ARCHIVES
of
ACOUSTICS

QUARTERLY

Vol. 50, No. 3, 2025

WARSAW

ARCHIVES of ACOUSTICS

Research Papers

Review Paper

Research Paper

Acoustic Decaphonic Piano: Calculating Safe Retunings from 12-TET to 10-TET and Beyond

Aleksander BOGUCKI[(1), (2)]* , Andrzej WŁODARCZYK[(3)], Paweł NUROWSKI[(1), (4)]

[(1)] *Center for Theoretical Physics, Polish Academy of Sciences*
Warsaw, Poland

[(2)] *Center for Quantum Nanoscience, Institute for Basic Science (IBS)*
Seoul, Republic of Korea

[(3)] *Pracownia Pianin i Fortepianów*
Słupno, Poland

[(4)] *Guangdong Technion – Israel Institute of Technology*
Shantou, China

*Corresponding Author e-mail: aleksander.bogucki@fuw.edu.pl

This paper presents a method for safe retuning of fixed-pitch string instruments to alternative musical scales with fewer degrees than their original design. Our approach uses a systematic monotonic surjective mapping to assign the existing set of strings to a new, smaller set of pitch classes. The primary goal is to preserve the instrument's timbre and structural integrity by keeping string tension changes within safe limits. We demonstrate the method on a grand piano and an upright piano retuned from 12-tone equal temperament (12-TET, 12EDO) to 10-tone equal temperament (10-TET, 10EDO). Presented approach may be generalized for retuning from N- to M-step scales ($N > M$) and to other fixed-pitch string instruments. A grand piano was safely retuned using the proposed method and successfully used in a professional concert.

Keywords: monotonic surjective mapping; decaphonic piano; 10-tone equal temperament; 10TET; 10EDO; alternative instrument tuning; xenharmonic; string tension.

1. Introduction

The interest in alternative scale tunings among musicians arises from their ability to enable more accurate performance in selected harmonic progressions, which historically led to the development of various temperaments in tuning. The microtonal approach, which divides the octave into more than the typical twelve semitones (*tasto spezzato*), has been explored since the 16th century, as seen in Nicola Vicentino's *archicembalo* with divided keys known as *tasto spezzato* (PILCH, TOPOROWSKI, 2014). Alternative musical scales can also originate from the characteristic spectrum of a given instrument, as in the case of gamelan music, which employs the *slendro*

and *pélog* scales (SETHARES, 1998). Some scales are based on repeating intervals larger than the typical octave, such as the Bohlen–Pierce scale, which uses a tritave (3:1 frequency ratio) as its fundamental unit (MATHEWS *et al.*, 1988), or the *hyperpiano*, which follows a hyperoctave structure with a 4:1 ratio (HOBBY, SETHARES, 2016).

Alternative tunings are gaining visibility beyond specialist contexts. Popular musicians such as Jacob Collier and Dua Lipa have incorporated microtonal elements in widely streamed songs, reaching millions of listeners (BANDY, 2025; FRASER, 2023). Online communities also play a growing role: YouTube creators including Adam Neely, David Bruce, Georg Vogel regularly explore harmony, tuning, and composi-

tional techniques – often presenting alternative temperaments in an accessible format. Microtonality is also present in mainstream entertainment; for instance, the track *Trees in the Depths of the Earth* from the 1996 video game *Kirby Super Star* uses the microtonal Maqam Rast scale and has reached a broad audience through a franchise that has sold over forty million copies (Martin, 2025). This growing engagement is rooted in a longer musical tradition. Early adopters such as Alois Hába and Julián Carrillo, along with 20th-century figures like Harry Partch, Ben Johnston, and Wendy Carlos, developed extensive bodies of microtonal work. Acoustic instruments such as the Sauter Microtone piano (Thomas, 1996) and the Clavemusicum Omnitonum by Krebs Cembalobau reflect continued interest in physical realizations of microtonal tuning systems.

Another musically interesting approach is to reduce the number of scale steps. In the case of the 10-tone equal temperament, where 'tone equal temperament' is abbreviated as TET, also referred to in the literature as the 10 equal division of the octave (10EDO), the octave is divided into ten equal steps, leading to larger intervals between each step compared to a standard semitone. The theoretical foundations of the 10-TET scale were discussed in (Sethares 1998, pp. 259–270). In particular, three structural features make 10-TET a musically functional example. This includes:

1) it supports neutral intervals, such as the neutral third and neutral sixth. These lie between traditional major and minor forms and allow for the construction of neutral chords – harmonic entities that extend the available vocabulary;

2) 10-TET enables chord cycles built on repeated neutral thirds. This structure forms a circle of thirds, functionally analogous to the circle of fifths in 12-TET. While the steps are different, the pattern supports harmonic progression and modulation in a coherent way;

3) 10-TET admits two types of tritone-based cadences, which can resolve to neutral chords in distinct ways. These structures provide multiple paths for harmonic motion and modulation, compensating for the absence of a major-minor dichotomy.

Several compositions have already been written for 10-TET (Xenharmonic Wiki, n.d.), with some specifically composed for a 10-TET piano by Hunt (2022), Senpai (2023), Sevish (2017), and Hideya (2021). The 10-TET scale also exhibits unique mathematical properties, which we demonstrate in Appendix. Despite its theoretical foundation and existing compositions, to our knowledge, no acoustic 10-TET piano has ever been built, and all performances in this tuning have relied on electronic synthesizers.

Retuning an acoustic piano to an alternative scale presents significant technical challenges. A change in tuning affects string tension and may risk breaking strings, altering the instrument's timbre, or making some strings too slack to vibrate properly.

A clear example of this challenge came when renowned (Stanevičiūtė, Janicka-Słysz, 2022) jazz pianist Leszek Możdżer approached us with a practical request. He wanted to perform on an acoustic piano tuned in 10-TET rather than the standard 12-tone equal temperament (12-TET, 12EDO). His aim was to achieve this new tuning without making major physical modifications – retuning alone should suffice. Although this idea may seem straightforward, direct methods of retuning can lead to extreme pitch deviations in the upper or lower registers, creating problems for both tone quality and instrument safety.

There are multiple reasons why acoustic 10-TET pianos have never been built. To construct an acoustic piano designed for the 10-TET scale, one must overcome all the challenges associated with designing a new standard piano, including significant economic costs and numerous design decisions specific to 10-TET, a largely unexplored field that can only be fully evaluated in a finished instrument. These same obstacles also contribute to the slow evolution of standard piano development.

In this paper, we have chosen the opposite approach: instead of building a new instrument from scratch, we start with an existing piano and introduce the minimum necessary modifications – exclusively through retuning – to achieve the desired effect: an acoustic 10-TET piano. Thus, we focus on a safe and practical method for working within the piano's existing mechanical limits.

We propose a method called monotonic surjective mapping that safely retunes a piano by preserving the original frequency range and maintaining string tensions within acceptable limits. The method can also be extended to other fixed-pitch string instruments, such as harpsichords and harps, and generalized to retune from any N-step scale to an M-step scale where $N > M$.

To validate our approach, we studied two instruments: the Nyström upright piano, which has 85 keys and served as a testing platform, and the Steinway Model B grand piano, which has the typical 88 keys. The presented method was successfully applied and demonstrated at a jazz concert performed by Leszek Możdżer on July 13, 2023 (Tomala 2024), and is used on his albums (Możdżer *et al.* 2024; 2025).

2. Retuning from 12-TET to 10-TET in a standard way demands extending the frequency range of an acoustic instrument

A key consequence of transitioning from 12-TET to 10-TET while utilizing all available strings is ex-
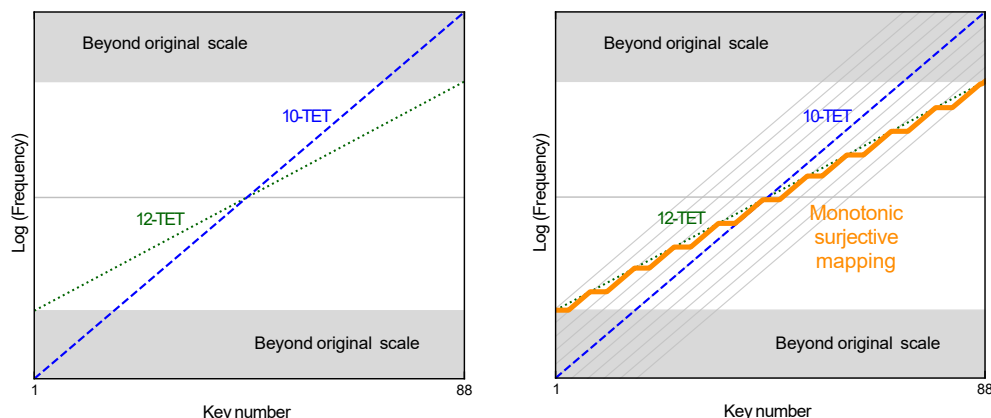
Fig. 1. Idea of monotonic surjective mappings: the problem of extending the original instrument scale when retuning from the 12-TET to the 10-TET scale (left panel); how to solve this problem using monotonic surjective mappings (right panel).

ceeding the instrument's total frequency range span. In 12-TET, the frequency progression follows a well-defined logarithmic slope, where each step corresponds to a fixed frequency ratio of $2^{1/12} \approx 1.0595$. However, in 10-TET, each step is larger, with a ratio of $2^{1/10} \approx 1.072$, meaning that for the same number of keys, the frequency span is stretched. As a result, if an instrument originally designed for 12-TET is simply retuned to 10-TET without additional constraints, its lowest strings may become too loose to function properly, while the highest strings can be subjected to excessive tension, increasing the risk of breakage.

To quantify this effect, consider a standard 12-TET piano, where the fundamental frequency range spans from 27.5 Hz (A_0) to 4186 Hz (C_8). Applying a standard mapping from 12-TET to 10-TET (centered at $C_4(40) = 261.63$ Hz[1]) shifts the lowest fundamental frequency to 17.53 Hz, far below the playable limit for an acoustic piano, while the highest frequency extends to 7288.3 Hz, well beyond the structural limits of typical piano strings. As the tension of a string is proportional to the square of its frequency, the tension ratio is given by $T/T_{\mathrm{orig}} = (f/f_{\mathrm{orig}})^2$, which results in lowering the tension of the first string to approximately 40 % of its original value and increasing the tension of the highest string to above 300 % of its original tension. This extreme expansion in the frequency range, leading to severe tension changes, is the fundamental reason why such retunings have, until now, only been implemented in electronic synthesizers rather than in acoustic instruments. An illustration of this problem is

presented in the left panel of Fig. 1. The left panel explains the problem of extending the original instrument scale when retuning from the 12-TET to the 10-TET scale. The vertical axis represents the frequencies of each note on a logarithmic scale. Both scales appear as straight lines but with different slopes. For the 10-TET scale (blue dashed line), each scale step is larger than for the original 12-TET scale (green dotted line). The gray areas at the bottom and top parts of the plot show the frequency range by which the instrument scale must be extended if a standard mapping is used for retuning from 12-TET to 10-TET. The right plot illustrates how to solve this problem using monotonic surjective mappings. One example of a monotonic surjective mapping is marked with an orange curve. This line connects the frequency of the first key in the original 12-TET tuning with the frequency of the last key in the original tuning by following either lines with the same slope as the alternative 10-TET tuning (examples of them are marked with gray lines) or remaining constant (horizontal).

3. Monotonic surjective mappings

The solution to the above problem proposed in this paper aims to possibly preserve the two outermost frequencies of the instrument's original scale while following the alternative 10-TET scale in between. However, this creates a contradiction, as the slope of the alternative scale is steeper than that of the original one. To resolve this, we allow for a monotonic surjective mapping, meaning that some pitches from the alternative scale can be repeated.

