In *Proceedings of the uk 13th conference: Microphones & loudspeakers.* Huntingdon, UK.

Bech, S., & Zacharov, N. (2006). *Perceptual audio evaluation—theory, method and application.* John Wiley & Sons Ltd.

Bertland, B. (1985). Loudspeaker directionality and the perception of reality. *Journal of the Audio Engineering Society*, *33*(5), 342–350.

Birnbaum, M. H. (2023). Testing transitivity of preference in individuals. *Decision*, *10*(2), 153–180. doi: 10.1037/dec0000185

Czesak, K. (2025). *Measurement analysis of distributed mode laudspeakers for the purpose of improving their sound properties* (Unpublished doctoral dissertation). AGH University of Krakow. (in Polish: Analiza pomiarowa głośników modów rozproszonych w celu udoskonalenia ich własności brzmieniowych)

Czesak, K., & Kleczkowski, P. (2023). Wybrane aspekty charakterystyk kierunkowości głośników modów rozproszonych (some aspects of directional characteristics of distributed mode loudspeakers). In *Postępy badań w inżynierii dźwięku i obrazu (advances in research on audio and video engineering)* (pp. 41–56). Retrieved from http://fbc.pionier.net.pl/id/oai:dbc.wroc.pl:125457

Czesak, K., Kleczkowski, P., & Król-Nowak, A. (2022). Metodyka wyznaczania charakterystyk kierunkowości głośników modów rozproszonych. In A. Król-Nowak (Ed.), *Postępy w inżynierii dźwięku i psychoakustyce.* Wydawnictwa AGH.

EBU Tech. (1998). *3276-listening conditions for the assessment of sound programme material: monophonic and two-channel stereophonic* (Tech. Rep.). Retrieved from https://tech.ebu.ch/docs/tech/tech3276.pdf

Evans, W., Dyreby, J., Bech, S., Zielinski, S., & Rumsey, F. (2009). Effects of loudspeaker directivity on perceived sound quality—a review of existing studies. In *Proceedings of the 126th audio engineering society convention.*

Ferralli, M. W., & Moulton, D. (1995). 360 degree dispersion frequency invariant acoustic transducer system. In *Proceedings of the 99th audio engineering society convention.*

Flanagan, S., & Harris, N. (1999, May). Loudness: A study of the subjective difference between dml and conventional loudspeakers. In *Proceedings of the 106th audio engineering society convention.* Munich, Germany. (May 8–11)

Flanagan, S., & Moore, B. C. J. (2000). The influence of loudspeaker type on timbre perception. In *Proceedings of the 109th audio engineering society convention.* Los Angeles, USA.

Flanagan, S., & Moore, B. C. J. (2001). The effect of loudspeaker type on the identification of vowel-like harmonic complexes. In *Proceedings of the 111th audio engineering society convention.*

Flindell, I. H., McKenzie, A. R., Negishi, H., Jewitt, M., & Ward, P. (1991). Subjective evaluations of preferred loudspeaker directivity. In *Proceedings of the 90th audio engineering society convention.*

Francombe, J., Brookes, T., Mason, R., Woodcock, J., et al. (2017). Evaluation of spatial audio reproduction methods (part 2): analysis of listener preference. *Journal of the Audio Engineering Society*, *65*(3), 212–225.

Gabrielsson, A., Lindström, B., & Till, O. (1991). Loudspeaker frequency response and perceived sound quality. *The Journal of the Acoustical Society of America*, *90*(2), 707–719.

Gontcharov, V., & Hill, N. (2000). Diffusivity properties of distributed mode loudspeakers. In *Proceedings of the 108th audio engineering society convention.*

Harris, N., Flanagan, S., & Hawksford, M. O. J. (1998). Stereophonic localization in rooms, comparing conventional and distributed-mode loudspeakers. In *Proceedings of the 105th audio engineering society convention.* San Francisco.

Harris, N., Gontcharov, V., & Hawksford, M. O. J. (2000). Measurement and simulation results comparing the acoustics of various direct radiators in the presence of a dominant specular reflection. In *Proceedings of the 109th audio engineering society convention.* Los Angeles.