This approach results in a large number of possible mappings, many of which do not align naturally with the standard keyboard layout. A logical way to introduce order into these mappings is to preserve the octave interval (12 key distance) on a normal keyboard. This has the advantage that trained pianists already

---

[1]In this paper, we use a simplified musical notation: $C_4(40)$ is written as C(40), meaning that the 40th key of the analyzed instrument corresponds to a key associated with the note C. This notation also emphasizes that we consider only the strings, which, in the case of a piano, are always connected to the piano action. The action is triggered by a key positioned within a keyboard that follows the fixed Halberstadt layout. In other fixed-pitch string instruments, such as a diatonic harp, the fixed pattern of the diatonic scale is represented by the colors of the strings.

Table 1. 12-TET and 10-TET scales expressed in cents.

a) 12-TET scale

Key name	C	C#	D	D#	E	F	F#	G	G#	A	A#	B
Interval [¢]	0	100	200	300	400	500	600	700	800	900	1000	1100

b) 10-TET scale

Step	1	2	3	4	5	6	7	8	9	10
Interval [¢]	0	120	240	360	480	600	720	840	960	1080

have the octave distance embedded in their muscle memory, making adaptation to the alternative mapping significantly easier.

This leads to a more formal definition of the assumptions that describe monotonic surjective mappings:

– the alternative scale is mapped onto the original scale surjectively. This ensures that no string (key) is omitted, meaning that every string set is assigned a pitch from the new scale;

– the mapping is monotonic, meaning that each subsequent string has an equal or higher frequency. This allows for repeated sounds in the alternative scale while maintaining the conventional left-to-right increasing pitch layout expected by pianists;

– the octave interval is preserved, ensuring that a pianist playing an octave in the original tuning will still play an octave after retuning.

The right panel of Fig. 1 demonstrates the idea how a monotonic surjective mapping resolves described problem by selectively repeating alternative scale pitches while maintaining the original frequency range.

4. Key definitions and computational framework for safe retuning

A piano consists of a set of strings (for a typical grand piano, like Steinway Model B, with 88 keys, there are 236 strings), each corresponding to a specific key on the keyboard. These keys are arranged in a repeating pattern of black and white keys known as the Halberstadt layout (MENDEL, 1949), which has been standardized for Western instruments. The pitch of each string follows a predefined tuning system, traditionally in modern Western culture based on the 12-tone equal temperament (12-TET), where each octave is divided into twelve equal steps. In this convention, the keys are assigned names based on letter notation (C, C#, D, D#, E, F, F#, G, G#, A, A#, B), with C often serving as a convenient reference point.

4.1. Expressing intervals in cents

A useful measure of the frequency ratio (musical interval) is the cent, a logarithmic unit that divides one octave (ratio 2:1) into 1200 cents. If f_1 and f_2 are two frequencies, their difference in cents, Δc, is given by $\Delta c = 1200 \log_2(f_2/f_1)$. A single semitone in 12-TET spans exactly 100 ¢. In contrast, in a 10-TET scale, each step between intervals is larger, measuring 120 ¢ as presented in Tables 1a and 1b.

4.2. Tuning point

To define a tuning system, one particular note is chosen as a reference, here called the tuning point (TP). This is the frequency from which all other pitches in the scale are derived. A common TP is $A_4 = 440$ Hz, which serves as the modern international tuning standard[2]. In this paper, unless explicitly stated otherwise, we assume the TP to be $A_4 = A(49) = 440$ Hz. However, this is an arbitrary choice, and tuning can be established from any key.

4.3. Pinning point

The pinning point (PP) we define as a main frequency alignment point between the original and alternative scales. Unlike the TP, which defines the frequency system, the PP is selected based on where the two tuning systems coincide at a particular key (string). At the PP, one note has the same frequency in both scales, ensuring that this pitch remains unchanged during the transition from 12-TET to 10-TET. The choice of PP affects how the mapping between old and new pitches relate to each other.

As illustrated in Fig. 2, the blue curve represents relative frequencies (intervals) in the 12-TET scale within one octave, where one note is chosen as a reference (called here TP) for tuning the entire system. A common example of a TP is the international standard pitch, $A_4 = 440$ Hz. Starting from this note, the frequencies of other steps in the 12-TET scale are calculated. In a 12-TET scale, each step corresponds to 100 ¢, whereas in a 10-TET scale, each step is larger, corresponding to 120 ¢. As a result, the relative frequency curve for the 10-TET scale is steeper (120 ¢/step) compared to the 12-TET scale (100 ¢/step). In the plot, the blue curve represents

---

[2]Other examples include scientific pitch (Verdi pitch) with $C_4 = 254$ Hz or French pitch (diapason normal) with $A_4 = 435$ Hz. In some historical instruments tuning was as low as $A_4 = 415$ Hz (ROSE, LAW, 2001).
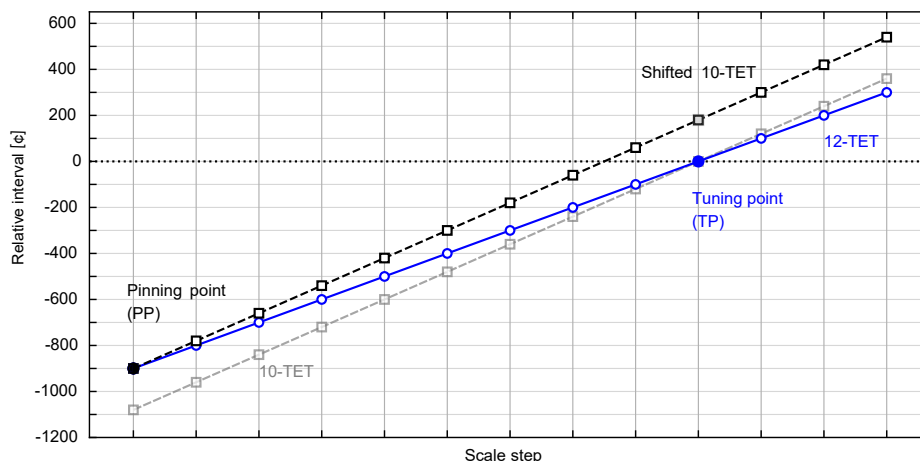
Fig. 2. Difference between the TP and the PP; the blue curve – 12-TET; gray – 10-TET from same TP;
black – shifted 10-TET intersecting the 12-TET curve at the PP.

the 12-TET scale, while the gray curve represents the 10-TET scale, both starting from the TP. However, the second curve can be arbitrarily shifted to start from a different step. This shift introduces another important point, referred to here as the PP. The note at this point has the same frequency in both the 12-TET and 10-TET scales, meaning that at this position, the two frequency curves intersect. The black curve represents the 10-TET scale shifted so now it intersects the blue 12-TET curve at the PP. The placement of the PP is independent of the TP, though in some cases, they may coincide. Selecting different PP values results in distinct mappings, influencing how the new scale aligns with the existing instrument layout.

Moreover, choosing a particular PP determines how the instrument interacts with other instruments tuned to 12-TET. A well-chosen PP increases the number of shared (common) notes between the two systems, allowing for better harmonic compatibility. If the PP is poorly chosen, the retuned instrument may lack critical common notes, making ensemble performance with 12-TET instruments more challenging. Thus, selecting an appropriate PP is not only a technical decision but also a musical one, balancing structural feasibility with practical usability.

4.4. Standard mapping

We define a standard mapping as the simplest way to retune from 12-TET to 10-TET. One selects a TP and a PP, then assigns each key to consecutive steps of the new scale, moving outward from the PP. Sethares (1998, pp. 259–270) and Hunt (2021) describe similar direct approaches for the 10-TET scale. Sethares suggests middle C as the starting point, PP = C_4 = C(40) = 261.63 Hz, while Hunt uses PP = C_2 = C(16) = 65.41 Hz. In both cases, they perform on electronic synthesizers rather than acoustic instruments.

4.5. Number of possible monotonic surjective mappings

As a consequence of the assumptions made regarding the monotonic surjective mappings, we retain the octave (12 keys apart) but distribute 10 steps of the alternative scale within it. This requires selecting two keys per octave for repeated sounds, leading to 66 possible key assignments for a given PP. This situation can be generalized to arbitrary scales by considering the number of ways to assign M-steps of the alternative scale to N-keys of the original scale, allowing for repeated steps while preserving order. The number of such mappings is given by the binomial coefficient $C(N, M)$:

$$C(N, M) = C(N - 1, M - 1) + C(N - 1, M)$$
$$= \frac{N!}{M!(N - M)!}, \tag{1}$$

where $N \geq M > 1$ and '!' denotes the factorial operation. The first term counts all monotonic mappings while excluding the 'cyclicity' of the musical scale, meaning that the last step is not equivalent to the first step. The second term accounts for mappings where the last step is equal to the first step due to the cyclic nature of the scale. While there are only 12 unique steps of the original scale, the physical properties of the instrument introduce additional complexity. Unlike the keyboard, which maintains translational symmetry across octaves – where shifting by an octave results in an equivalent musical structure – the strings do not share this symmetry. Each string has a unique tension. Consequently, there are 88 (number of keys) unique PP, leading to a total of 88 × 66 = 5808 possible mappings that fulfill our assumptions. In this paper, for simplicity, we discuss subset of PP from middle octave (from C(40) to B(51)).

4.6. Operating point

The operating point (OP) of a given string is defined as the ratio, expressed as a percentage, of the string tension to the breaking tension of that string, based on manufacturer data: $\text{OP} = T/T_{\text{break}} \times 100\,\%$. For an instrument designed for a specific tuning system, such as 12-TET, the original OP values correspond to an optimal tension that ensures the desired timbre and sound quality.

4.7. Mapping signature
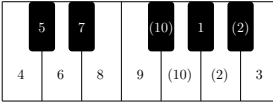
As briefly explained in Fig. 1, a monotonic surjective mapping describes how to assign 10 steps of the 10-TET scale to 12 keys of the 12-TET scale. To unambiguously identify the mapping, we must specify the PP, which indicates which step of the original 12-TET scale corresponds to the 1st step of the alternative 10-TET scale. We also need to identify which steps 'break' the ascending sequence by repeating the previous steps from the alternative scale. For example, for PP = G#(48) and choosing 2nd and 10th steps of alternative scale, the full mapping is fully defined, resulting in the sequence —1, 2, 2, 3, 4, 5, 6, 7, 8, 9, 10, 10— assigning them to the 12-TET-based keyboard. This means that key G# plays the 1st step of the 10-TET scale, A plays the 2nd step, A# also plays the 2nd step, B the 3rd, C the 4th, C# the 5th, D the 6th, D# the 7th, E the 8th, F the 9th, F# the 10th, and G the 10th step. Noticeably, keys A and A# play the same pitch that corresponds to 2nd step of the 10-TET scale. Similarly keys F# and G play now the same sound that is 10th step of the 10-TET scale as presented in Table 2a. To make the notation easier to read, we refer to this mapping as PP = G#(48), MS = '2':(A, A#); '10':(F#, G), where MS is mapping signature. For convenience, we also assign each mapping a numerical label (e.g., mapping no. 18). This label is arbitrary but helps us quickly refer to different mappings in software or in larger plots such as presented in Fig. 9 without spelling out the entire step sequence each time.

Table 3 demonstrates a special repeated-key scenario, where the same new step is assigned multiple times in a row – including the possibility of three repetitions – while still respecting monotonic surjective criteria. Finally, Table 4 shows how cyclicity can come

Table 2. Typical mapping cases.

a) Mapping no. 18: PP = G#, MS = '2':(A, A#); '10':(F#, G)

Original key	C	C#	D	D#	E	F	**F#**	**G**	G#	**A**	**A#**	B
Alternative	4	5	6	7	8	9	**10**	**10**	1	**2**	**2**	3

b) Mapping no. 31: PP = G, MS = '4':(A#, B); '8':(D#, E)

Original key	C	C#	D	**D#**	**E**	F	F#	G	G#	A	**A#**	**B**
Alternative	5	6	7	**8**	**8**	9	10	1	2	3	**4**	**4**
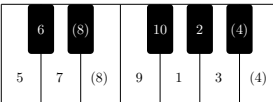
Table 3. Special double-key case, mapping no. 10: PP = C, MS = '2':(C#, D, D#).

Original key	C	**C#**	**D**	**D#**	E	F	F#	G	G#	A	A#	B
Alternative	1	**2**	**2**	**2**	3	4	5	6	7	8	9	10

Table 4. Cyclicity cases.

a) Mapping no. 55, MS = '1':(B, C, C#)

Original key	**C**	**C#**	D	D#	E	F	F#	G	G#	A	A#	**B**
Alternative	**1**	**1**	2	3	4	5	6	7	8	9	10	**1**

b) Steinway, TP = A(49) 440 Hz, PP = C(40), mapping no. 65, MS = '1':(A#, B, C)

Key	40	41	42	43	44	45	46	47	48	49	50	51
Original key	**C**	C#	D	D#	E	F	F#	G	G#	A	**A#**	**B**
Alternative	**1**	2	3	4	5	6	7	8	9	10	**1**	**1**
f_{orig} [Hz]	**261.63**	277.18	293.66	311.13	329.63	349.23	369.99	392.00	415.30	440.00	466.16	493.88
f_{alt} [Hz]	**261.63**	280.40	300.53	322.10	345.22	369.99	396.55	425.01	455.52	488.21	523.25	523.25

c) Steinway, TP = A(49) 440 Hz, PP = F#(46), mapping no. 34, MS = '5':(A#, B, C)

Key	40	41	42	43	44	45	46	47	48	49	50	51
Original key	**C**	C#	D	D#	E	F	**F#**	G	G#	A	**A#**	**B**
Alternative	**5**	6	7	8	9	10	1	2	3	4	**5**	**5**
f_{orig} [Hz]	261.63	277.18	293.66	311.13	329.63	349.23	**369.99**	392.00	415.30	440.00	466.16	493.88
f_{alt} [Hz]	244.11	261.63	280.40	300.53	322.10	345.22	**369.99**	396.55	425.01	455.52	488.21	488.21

into play, allowing repeated or tripled steps to extend beyond a single octave. These examples confirm that as long as pitch assignments remain non-decreasing (or remain constant across a short span), the mapping fulfills the monotonic surjective definition, even in more complex cases involving octave equivalence. The mappings presented in Tables 4b and 4c are particularly interesting because, at first glance, the mapping MS = '1':(A#, B, C) with PP = C(40) and MS = '5':(A#, B, C) with PP = F#(46) leads to the same assignment of new steps from the 10-TET scale to the notes on the 12-TET keyboard – the three keys A#, B, and C play identical pitch. However, the choice of PP results in different frequencies being assigned to the strings for those mappings (compare last rows in Tables 4b and 4c). As we will see later, this leads to significantly different total tension and OPs.

4.8. Calculations

Figure 3 illustrates the sequence of calculations needed to evaluate each mapping. The 'constants' box provides instrument-specific data, such as the reference frequency at the TP and each string's physical properties. The 'variables' box defines the PP and MS, which change with each tested mapping. The 'original tuning' and 'alternative tuning' boxes list values calculated for each string (for example, tension and OP) and single scalar values for the entire instrument (for instance, total tension). Finally, the 'differences' box compiles parameter differences or ratios (e.g., ΔT, Δf, etc.) that allow us to compare the original and alternative tunings under chosen criteria.

Figure 4 shows the input data used for the calculations for the Steinway Model B grand piano and the Nyström upright piano (both instruments originally were designed for TP = A(49) = 440 Hz). Those parameters correspond to the 'constants' box in Fig. 3. The values for Nyström (e.g., string lengths, diameters) were measured directly, and for the Steinway are taken from MATTHIAS (1990). Notable differences include the longer speaking lengths (the lengths of vibrating part of the strings) for the grand piano (compared to the upright) and distinct transitions between
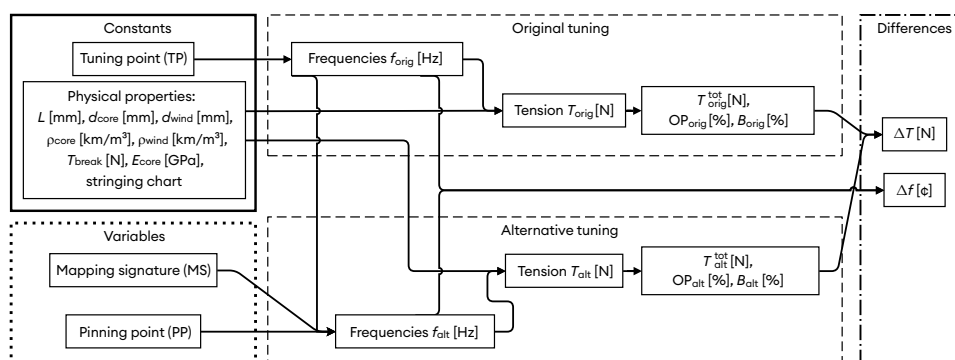
Fig. 3. Flowchart of calculations for mapping an instrument's tuning from an original scale with N-steps (e.g., 12-TET) to an alternative scale with M-steps.
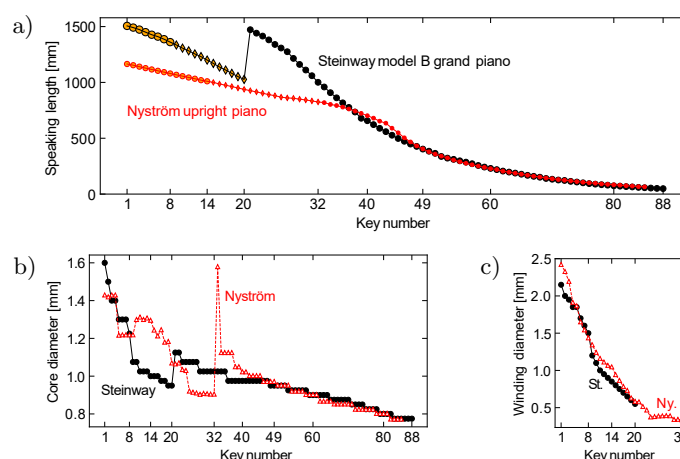
Fig. 4. Stringing scale parameters of the studied instruments: a) speaking length of the strings for each key. Each curve consists of three segments, distinguished by different symbols. The leftmost segment represents the overspun (wound) strings with a single string per key. The middle segment corresponds to the section where there are two overspun strings (bichord) per key. The final part represents the section with three strings per key (trichord) without winding; b) string core wire diameter for each key; c) copper winding wire diameter for the overspun strings.

unichord, bichord-wound, and trichord string sections. Additional windings are used in shorter bass strings to compensate for the reduced length (Rose, Law, 2001). Other parameters like maximum tension, density properties are taken from datasheets provided by manufacturer of piano rounded steel wire (Stahl- und Drahtwerk Röslau GmbH, n.d.).

4.9. String tension

For an idealized perfectly flexible, plain (unwound) string of density ρ and diameter d, vibrating at frequency f over a speaking length L, the tension T is often approximated by $T = 4\,\mu\,f^2\,L^2$, where $\mu = \pi\,\rho\,d^2/4$ is the linear mass density (kg/m). This expression assumes negligible bending stiffness. In case of wounded (overspun) strings the tension is determined by the core wire as the winding is made from soft (compared to core made out of steel) copper wire (Bucur, 2016).

4.10. Overspun (wound) strings

For copper-wound bass strings, the tension formula (Louchet, 2021) generalizes to

$$T = \frac{\pi\,f^2\,L^2}{10^{12}}\left[\rho_{\text{core}}\,(d_{\text{core}})^2 + \left(\frac{\pi\,\rho_{\text{wind}}}{4}\right)(\Omega^2 - d_{\text{core}}^2)\right],$$

where d_{core} and d_{wind} are the core and winding diameters, $\Omega = d_{\text{core}} + 2\,d_{\text{wind}}$ is the outer diameter, and ρ_{core}, ρ_{wind} are material densities (e.g., $7750\,\text{kg/m}^3$ for steel, $8920\,\text{kg/m}^3$ for copper). The factor 10^{12} accounts for unit conversions from millimeters to meters.

4.11. Inharmonicity for finite stiffness

Real musical strings have finite stiffness, so their partial frequencies deviate from integer multiples of the fundamental. If the fundamental frequency is f_0, then the n-th partial can be approximated by $f_n = n f_0 \sqrt{1 + Bn^2}$, where B is the inharmonicity coefficient. According to Fletcher (1964), for a solid steel string with diameter d (in cm), speaking length L (in cm), and fundamental frequency f_0 (in Hz), the inharmonicity coefficient is given by $B \approx 3.95 \times 10^{10}(d^2/(L^4 f_0^2))$. For copper-wound steel strings, where d_{core} is the core diameter and d_{total} is the total diameter (including winding), Fletcher (1964) provides $B \approx 4.6 \times 10^{10}(d_{\text{core}}^4/(d_{\text{total}}^2 L^4 f_0^2))$. In all cases, a sufficiently small B is important to preserve the instrument's characteristic timbre (Louchet, 2021).

5. Results and discussion

The method described in the previous section allows us to evaluate various monotonic surjective mappings by computing key parameters such as string tension, OPs, and inharmonicity coefficients. Before comparing different mappings in detail, we first examine these parameters for selected cases to illustrate how individual mappings affect the instrument.

Figure 5 presents the OPs of each string, expressed as a percentage of the breaking force. This provides a reference for further comparisons by showing how tension varies across the keyboard. The figure includes results for the original 12-TET tuning (empty markers), the standard mappings (thick curves), and two selected surjective mappings (full markers connected by thin curves). The plot reveals that the original OPs are significantly different between the two instruments which is expected as all string parameters are different. The Steinway grand piano exhibits a smoother distribution of tension, with a maximum OP around 50%, while the Nyström upright piano reaches above 70%. This difference shows that the optimal mapping may not be the same for both instruments. In particular, since the Nyström upright piano is already closer to the breaking point near the 31st key, one would prioritize mappings that minimize additional tension.
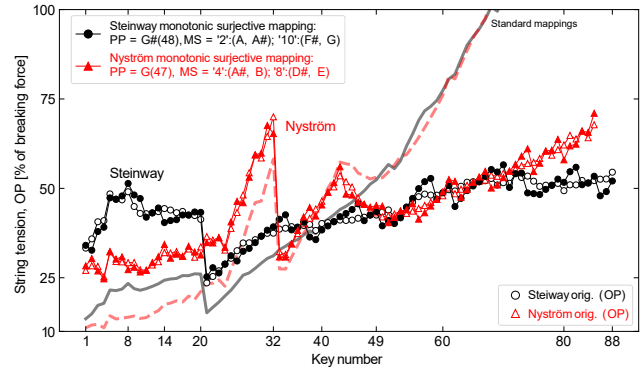
Fig. 5. String tension as a percentage of the breaking force (OPs). The empty markers represent the tension levels for the Steinway (empty black dots) and the Nyström upright piano (empty red triangles), expressed as a percentage of the breaking force for each string. These points define the original OPs (OP_orig) of the strings. The Nyström upright piano reaches 70% of the breaking force around the 31st key. The thick red and black curves represent the tension values when a standard mapping is used to retune from 12-TET to 10-TET. The breaking force is exceeded for both instruments near the 67th key. The filled markers connected by lines represent the calculated (OP_alt) values for selected monotonic surjective mappings, which closely align with the original values.

We also observe that the standard mapping leads to significant imbalances in OPs, as predicted earlier. However, Fig. 5 now quantifies this effect across all strings, confirming that standard mapping (for PP = C(40)) introduces excessive tension in the upper range – maximum allowed tension is exceeded around 67th key which leads to string breaking. Unlike the standard mapping, the OP values obtained for the selected monotonic surjective mappings closely match the original ones for both instruments.