Harris, N. J., & Hawksword, M. O. J. (2000). Introduction to distributed mode loudspeaker (dml) with first-order behavioural modeling. *IEEE Proceedings - Circuits, Devices and Systems*, *147*(3), 153–157.

Heilemann, M., Anderson, D., Roessner, S., & Bocko, M. F. (2018). Quantifying listener preference of flat-panel loudspeakers. In *Proceedings of the 145th audio engineering society convention.* New York, USA. (October 17–20)

Heilemann, M. C., Anderson, D., & Bocko, M. F. (2017). Equalization of localized sources on flat-panel audio displays. In *Proceedings of the 143rd audio engineering society convention.* New York, USA. (October 18–21)

Heilemann, M. C., Anderson, D. A., Roessner, S., & Bocko, M. F. (2021). The evolution and design of flat-panel loudspeakers for audio reproduction. *Journal of the Audio Engineering Society*, *69*(1-2), 27–39. doi: 10.17743/JAES.2020.0057

Ho, J.-H., & Berkhoff, A. P. (2015). Flat acoustic sources with frequency response correction based on feedback and feed-forward distributed control. *The Journal of* *the Acoustical Society of America*, *137*(4), 1948–1957.

Hörchens, L., & de Vries, D. (2011). Comparison of measurement methods for the equalization of loudspeaker panels based on bending wave radiation. In *Proceedings of the 130th audio engineering society convention.* London, UK.

ITU-R. (2015). *Bs.1116-3 methods for the subjective assessment of small impairments in audio systems* (Tech. Rep.). International Telecommunication Union.

ITU-R. (2019). *Recommendation bs.1284-2 general methods for the subjective assessment of sound quality* (Tech. Rep.). International Telecommunication Union.

Jeon, O., Ryu, H., Kim, H.-G., & Wang, S. (2020). Vibration localization prediction and optimal exciter placement for improving the sound field optimization performance of multi-channel distributed mode loudspeakers. *Journal of Sound and Vibration*, *481*, 115424. doi: 10.1016/j.jsv.2020.115424

Jung, J., Jensen, J. S., Jeong, C.-H., Jeon, O., & Wang, S. (2021). Optimizing a distribution of resonators on a thin plate for the desired sound radiation. *Journal of Sound and Vibration*, *496*, 115926. doi: 10.1016/j.jsv.2020.115926

Karjalainen, M., Piirila, E., Järvinen, A., & Huopaniemi, J. (1999). Comparison of loudspeaker equalization methods based on dsp techniques. *Journal of the Audio Engineering Society*, *47*(1/2), 3–27.

Kleczkowski, P., Król, A., & Małecki, P. (2015). Reproduction of phantom sources improves with separation of direct and reflected sounds. *Archives of Acoustics*, *40*(4), 575–584.

Kleczkowski, P., Makuch, T., Król-Nowak, A., & Czesak, K. (2025). Equalized distributed mode loudspeakers rival conventional ones in listening tests. *The Journal of the Acoustical Society of America*, *157*(4), 2528–2541. doi: 10.1121/10.0036373

Koczkodaj, W. W. (2016). Pairwise comparisons rating scale paradox. In N. T. Nguyen & R. Kowalczyk (Eds.), *Transactions on computational collective intelligence xxii* (Vol. 9655, pp. 1–10). Springer. doi: 10.1007/978-3-662-49619-0$_1$

Koehl, V., & Paquier, M. (2013). A comparative study on different assessment procedures applied to loudspeaker sound quality. *Applied Acoustics*, *74*, 1448–1457. doi:

10.1016/j.apacoust.2013.06.008

Lee, H. K., & Rumsey, F. (2004). Elicitation and grading of subjective attributes of two-channel phantom images. In *116th aes convention.*

Linkwitz, S. (2007). Room reflections misunderstood? In *Proceedings of the 123rd audio engineering society convention.*

Lu, G., & Shen, Y. (2009). Model optimization of orthotropic distributed-mode loudspeaker using attached masses. *The Journal of the Acoustical Society of America, 126*, 2294–2300. doi: 10.1121/1.3212943

Lu, G., Shen, Y., & Liu, Z. (2012). Optimization of orthotropic distributed-mode loudspeaker using attached masses and multi-exciters. *The Journal of the Acoustical Society of America, 131*. doi: 10.1121/1.3672642

Moulton, D. (1986). The significance of early high-frequency reflections from loudspeakers in listening rooms. In *Proceedings of the 81st audio engineering society convention.*

Newell, P., & Holland, K. (2019). *Loudspeakers for music recording and reproduction.* Routledge.