Similarly, Fig. 6 presents the inharmonicity coefficients B calculated for the original tuning (markers). In addition, we include values for the selected monotonic surjective mappings (thick black and red curves). The dotted lines represent the inharmonicity coefficients for the standard mapping. The results show an order-of-magnitude difference between the original and standard mappings, highlighting the necessity of careful mapping to preserve the instrument's intended timbre.
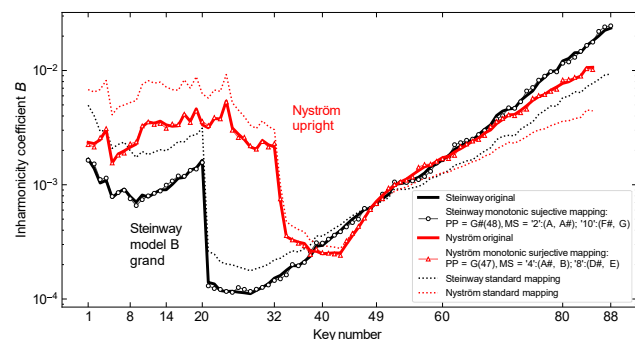
Fig. 6. Inharmonicity. The thick solid lines represent the inharmonicity coefficients B calculated for the original OP. The markers show the calculated inharmonicity coefficients for the optimal mappings: PP = G(47), MS = '4':(A#, B); '8':(D#, E) for Nyström and PP = G#(48), MS = '2':(A, A#); '10':(F#, G) for Steinway. The dotted lines represent the inharmonicity coefficients B calculated for the standard mapping. The difference between the original and standard mapping reaches an order of magnitude.

The details of presented before two selected monotonic surjective mappings, PP = G(47), MS = '4':(A#, B); '8':(D#, E) and PP = G#(48), MS = '2':(A, A#); '10':(F#, G), are explained in Fig. 7. In this figure, unlike in Figs. 1 and 2, the slope of the original 12-TET scale has been subtracted. As a result, the original tuning appears as a horizontal line, and the plot shows deviations from that tuning. This plot serves as a practical reference for piano tuners performing the transition from 12-TET to 10-TET, as it directly indicates how many cents each key must be retuned from the original 12-TET tuning. The inset presents zoomed main plot for one octave around PPs G(47) in case of Nyström and G#(48) for Steinway. For this points the frequency difference is by definition zero. The jumps between 43rd, 44th and 50th, 51st keys for Nyström and between 46th, 47th and 49th, 50th key for Steinway are fingerprints of the presented signatures.

To find the optimal mapping, one must establish suitable criteria for ranking mappings from the most to the least optimal. As mentioned earlier, multiple parameters have conflicting requirements. For the pinblock, where tuning pins are placed, lower total tension is preferable. Similarly, individual pins benefit from reduced tension. From a timbre perspective, a lower
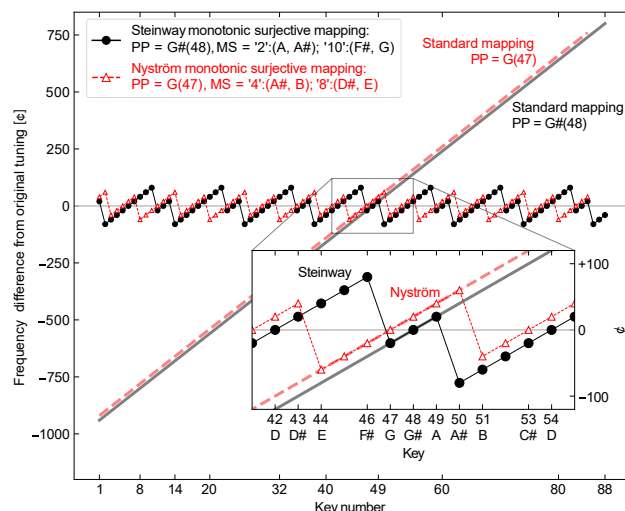
Fig. 7. PP and MS. Detuning from oryginal 12-TET scale expressed in cents for all keys of keyboard. The gray solid line at 0 ¢ represents oryginal tuning in 12-TET scale. The black and red markers show selected monotonic surjective mappings. The rising diagonal thick black and red dashed curves are result of standard mappings from the same PPs as those for optimal mapping. The slope of of diagonal lines is a difference between 12-TETs slope (100 ¢ per key) and 10-TETs slope (120 ¢ per key) resulting in 20 ¢ per step.

inharmonicity coefficient is desirable; however, the inharmonicity coefficient is inversely proportional to tension. Moreover, reducing inharmonicity below its original value is not always the goal, as preserving the original timbre may be a higher priority. This consideration is reflected in the maximum change of the OP criterion. Furthermore, starting from different PP results in varying numbers of shared frequencies (notes) with the original scale, which may be important when performing with other instruments.

Ultimately, the criteria for selecting an optimal mapping are subjective. The only truly objective requirement is that each string's tension must remain below its ultimate breaking point. A nearly objective criterion is ensuring tuning stability, but the literature does not define a single reference value for this. Tuning stability depends on factors such as the instrument's materials, environmental conditions (temperature, humidity), and playing intensity. A commonly used guideline is to limit frequency changes for each string to no more than 100 ¢.

Table 5 presents the results of calculations for standard mappings with PPs C(40) and A(49). These mappings were evaluated based on their impact on total string tension and OPs, which measure the string tension as a percentage of the breaking force. The alternative tension values (T_{alt}) and the changes in tension (ΔT) are listed alongside the maximum and minimum OPs for each case.

The results clearly demonstrate that standard mappings, regardless of the chosen PP, introduce severe

Table 5. Summary of tension values for standard mappings with PPs from C(40) and A(49). The original tension is 161.684 kN for Nyström and 168.735 kN for Steinway. The original maximum $\mathrm{OP}_{\mathrm{orig}}^{\max}$ is 70 % (Nyström) and 54.5 % (Steinway). Similarly the original minimum $\mathrm{OP}_{\mathrm{orig}}^{\min}$ is 25.3 % for Nyström and 23.6 % for Steinway.

Instr.	PP	T_{alt} [kN]	ΔT [kN]	$\mathrm{OP}_{\mathrm{alt}}^{\max}$ [%]	$\mathrm{OP}_{\mathrm{alt}}^{\min}$ [%]
Nyström	A(49)	180.955	+19.271	155.7	8.9
Nyström	C(40)	222.781	+61.097	191.7	11.0
Steinway	A(49)	186.174	+17.438	134.3	11.0
Steinway	C(40)	229.207	+60.471	165.3	13.5

structural and acoustic issues. In every case, the maximum $\mathrm{OP}_{\mathrm{alt}}^{\max}$ significantly exceeds the instrument's limit (is over 100 %). Simultaneously, the minimum $\mathrm{OP}_{\mathrm{alt}}^{\min}$ drops drastically from original values of 25.3 % for Nyström and 23.6 % for Steinway, indicating that some strings become too loose to function properly, compromising pitch stability and timbre. The uneven distribution of tension, combined with the extreme departure from the instrument's original inharmonicity characteristics, results in an unbalanced tonal spectrum, rendering the instrument practically unplayable.

On the other hand, Table 6 summarizes the total string tension values for selected monotonic surjective mappings. The mappings listed here represent a subset of all possible monotonic surjective mappings, chosen based on specific PPs and mapping structures previously presented in Tables 2, 3, and 4 as examples of MSs. As shown, some of the presented mappings exhibit a total tension difference between the alternative and original tuning that is close to zero. It is worth noting that even in some cases total tension difference is relatively large (e.g., 21.444 kN) the maximum OP – $\mathrm{OP}_{\mathrm{alt}}^{\max}$ is well below breaking value and minimum OP value $\mathrm{OP}_{\mathrm{alt}}^{\min}$ is close to the original one.

While Table 6 provides a numerical overview of specific mappings, manually comparing each case is inefficient given the large number of possible mappings. A more effective approach is to visualize all computed mappings and sort them based on different criteria to identify the most favorable ones.

Figure 8 presents an extensive analysis of 792 possible monotonic surjective mappings, covering 12 different PP from PP = C(40) to PP = B(51). The plots show absolute total tension change $|\Delta T^{\mathrm{tot}}|$, signed total tension difference ΔT^{tot}, maximum and mean
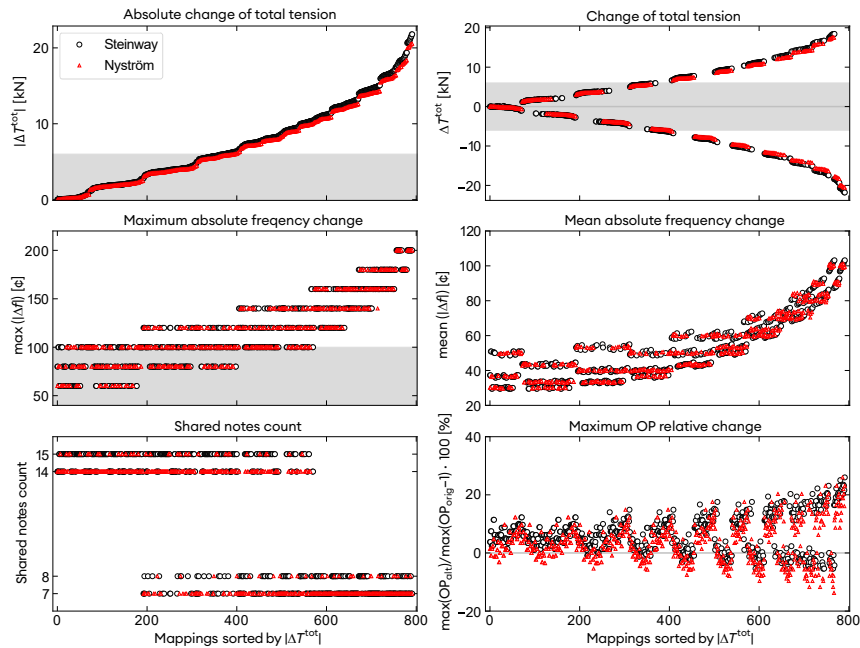
Fig. 8. Criteria for selecting an optimal mapping.

Table 6. Summary of tension values for selected monotonic surjective mappings. The original tension is 161.684 kN for Nyström and 168.735 kN for Steinway. The original maximum $\mathrm{OP}_{\mathrm{orig}}^{\max}$ is 70 % (Nyström) and 54.5 % (Steinway). Similarly the original minimum $\mathrm{OP}_{\mathrm{orig}}^{\min}$ is 25.3 % for Nyström and 23.6 % for Steinway.

Instr.	PP	MS	T_{alt} [kN]	ΔT [kN]	$\mathrm{OP}_{\mathrm{alt}}^{\max}$ [%]	$\mathrm{OP}_{\mathrm{alt}}^{\min}$ [%]
Nyström	G(47)	no. 31: '4':(A#, B); '8':(D#, E)	161.684	−0.000	71.0	24.7
Steinway	G#(48)	no. 18: '2':(A, A#); '10':(F#, G)	168.730	−0.006	56.6	25.3
Steinway	C(40)	no. 34: '1':(A#, B, C)	165.560	−3.175	58.6	23.0
Steinway	C(40)	no. 55: '1':(B, C, C#)	169.724	+0.988	59.1	23.0
Steinway	F#(46)	no. 65: '1':(A#, B, C)	190.179	+21.444	67.3	26.5

absolute frequency change, the number of shared notes with the original tuning, and the relative change in the maximum OP. By sorting mappings based on $|\Delta T^{\mathrm{tot}}|$ (top-left panel), it becomes evident that mappings optimized for one criterion are not necessarily optimal for others. Additionally, mappings for different instruments (Steinway and Nyström) do not overlap, suggesting that a universally optimal mapping does not exist across different instruments.

A key finding of this analysis is that all 792 examined monotonic surjective mappings are structurally safe. This means that none of the mappings exceed the maximum breaking tension for any string. Moreover, as shown in Fig. 8, particularly in the panel displaying the maximum absolute frequency change, no string undergoes a retuning greater than 200 ¢, ensuring that all modifications remain within a reasonable tuning range. The plot in Fig. 8 presents 792 possible monotonic surjective mappings for Steinway (black circles) and Nyström (red triangles), calculated across 12 PP ranging from PP = C(40) to PP = B(51), resulting in a total of $12 \times 66 = 792$ mappings. Each panel corresponds to a different ranking criterion: absolute total tension change $|\Delta T^{\mathrm{tot}}|$, total tension difference ΔT^{tot}, maximum absolute frequency change, mean absolute frequency change, number of shared notes between the original and mapped tuning, and the relative change in the maximum OP. The mappings are sorted by $|\Delta T^{\mathrm{tot}}|$, as shown in the top-left panel. The results demonstrate

that mappings optimized for one criterion (±100 ¢ or ±6000 kN, shown as gray regions) are not necessarily optimal for others.

We propose a useful tool for selecting an optimal monotonic surjective mapping – a plot that visualizes mapping-related scalar values using a color scale, with PP on the vertical axis and MSs on the horizontal axis. Black dots indicate mappings that fall below an arbitrarily chosen threshold, while the top 20 mappings are ranked and labeled with red numbers.

This type of chart is particularly useful in several ways. First, it helps identify which keys are doubled in a given mapping. To interpret the chart, one starts by selecting a specific field in the 2D map, then reads the corresponding PP on the left and the MS at the bottom. For example, in Fig. 9, which presents total tension differences for Steinway, the best mapping according to this criterion is marked with a red '1'. This corresponds to PP = G# on the left and mapping number 18, which yields the alternative scale step sequence —1, 2, 2, 3, 4, 5, 6, 7, 8, 9, 10, 10—. This means that, starting from the chosen PP key, the sequence follows the structure presented in Table 11a, with repeated key pairs **A, A#** and **F#, G**. This mapping can be expressed as PP = G#(48), MS = '2':(A, A#); '10':(F#, G). Similarly for Nyström with this method we identify mapping PP = G(47), MS = '4':(A#, B); '8':(D#, E) as optimal from the point of view of total tension.
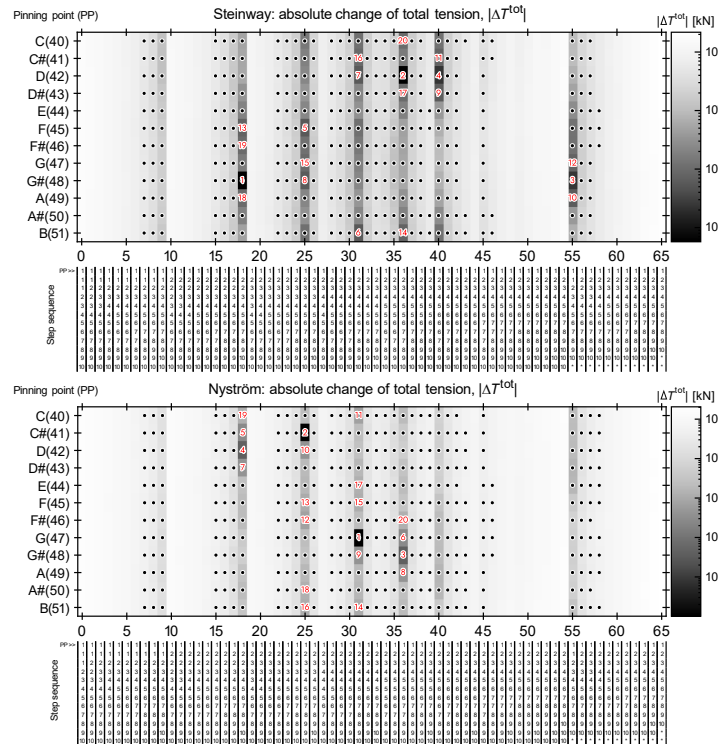
Fig. 9. Absolute total tension change for: a) Steinway; b) Nyström. The color scale represents the absolute total tension change $|\Delta T^{\mathrm{tot}}|$ in kN, using a logarithmic color scale. The black dots indicate mappings where $|\Delta T^{\mathrm{tot}}|$ is below the threshold of 6000 kN. The top 20 mappings are ranked and labeled with red numbers.
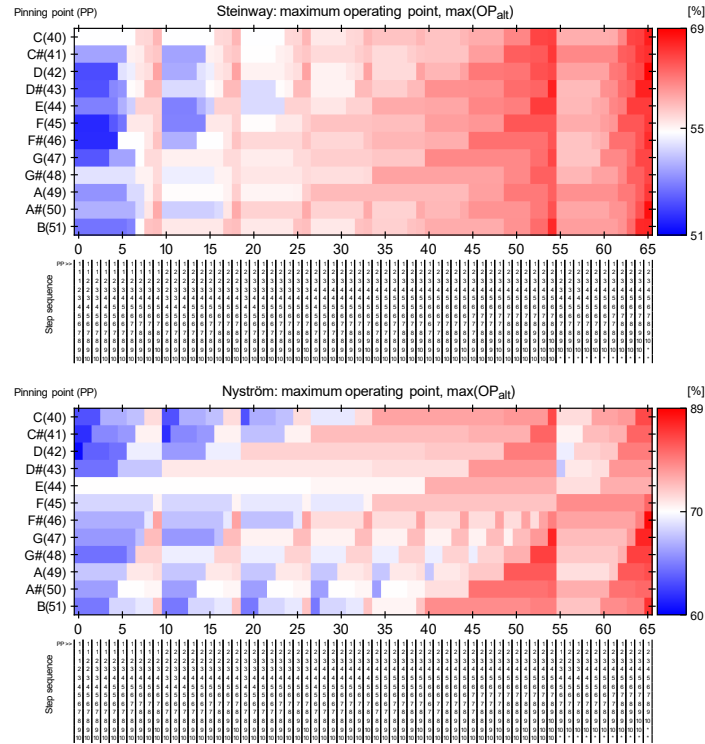
Fig. 10. Maximum operating point for: a) Steinway; b) Nyström. The color scale represents the maximum operating point – $\text{OP}_{\text{alt}}^{\max}$, where white color is assigned to maximum value of original operating point – $\text{OP}_{\text{orig}}^{\max}$.

While all analyzed mappings are structurally safe and do not compromise the instrument, we can further refine our selection by identifying the most optimal ones based on additional criteria. Since total tension remains within acceptable limits and does not threaten the integrity of the pinblock (Fig. 9), we can instead prioritize mappings that minimize changes in the maximum or minimum OPs, as these influence inharmonicity (Fig. 10). Here, we present selected examples, but the final choice of mapping should always be based on a similar analysis tailored to each individual instrument. For this reason, we do not provide 2D visualizations for all possible parameters.

6. Generalization for other fixed-pitch string instruments

The methodology for identifying monotonic surjective mappings, as demonstrated with pianos, can be extended to other fixed-pitch string instruments, such as harps, celestas, cimbaloms, harpsichords, clavichords, lyres, dulcimers, kanteles, spinets, psalteries, virginals, and zithers. In these applications, the traditional octave (2:1 frequency ratio) can be replaced by alternative interval structures, such as the tritave (3:1 ratio) found in the Bohlen–Pierce scale (MATHEWS *et al.*, 1988) or the hyperoctave (4:1 ratio) used in experimental tuning systems, including some hyperpiano concepts (HOBBY, SETHARES, 2016). The interval being divided, sometimes called an equave, can be split into unequal steps. Therefore, the approach presented in this paper provides a systematic method for determining safe retuning mappings from an arbitrary N-step scale to a smaller M-step scale, provided that $M < N$.

An example of a non-octave-based system is the set of tuning scales proposed by Wendy Carlos, known as the alpha, beta, and gamma scales. These systems divide the perfect fifth, rather than the octave, into 9, 11, and 20 equal parts, respectively. While our method is not directly applicable to retuning a 12-TET instrument to one of these scales, because the basic periodic unit is not the octave, it becomes fully applicable when retuning between Carlos-type scales. For instance, an instrument tuned in gamma ($N = 20$) can be retuned safely to beta ($M = 11$) or alpha ($M = 9$), and beta can be retuned to alpha. These transformations preserve the periodicity of the fifth and meet all assumptions of the monotonic surjective mapping framework. In this way, retuning between scales with different step counts becomes feasible even when the equave is not the octave but another interval, such as the perfect fifth in Carlos-type tunings.

To illustrate the applicability of the method in a more traditional octave-based context, we turn to the example of a folk diatonic harp (not to be confused with a harmonica). This instrument traditionally follows a heptatonic (seven-note) scale (original scale) consisting of the notes C, D, E, F, G, A, and B, mean-

ing $N = 7$. If the goal is to adapt the instrument to a pentatonic (five-note) scale (alternative scale) consisting of C, D, E, G, and A ($M = 5$), then the number of possible monotonic surjective mappings for a given PP can be determined using Eq. (1), yielding 21 potential mappings. Similar to the fixed keyboard pattern of the piano, diatonic harps also incorporate a structured system to indicate pitch relationships. In this case, strings are color-coded, with C strings typically colored red and F strings colored blue or black. This provides a visual reference for the player and reinforces the structural role of the octave. This suggests that the assumption of octave preservation remains valid in this case. Consequently, when applying monotonic surjective mappings to the diatonic harp, maintaining octave equivalence remains a natural and practical choice. The remainder of the analysis would proceed similarly to that of the piano, requiring an evaluation of breaking tensions and other parameters, such as speaking length, to ensure both structural integrity and tonal stability in the alternative tuning.

7. Conclusions – Final selection of mapping

The proposed method of monotonic surjective mappings provides a systematic way to retune a 12-TET instrument to the alternative 10-TET scale while preserving the original frequency range and ensuring that no string exceeds its breaking tension. Our computational analysis confirms that every mapping generated by this method is structurally safe. Nonetheless, various factors – such as absolute frequency changes, total tension differences, and the number of shared notes between the original and alternative tunings – must be taken into account when selecting the optimal mapping.

The final choice of mapping ultimately depends on personal preference and performance context. One important consequence of the monotonic surjective mapping approach is that some neighboring keys on the piano keyboard may be assigned the same pitch. This feature allows for a rapid, percussive repetitions of the same sound during performance, which may influence an artist's preference for a particular mapping. While many mappings can be optimized to minimize tension differences or frequency deviations, the selection remains subjective. For instance, in a somewhat humorous twist, the professional pianist Leszek Możdżer chose not to adopt the mapping with the smallest overall tension change. Instead, he preferred a different signature that maintained symmetry in the keyboard layout. His decision was to repeat the keys F and F# as well as A and A#, while sharing the note A between the original and alternative scales. In his case, his selection corresponds to mapping no. 7 with PP = A(49), where the MS is given by: MS = '1':(A, A#); '8':(F, F#).

In conclusion, the monotonic surjective mapping approach offers a viable solution for retuning acoustic instruments to alternative scales. Although the method ensures structural safety, the optimal mapping must be chosen by balancing technical criteria with the musician's individual artistic taste. Moreover, the concept can be generalized beyond pianos to other fixed-pitch string instruments, such as harps, harpsichords, and dulcimers, thereby broadening its potential applications in diverse musical contexts.

Appendix – Why 10-TET is a unique choice

In the main text, we have focused on dividing the octave into equal steps (TET, EDO). However, there is also broad interest in other scales (temperaments), seeking to resolve the small comma mismatch that arises when combining integer frequency ratios (PILCH, TOPOROWSKI, 2014; RASCH, 1984).

The Pythagorean tuning, for example, cycles through perfect fifths (frequency ratio 3:2) while treating octave-equivalent tones (2:1) as musically identical. This produces pure, consonant fifths but also leads to the famous Pythagorean comma, which quantifies the mismatch between twelve stacked just fifths and seven octaves.

The Pythagorean comma is typically defined as the cumulative discrepancy between twelve just fifths and seven octaves:

$$\frac{(3/2)^{12}}{2^7} = \frac{531441}{524288} \approx 1.01364,$$

which corresponds to approximately 23.46 cents. This classical definition is dimensionless and expresses the ratio between the two paths through pitch space.

In this work, however, we use a normalized (per-octave) version of the comma:

$$c_{(3,12,7)} = \left|(3/2)^{12/7} - 2\right| \approx 0.00388,$$

where 3 is the frequency ratio of a perfect fifth, 12 is the number of such intervals stacked, and 7 is the number of octaves they are expected to span. By raising the full stack to the power $1/7$, we calculate the average interval needed per octave. The result is then compared to the exact doubling of frequency (ratio 2) expected for one octave. This way, the value $c_{(3,12,7)}$ captures how much the average fifth-based step deviates from the ideal octave size, making it easier to generalize and compare different (t, s) scale configurations.