Norcross, S. G., Soulodre, G. A., & Lavoie, M. C. (2004). Subjective investigations of inverse filtering. *Journal of the Audio Engineering Society, 52*(10), 1020–1033.

Olive, S. E. (2004a). A multiple regression model for predicting loudspeaker preference using objective measurements: Part ii—development of the model. In *Proceedings of the 117th audio engineering society convention.*

Olive, S. E. (2004b). A multiple regression model for predicting loudspeaker preference using objective measurements: Part ii—development of the model. In *Proceedings of the 117th audio engineering society convention.* Northridge, CA, USA: Harman International Industries, Inc.

Olive, S. E. (2004c). A multiple regression model for predicting loudspeaker preference using objective measurements: Part i—listening test results. In *Proceedings of the 116th audio engineering society convention.* Berlin.

Perez-Ortiz, M., Mikhailiuk, A., Zerman, E., Hulusic, V., Valenzise, G., & Mantiuk, R. K. (2019). From pairwise comparisons and rating to a unified quality scale. *IEEE*

Transactions on Image Processing, *29*, 1139–1151. doi: 10.1109/TIP.2019.2936103

Pilch, A., & Kamisiński, T. (2011). The effect of geometrical and material modification of sound diffusers on their acoustic parameters. *Archives of Acoustics*, *36*, 955–966. doi: 10.2478/v10168-011-0065-1

Pueo, B., López, J. J., Ramos, G., & Escolano, J. (2009). Efficient equalization of multi-exciter distributed mode loudspeakers. *Applied Acoustics*, *70*, 737–746.

Roessner, S., Heilemann, M. C., & Bocko, M. F. (2019). Evaluating listener preference of flat-panel loudspeakers. In *Proceedings of the 147th audio engineering society convention.* New York, USA. (October 21–24)

Schuck, P. L., Olive, S. E., Ryan, J. G., Toole, F. E., Sally, S. L., Bonneville, M. E., & Verreault, E. S. (1993, June). Perception of perceived sound in rooms: some results of the athena project. In *Audio engineering society conference: 12th international conference: The perception of reproduced sound.*

Thurstone, L. L. (1927a). A law of comparative judgement. *Psychological Review*, *34*, 273–286.

Thurstone, L. L. (1927b). The method of paired comparisons for social values. *Journal of Abnormal and Social Psychology*, *21*, 384–400.

Toole, F. E. (1986a). Loudspeaker measurements and their relationship to listener preferences: Part 1. *Journal of the Audio Engineering Society*, *34*(4), 227–235.

Toole, F. E. (1986b). Loudspeaker measurements and their relationship to listener preferences: Part 2. *Journal of the Audio Engineering Society*, *34*(5), 323–348.

Toole, F. E. (2018). *Sound reproduction: The acoustics and psychoacoustics of loudspeakers and rooms* (3rd ed.). Routledge.

Tsukida, K., & Gupta, M. R. (2011). *How to analyze paired comparison data* (Technical Report No. UWEETR-2011-0004). University of Washington.

Wendt, F., & Höldrich, R. (2021). Precedence effect for specular and diffuse reflections. *Acta Acustica*, *5*, 1–9.

Yu, Z., Zhu, Q., Wu, M., & Yang, J. (2023). Exploring the limits of virtual source localization with amplitude panning on a flat panel with actuator array: Implications for future research. *The Journal of the Acoustical Society of America*, *154*, 1362–1371. doi: 10.1121/10.0020827

Zacharov, N. (1998). Subjective appraisal of loudspeaker directivity for multichannel reproduction. *Journal of the Audio Engineering Society*, *46*(4), 288–303.

Zenker, B., Schurmann, R., Merchel, S., & Altinsoy, M. E. (2020). Improved sound radiation of flat panel loudspeakers using the local air spring effect. *Applied Sciences*, *10*(24), 1–18. doi: 10.3390/app10248926