Although small, this discrepancy confirms that the cycle does not close perfectly, and the frequency does not return exactly to the starting note.

General Pythagorean t-step scales

A more general version of Pythagorean tuning picks an odd harmonic integer denoted n_k and an integer

m_0 that 'folds' the n_k-based interval into the base octave, ensuring the resulting frequency ratio remains within $[1, 2]$:

$$h_1 = \frac{n_k}{2^{m_0}}, \quad \text{with} \quad 1 < h_1 < 2.$$

Repeatedly multiplying (or dividing) by h_1, and dividing (or multiplying) by 2 whenever the resulting frequency lies outside $[1, 2]$, generates t steps within each octave. If this process nearly reconstructs the octave after t steps across s octaves, the mismatch – called the comma – can be written as:

$$c = \left|(h_1)^{t/s} - 2\right| < \epsilon,$$

where t is the number of steps per octave, s is the number of octaves required to close the cycle, and ϵ is a small upper bound for acceptable mistuning. If the comma is smaller than the Pythagorean $c_{(3,12,7)}$, then the triple (n_k, t, s) is said to define a good comma. This indicates that the generated scale closely approximates octave closure while using a consistent generator interval.

We restrict attention to manageable scales by requiring $t \cdot s < 500$, which limits the total number of notes (i.e., the number of steps per octave times the number of octaves needed to complete the cycle), and limit $n_k \le 21$ to avoid using excessively high harmonics.

Under these constraints, one finds that the smallest comma c – i.e., the best result among all possibilities within these limits – arises for $(n_k, t, s) = (13, 10, 7)$, giving a generating interval $h_1 = 13/8$, often called a 'neutral sixth'. Since here $t = 10$, this results in a 10-step Pythagorean scale whose comma

$$c_{(13,10,7)} \approx 0.00087$$

is more than four times smaller than that of the usual 12-tone Pythagorean system $(3, 12, 7)$, with comma $c_{(3,12,7)}$.

Angles on a musical circle

An insightful way to handle such cyclic issues is to place each frequency ratio on a circle of angles. Specifically, if h_ℓ is the frequency multiplier for the ℓ-th step (where $\ell = 0, 1, \ldots, t - 1$), we define the Pythagorean angle by

$$\varphi_\ell = 2\pi \log_2(h_\ell).$$

Because multiplying a frequency by 2 shifts its angle by 2π (one full turn), the angle φ_ℓ neatly captures where h_ℓ lies 'modulo octaves'. In that sense, going once around the circle corresponds to going up by one full octave in frequency.

Comparing with equal temperament

Another way to assess how well a Pythagorean scale approximates its equal-tempered counterpart is the tempered index:

$$\delta_{(n_k,t,s)} = \frac{1}{t-1} \sum_{\ell=0}^{t-1} \left|\Phi_\ell - \varphi_\ell\right|,$$

where $\varphi_\ell = 2\pi \log_2(h_\ell)$ and $\Phi_\ell = 2\pi \frac{s}{t}\ell$ are the equally tempered angles for t notes in each of s octaves. Numerically,

$$\delta_{(3,12,7)} \approx 6.40 \ \text{¢/step}$$

and

$$\delta_{(13,10,7)} \approx 1.46 \ \text{¢/step},$$

where radians per step were recalculated to cents per step.

Hence, from a purely mathematical viewpoint, the 10-TET scale is, at the same time, almost perfectly the 10-step Pythagorean scale built using the 13:8 ratio interval. Moreover, this 10-step Pythagorean scale has the smallest comma c among all general Pythagorean scales calculated with reasonable assumptions.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. However, part of the research described in this article is the subject of a patent application

Authors' contribution

Aleksander Bogucki developed the method, prepared figures, and wrote the manuscript. Andrzej Włodarczyk conducted string measurements and provided technical validation. Paweł Nurowski initiated and coordinated the project, developed the mathematical framework for 10-TET scale. All authors reviewed and approved the final manuscript.

References

1. Bandy C. (2025), *A study in reharmonization: Jacob Collier's subversive vocal arrangements of popu-*

lar tunes, Ph.D. Thesis, Kent State University, United States, Ohio.

2. BUCUR V. (2016), *Handbook of Materials for String Musical Instruments*, Springer Cham, https://doi.org/10.1007/978-3-319-32080-9.

3. FLETCHER H. (1964), Normal vibration frequencies of a stiff piano string, *The Journal of the Acoustical Society of America*, **36**(1): 203–209, https://doi.org/10.1121/1.1918933.

4. FRASER J. (2023), Microtonality in a Western setting, [in:] *Music Talks: Proceedings of the Western Sydney University Undergraduate Musicology Conference 2022*, Williams J., Stead M., Leonard J. [Eds.], Western Sydney University, https://doi.org/10.26183/X0RY-NF79.

5. HIDEYA A. (2021), *Like Spring Sea*, 10 tone equal temperament song (10-TET 10EDO), https://www.schwaben.jp/ongaku08.html.

6. HOBBY K. SETHARES W.A. (2016), Inharmonic strings and the hyperpiano, *Applied Acoustics*, **114**: 317–327, https://doi.org/10.1016/j.apacoust.2016.07.029.

7. HUNT A.A. (2021), *Foreword to The Equal-Tempered Keyboard*, https://zwillinge.zentral.zone/vorschau/etk_frontmatter.pdf.

8. HUNT A.A. (2022), *The Equal-Tempered Keyboard* [digital album], (recorded in 2021), https://aaronandrewhunt.bandcamp.com/album/the-equal-tempered-keyboard.

9. LOUCHET J. (2021), *The Keyboard Stringing Guide for the Restoration of Pianos, Harpsichords and Clavichords*.

10. MARTIN G. (2025), The best-selling Nintendo franchises of all time, *Endlessmode*, https://www.paste magazine.com/games/nintendo/best-selling-nintendo-franchises (access: 12.06.2025).

11. MATHEWS M.V., PIERCE J.R., REEVES A., ROBERTS L.A. (1988), Theoretical and experimental explorations of the Bohlen–Pierce scale, *The Journal of the Acoustical Society of America*, **84**(4): 1214–1222, https://doi.org/10.1121/1.396622.

12. MATTHIAS M. (1990), *Steinway Service Manual: Guide to the Care and Maintenance of a Steinway*, E. Bochinsky, Germany, https://openlibrary.org/books/OL182 38258M/Steinway_service_manual.

13. MENDEL A. (1949), Devices for transposition in the organ before 1600, *Acta Musicologica*, **21**: 24–40, https://doi.org/10.2307/931533.

14. MOŻDŻER L., DANIELSSON L., FRESCO Z. (2024), *Beamo* [CD], Teraz Teraz, https://www.discogs.com/release/33164979-Leszek-Mo%C5%BCd%C5%Bcer-Lars-Danielsson-3-Zohar-Fresco-Beamo.

15. MOŻDŻER L., DANIELSSON L., FRESCO Z. (2025), *Beamo* [CD], ACT, https://www.prestomusic.com/jazz/products/9752880–beamo.

16. PILCH M., TOPOROWSKI M. (2014), *Former Temperations: Acoustic Basics and Practical Use* [in Polish: *Dawne Temperacje: Podstawy Akustyczne i Praktyczne Wykorzystanie*], Akademia Muzyczna im. Karola Szymanowskiego, Frodo, Poland.

17. RASCH R.A. (1984), Approaches to tuning and temperament, *The Journal of the Acoustical Society of America*, **76**(S1): S40–S41, https://doi.org/10.1121/1.2021849.

18. ROSE M., LAW D. (2001), *A Handbook of Historical Stringing Practice of Early Keyboard Instruments*, Vogel & Scheer.

19. SENPAI Z. (2023), *Chromatic Etude for 10 EDO Piano No. 1 "Intervals"*, https://scorefol.io/w/cllchoxyp 000kl90f15j182bm (access: 5.03.2025).

20. SETHARES W.A. (1998), *Tuning, Timbre, Spectrum, Scale*, Springer-Verlag London, https://doi.org/10.1007/978-1-4471-4177-8_12.

21. SEVISH (2017), *Vidya*, [on:] Harmony hacker [digital album], https://sevish.bandcamp.com/track/vidya.

22. Stahl- und Drahtwerk Röslau GmbH (n.d.), Steel wire for strings "Röslau Blue Label", https://www.roeslau-draht.com/fileadmin/user_upload/PDF/Datenblaetter/EN/Important_parameters_blau.pdf (access: 2.03.2025).

23. STANEVIČIŪTĖ R., JANICKA-SŁYSZ M. (2022), *Music and Change in the Eastern Baltics Before and After 1989*, Academic Studies Press, https://doi.org/10.2307/j.ctv2xszr6z.

24. THOMAS P.E. (1996), *Music for one or more alternately tuned acoustic pianos, 1920–1993: Trends in melody, harmony, and technique*, Ph.D. Thesis, University of Cincinnati, United States.

25. TOMALA L. (2024), Octave divided into 10 parts. Scientists prepare decaphonic piano for jazz legend Leszek Możdżer, *Science in Poland*, https://sciencein poland.pl/en/news/news%2C100700%2Coctave-divided-10-parts-scientists-prepare-decaphonic-piano-jazz-legend-leszek (access: 5.03.2025).

26. Xenharmonic Wiki (n.d.), 10edo/Music, https://en.xen.wiki/w/10edo/Music (access: 5.03.2025).

Research Paper

Sensitivity Analysis of Acoustic Parameters in a Theatre Hall: A Case Study of the Maria Zankovetska Theatre in Lviv

Jarosław RUBACHA[(1)*], Adam PILCH[(1)], Roman KINASZ[(2)], Wojciech BINEK[(1)],
Tadeusz KAMISIŃSKI[(1)], Mykhaylo MELNYK[(3)]

[(1)] *Department of Mechanics and Vibroacoustics, Faculty of Mechanical Engineering and Robotics*
AGH University of Krakow
Kraków, Poland

[(2)] *Department of Geomechanics, Civil Engineering and Geotechnics*
Faculty of Civil Engineering and Resource Management, AGH University of Krakow
Kraków, Poland

[(3)] *Department of Computer Aided Design Systems, Lviv Polytechnic National University*
Lviv, Ukraine

[*]Corresponding Author e-mail: jrubacha@agh.edu.pl

This article explores the potential for modifying acoustic parameters within a theatre hall, using the Maria Zankovetska Theatre in Lviv, as a case study. The study used sensitivity analysis to evaluate how changes in the sound absorption of specific surfaces affect selected acoustic parameters in the horseshoe-shaped theatre hall. A numerical model of the hall is developed and calibrated based on in-situ acoustic measurements to assess the sensitivity of parameters such as reverberation time (T_{20}), early decay time (EDT), clarity (C_{50}), and early sound strength (G_{80}).

The analysis reveals that surfaces such as the stage tower ceiling, stage walls with curtains, audience walls, and seating have the most significant impact on the acoustic parameters. Modifying the sound absorption of these surfaces can affect T_{20}, EDT, C_{50}, and G_{80}. Notably, increasing the absorption of a single surface might not significantly alter C_{50} and G_{80} values, whereas reducing the absorption of surfaces such as seating can lead to noticeable changes in these parameters. These findings provide valuable insights for future renovations and acoustic adjustments, aiming to optimize the theatre's acoustic performance while preserving its historical character.

Keywords: room acoustics; sensitivity analysis; sound absorption coefficient; reverberation time.

1. Introduction

The aim of this article is to investigate the potential for adjusting acoustic parameters in a horseshoe-shaped theatre hall, using the Maria Zankovetska Theatre in Lviv as a case study. The theatre, constructed between 1837 and 1842, stands as a testament to the architectural vision of Ludwig Pichla and Johann Zalcman. Designed with an uncompromising architectural layout, it became one of the largest theatre in Europe, serving as a pivotal hub for artists and cultural institutions. From its inception, the theatre has been the epicentre of cultural life in eastern Galicia, boasting an audience capacity of approximately 1460 seats distributed across the ground floor, side boxes, and four balconies. The theatre has undergone significant changes throughout its history. Between 1941 and 1944, construction damage caused by pile foundation problems required a partial rebuilding of the structure. As a result of this reconstruction, the original balconies

and side boxes were replaced with two amphitheatrical balconies, which greatly altered the theatre's interior. In 2017, the theatre underwent a major modernization, which included replacing the flooring in the audience area and installing new seats. The previous seats, heavily upholstered on the bottom and back, had been the primary sound-absorbing elements in the hall since the 1940s rebuild. This resulted in an excessively long reverberation time that was unsuitable for theatrical performances. To address this issue, the decision was made to install new seats of a similar construction to those introduced in the 1940s, aiming to prevent any further increase in reverberation time.

Horseshoe-shaped halls provide good stage visibility and acoustic proximity for the audience, but exhibit acoustic limitations. The construction of multi-tiered boxes, with rich ornamentation and carvings, improves sound dispersion, but at the same time increases acoustic absorption, thereby reducing the reverberation time. The audience is primarily reached via direct sound, while a scarcity of reflections from the side walls reduces the impression of spaciousness. As previous studies have shown, listening conditions are worse in the depths of the boxes due to sound screening by balustrades and attenuation by walls. This problem has been addressed by designing shallow galleries and balconies without dividing the lodges (Barron, 2009) or by using Schroeder diffusers on the rear walls of sub-balcony lodges (Kamisiński, 2012).

Analyzing the acoustic parameters of historic theatres requires both in-situ measurements and numerical model studies (Kamisiński, 2010). In order to develop a numerical model of a theatre hall, it is necessary to determine the surface parameters inside the hall for proper calibration (Pilch, 2020). The basic parameter is the sound absorption coefficient of finishing materials (Prodi, Pompoli, 2016; Rubacha et al., 2019) and auditorium seats (Beranek, Hidaka, 1998; Rubacha et al., 2012). It is also important to consider sound diffusion (Binek et al., 2022; Pilch, 2021; Shtrepi, 2019) as well as sound-reflecting elements (Szeląg et al., 2014; 2020) in the model. Such elements help to control early reflections and eliminate acoustic defects.

The research conducted encompasses the following key areas. Firstly, the preparation of a numerical model of the theatre and its calibration based on acoustic measurements. This involved creating a digital representation of the theatre's interior that accounted for its architectural features and materials. Calibration was conducted by comparing the model's predictions with measured acoustic parameters collected within the theatre, thereby ensuring accuracy and reliability for further analysis. Secondly, the determination of the values of selected acoustic parameters of the interior under conditions of variable surface absorption using the model. By manipulating the absorption coefficients of various surfaces within the theatre, such as walls, ceilings, and seating, we could observe changes in key acoustic parameters, including reverberation time (T_{20}), clarity (C_{50}), and sound strength (G_{80}). This step was necessary for conducting a sensitivity analysis. The sensitivity analysis was conducted to examine the relationship between the acoustic absorption of selected surfaces and the acoustic parameters, identifying the surfaces that have the greatest impact on shaping these acoustic parameters. This analysis systematically varied the absorption properties of individual surfaces to observe the resulting changes in acoustic parameters. The findings highlighted which surfaces are most critical in influencing the theatre's acoustics, providing valuable insights for future renovations and acoustic improvements.

Through this study, we aim to gain a deeper understanding of the acoustics of horseshoe theatres and analyse the possibility of shaping their acoustic parameters while preserving their historical character.

2. Sensitivity analysis

2.1. Local and global sensitivity analysis

The purpose of sensitivity analysis is to examine the relationship between the input variables x of a model and the output variables y, where $y = g(x)$ and g is the model that maps inputs to outputs (Borgonovo, Plischke, 2016). Two primary techniques for sensitivity analysis are distinguished: local and global. Local sensitivity analysis involves changing the model parameters around specific reference values to determine how small variations in the inputs affect the model's response. According to derivative-based sensitivity analysis (Borgonovo, 2008; Pianosi et al., 2016), the output sensitivity index S_i of the i-th input factor x_i can be calculated using the partial derivative $\frac{\sigma_y}{\sigma_{x_i}}$ evaluated at the nominal value $\overline{x}$ of the input vector $(x_1, x_2, ...)$. The goal of derivative-based sensitivity analysis is to identify which parameters have the most significant impact on the model's outcomes and to understand the nature of that impact.

The advantages of local sensitivity analysis are its ease of use and low computational requirements. For this reason, this approach is widely used in the literature; however, it also has significant limitations (Saltelli, Annoni, 2010). If the model is nonlinear, the results of local sensitivity analysis can be highly biased, as it assumes independence among model's input variables (Tang et al., 2007). If the model inputs are not independent (i.e., when they interact with each other), local sensitivity analysis will underestimate their importance as it does not account for the effects of mutual interaction (Hamm et al., 2006). In such a case, global sensitivity analysis is applicable.

2.2. Regression-based methods

Regression-based sensitivity analysis is a statistical method used to evaluate the impact of changes in the values of independent variables on the outcome of dependent variables in a regression model (Iooss, Lemaître, 2015; Manache, Melching, 2008; Saltelli *et al.*, 2004). This method is particularly useful for identifying the factors that have the greatest influence on the model and for supporting informed decision-making.

The primary step in regression-based sensitivity analysis is the computation of the sensitivity coefficient for each independent variable. This coefficient measures how much a change in a given variable affects the model's outcome. Thereby, the value of this coefficient indicates whether that variable has a strong or weak impact on the model.

One of the main sensitivity indicators in this category is the standardized regression coefficient (SRC). To calculate the SRC, a regression model between the input vector x_i and the output variable y is fitted using the least squares method:

$$y = b_0 + \sum_{i=1}^{N} b_i x_i, \qquad (1)$$

where b_0 and b_i are the regression coefficients corresponding to the i-th input variable of the model. Equation (2) can be used to calculate the SRCs for different input values:

$$S_i = b_i \frac{\sigma_{x_i}}{\sigma_y}, \qquad (2)$$

where σ_{x_i} is the standard deviation of the i-th model input and σ_y is the standard deviation of the model output.

Regression-based sensitivity analysis methods are global in nature and can examine the entire input space for variations. However, their actual level of comprehensiveness depends on the experimental design and the number of simulations providing data to establish the regression relationships. Although these methods are generally computationally efficient, they do not provide significant information on parametric interactions.

In the present study, sensitivity analysis methods based on regression were applied. The aim was to assess the impact of changes in the acoustic parameters of the Maria Zankovetska Theatre hall in Lviv on the changes in the sound-absorbing properties of individual surfaces within the hall. Based on the model results, inferences were drawn, and the surfaces and groups of surfaces exhibiting the greatest impact on the interior acoustics were identified.

In the studied model, the input parameters were the sound absorption coefficients of individual surfaces and groups of surfaces. The surfaces were grouped based on the type of material used for finishing and the location of the surfaces within the hall. Overall, the surfaces were divided into 14 groups, distinguishing between their location on the stage and in the audience area.

3. Analysis of acoustic parameters in the theatre hall

3.1. Description of the theatre hall

The Maria Zankovetska Theatre accommodates a total of 799 seats, with 531 seats located on the ground floor and an additional 268 seats distributed across two balconies. The main hall, encompassing approximately $5400\,\text{m}^3$, is connected to an $8000\,\text{m}^3$ scenic box, providing ample space for stage setups and scenery changes. For detailed parameters of the theatre hall, see Figs. 1 and 2, and Table 1.

Table 1. Parameters of the Maria Zankovetska Theatre in Lviv.

No.	Parameter	Value
1	Total volume V [m^3]	13 400
2	Audience hall volume V_w [m^3]	5400
3	Stage volume V_{sc} [m^3]	8000
4	Orchestra pit volume V_{or} [m^3]	80
5	Stage surface area S_{sc} [m^2]	410
6	Orchestra pit surface area S_{or} [m^2]	40
7	Total number of seats N	799
8	Number of seats on the ground floor N_p	531
9	Number of seats in balconies N_l	268
10	Audience volume per person [m^3/person]	5.63

3.2. Measurements of acoustic parameters in the theatre hall

The theatre's acoustic parameters were selected for analysis according to general literature recommendations (Barron, 2009) and International Standard ISO 3382-1 (International Organization for Standardization, 2009). The evaluation focused on early decay time (EDT), reverberation time (T_{20}), clarity (C_{50}), and early sound strength (G_{80}), with the curtain open and the stage empty. Impulse response measurements in the audience area were conducted according to the standard, using software for simultaneous recording from two microphones. All measurements were taken at a height of $1.2\,\text{m}$ above the floor surface, while omnidirectional sound sources were positioned at three locations on the stage at a height of $1.5\,\text{m}$. The location of the sound source was chosen to cover all characteristic points on the stage: proscenium near the axis, deep side stage, and deep stage near the axis of the hall (Fig. 1).

The analysis of impulse responses was performed using software to determine all necessary acoustic pa-
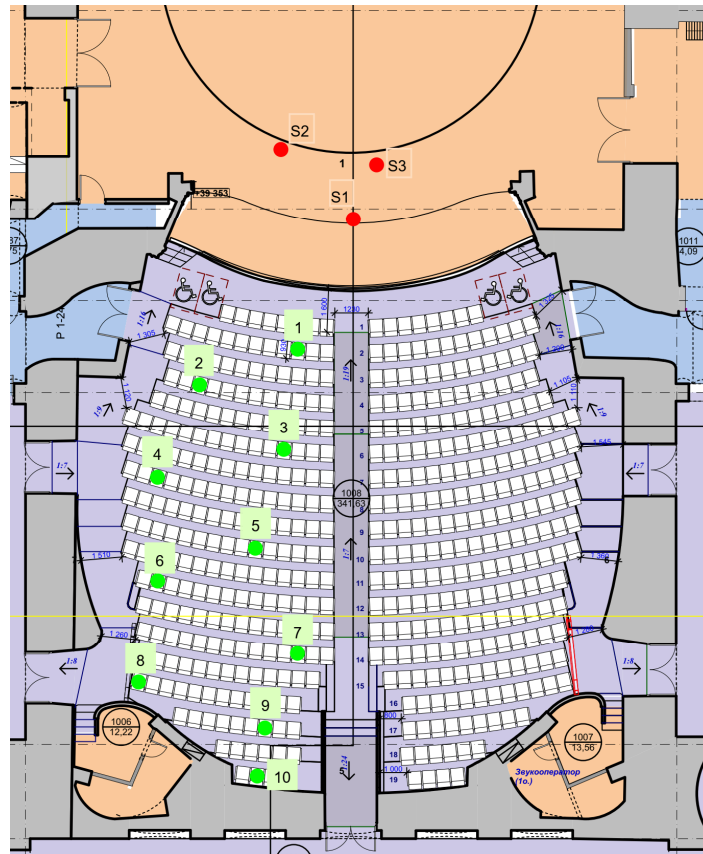
Fig. 1. Positions of the sound sources and microphones in the theatre's hall.
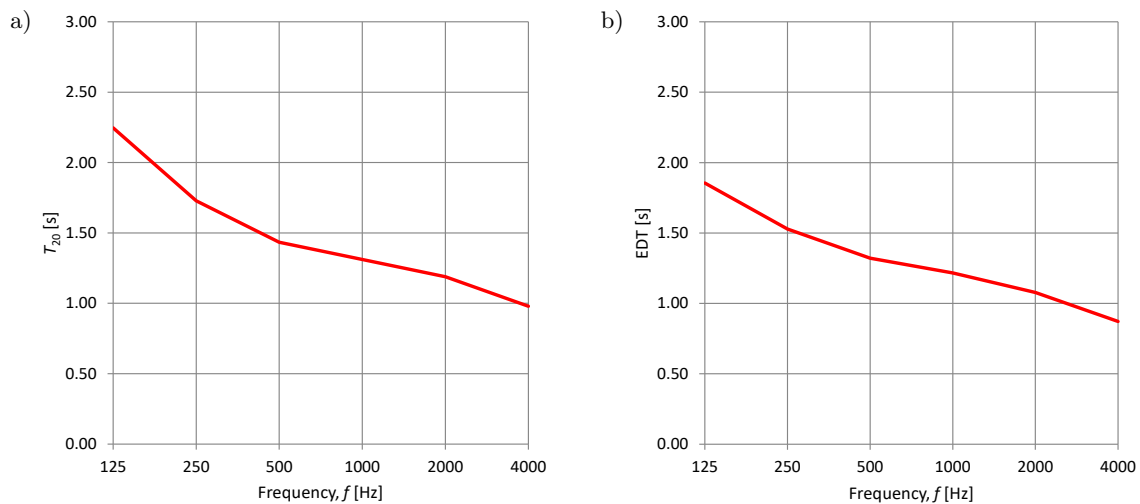
Fig. 2. Measured values of: a) T_{20} and b) EDT.

rameters. The determined values of T_{20} and EDT are shown in Fig. 2.

    The reverberation time characteristics, specifically T_{20} and EDT, are affected by the finishing materials used in the hall. The primary sound-absorbing element is the audience seating. An overhanging auditorium layout has been designed for the hall, featuring auditorium seats with medium-thickness upholstery mounted on stepped platforms. This arrangement allows for high

sound absorption coefficients in the mid- and high-frequency ranges. However, the absorption coefficient values at low frequencies are lower.

    Another significant sound-absorbing element is the stage decorations suspended in the stage box, which also provide high sound absorption in the mid and high-frequency ranges but low absorption at low frequencies. The walls, ceiling, and balcony balustrades are made of concrete, resulting in very low absorp-

tion coefficients across the entire frequency spectrum. Consequently, the hall lacks low-frequency sound-absorbing elements, leading to significantly higher reverberation times at 125 Hz and 250 Hz.

A more detailed analysis of the hall's reverberation was conducted based on the average EDT values measured at each point (see Table 2).

Table 2. Mean values of the EDT for the frequency ranges 500 Hz–1000 Hz.

	EDT_m [s]		
Point no.	Source position S1	Source position S2	Source position S3
1	1.80	1.86	1.67
2	1.45	1.90	1.64
3	1.38	1.66	1.56
4	1.15	1.63	1.46
5	1.16	1.32	1.16
6	1.19	1.24	1.19
7	1.17	1.34	1.21
8	1.01	1.09	1.12
9	1.01	0.98	1.04
10	0.97	1.12	1.22
Mean	1.23	1.41	1.33
RMSD	0.24	0.31	0.22

Analysis of the mean EDT values revealed that the highest values were recorded in areas closest to the stage, while the lowest values were found at the rear of the auditorium, particularly in the space beneath the balcony. This suggests that the stage area exhibits greater reverberation compared to the auditorium, primarily because it has a volume that is 50 % larger than that of the auditorium. Significant changes in EDT were also observed depending on the depth of the sound source on stage. For the sound source positions S2 and S3, located deep on the stage behind the proscenium, the EDT values were significantly higher compared to source position S1, located at the proscenium. This indicates that the acoustic parameters of the stage and hall differ and that the position of the sound source is considerably affected by the stage decorations.

Analysis of the C_{50}, parameter responsible for assessing music clarity was also carried out. The measured values of this acoustic parameter are presented for three sound source positions in Table 3.

The average value of the clarity index C_{50} is in the range of 1.3 dB–1.9 dB. A high root mean square deviation (RMSD) indicates significant variation in this parameter across different audience locations. Analysis of the measurements reveals that the lowest values occur at the centre of the hall, while higher values are found near the stage and at the rear wall. Overall, the measured values at the specific points are generally favourable for speech intelligibility.

Table 3. Mean values of the clarity index C_{50} for the frequency range 500 Hz–2000 Hz.

	C_{50m} [dB]		
Point no.	Source position S1	Source position S2	Source position S3
1	5.2	3.0	1.7
2	1.9	4.0	0.8
3	1.2	1.3	0.9
4	1.9	2.5	−0.5
5	0.1	−0.7	0.0
6	−0.5	0.7	0.3
7	1.0	1.6	1.8
8	0.0	0.4	2.0
9	3.4	2.2	3.6
10	3.2	3.8	2.2
Mean	1.7	1.9	1.3
RMSD	1.7	1.4	1.2

An analysis of the early G_{80} sound strength distribution is shown in Table 4.

Table 4. Mean values of the sound early strength G_{80} for the frequency range 500 Hz–1000 Hz.

	G_{80m} [dB]		
Point no.	Source position S1	Source position S2	Source position S3
1	6.8	1.8	2.0
2	4.1	2.9	0.4
3	5.6	0.8	0.6
4	4.4	−0.9	−1.5
5	4.7	0.6	2.7
6	1.5	−0.6	−0.4
7	3.8	1.2	2.2
8	2.3	0.1	1.1
9	2.9	1.4	1.9
10	2.1	0.4	−0.9
Mean	3.8	0.8	0.8
RMSD	1.6	1.1	1.3

The results indicate that the early G_{80} sound strength is highest near the sound source and decreases with distance. The location of the sound source significantly affects the G_{80} sound strength. When the sound source is positioned deep within the stage behind the proscenium, there is a notable reduction in sound strength of up to 5 dB at points closest to the stage. This reduction occurs partly because the sound source is farther from the audience, leading to lower energy in the direct sound, but it is primarily due to the absence of early sound reflections. The depth of the stage lacks surfaces that could support and transmit these early reflections to the audience, so listeners primarily receive direct sound. The analysis of acoustic parameters shows that the sound source's position on stage greatly impacts the acoustic parameters. The G_{80} analysis re-

veals that the stage layout in the analysed configuration (without scenery) offers no support for early reflections.

These results indicate that EDT, C_{50}, and G_{80} values significantly depend on the location of the measurement point in the auditorium as well as on the position of the sound source on the stage. For further research, average values of the acoustic parameters calculated from all observation points in the auditorium will be used to provide a comprehensive assessment of the hall's acoustic properties. Additionally, to eliminate the influence of the sound source position on stage, future studies will focus only on measurements from the proscenium location.

3.3. Numerical model of the theatre hall

To estimate the acoustic parameters of the theatre with a full audience, we developed a numerical model using ISIMPA software (Fig. 3). This application employs an advanced geometrical model based on ray tracing and the image sources method to simulate sound propagation within the theatre space. The model was constructed using geometric data obtained through photogrammetric methods and incorporates the specific sound absorption and scattering coefficients of the interior surfaces.
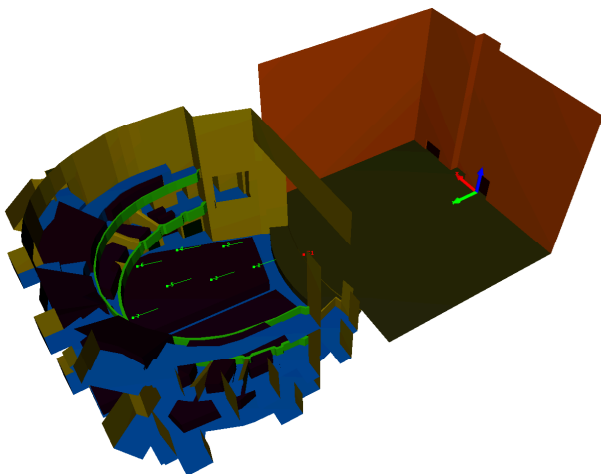

Fig. 3. View of the interior of the Maria Zankovetska Theatre showing the acoustic model in its actual state.

The absorption coefficients for the floors and walls were adapted from values measured in the laboratories of the Department of Mechanics and Vibroacoustics at AGH University of Krakow, Poland for similar venues, such as the Lviv Opera (KAMISIŃSKI *et al.*, 2009). Additionally, absorption coefficients for the stage were determined through direct in-situ measurements taken within the theatre.

To ensure the accuracy of our predictions, the numerical model underwent calibration to match the measured reverberation times obtained within the theatre. This calibration process ensures that the simulated acoustic environment closely reflects real-world conditions, providing a reliable basis for analysing and optimizing the theatre's acoustic performance.

The calibration of the model involved a two-stage process to ensure that the results obtained from the developed acoustic model were consistent with the results of measurements of acoustic parameters in the hall of the Maria Zankovetska Theatre in Lviv. The first stage involved a preliminary selection of sound absorption coefficients for the individual surfaces in the analysed hall. This process was based on a detailed visual inspection of the space, identification of the finishing materials used, laboratory testing of selected material samples, and analysis of available and relevant literature data. The collected data provided the basis for the initial setting of model parameters.

The second stage of validation involved iterative tuning of the model. Each step of the iteration involved fine-tuning of the sound absorption coefficients and then verifying that the calculation results obtained from the model are consistent with the results of measurements taken in the theatre hall. Corrections to the absorption coefficients were carried out until the compatibility criterion was met, that is, when the difference between the average T_{20} reverberation times calculated from the model and measured in the hall did not exceed 5 %. The criterion adopted corresponds to the just-noticeable difference (JND) for reverberation time. It was assumed that this approach would ensure that the model is consistent with the actual acoustic conditions in the theatre hall. The validated model then served as the starting point for further analysis, which is the focus of this article.

The values of sound absorption and scattering coefficients of each surface group after model calibration are shown in Table 5.

Figure 4 illustrates the percentage of sound absorption contributed by each surface group within the total sound absorption within the theatre. The values presented are averages for the 500 Hz and 1000 Hz frequency bands. Upholstered seats and the stage ceiling, which features textile stage decorations, demonstrate the highest levels of sound absorption. These two surface groups possess both high sound absorption coefficients and cover extensive surface areas, collectively accounting for over 50 % of the overall sound absorption within the hall. In contrast, the auditorium and stage walls have relatively low sound absorption coefficients; however, their large surface areas significantly contribute to sound absorption, approximately 20 % (see Fig. 4).

3.4. Sensitivity analysis

3.4.1. Input parameters to sensitivity analysis

Sensitivity analysis of changes in the acoustic parameters of the theatre hall in response to variations

Table 5. Sound absorption and scattering coefficients of surfaces and surface groups used in the theatre hall model: a refers to auditorium, while s indicates stage materials.

| No.	Surface group	S [m²]	a [–] s [–]	f [Hz]					
				125	250	500	1000	2000	4000
1	A:balcony ceiling	341.8	a	0.13	0.16	0.16	0.18	0.19	0.19
			s	0.10	0.12	0.15	0.20	0.25	0.30
2	A:ceiling	338.2	a	0.13	0.16	0.16	0.18	0.19	0.19
			s	0.10	0.12	0.15	0.20	0.25	0.30
3	S:ceiling	371.8	a	0.24	0.48	0.99	0.99	0.99	0.99
			s	0.10	0.12	0.15	0.20	0.25	0.30
4	A:walls	939.8	a	0.13	0.16	0.16	0.18	0.19	0.19
			s	0.10	0.12	0.15	0.20	0.25	0.30
5	S:walls	750.0	a	0.13	0.16	0.16	0.18	0.19	0.19
			s	0.10	0.12	0.15	0.20	0.25	0.30
6	A:front box barriers	111.1	a	0.13	0.16	0.16	0.18	0.19	0.19
			s	0.10	0.12	0.15	0.20	0.25	0.30
7	A:floor	358.2	a	0.13	0.16	0.16	0.18	0.19	0.19
			s	0.10	0.12	0.15	0.20	0.25	0.30
8	S:floor	422.4	a	0.21	0.19	0.11	0.12	0.13	0.15
			s	0.10	0.12	0.15	0.20	0.25	0.30
9	Portal	84.9	a	0.13	0.16	0.16	0.18	0.19	0.19
			s	0.10	0.12	0.15	0.20	0.25	0.30
10	A:windows	22.0	a	0.25	0.16	0.16	0.18	0.19	0.20
			s	0.10	0.12	0.15	0.20	0.25	0.30
11	S:back curtain	327.5	a	0.13	0.16	0.16	0.18	0.19	0.19
			s	0.10	0.12	0.15	0.20	0.25	0.30
12	A:seats	567.8	a	0.70	0.78	0.81	0.85	0.84	0.81
			s	0.30	0.40	0.50	0.60	0.70	0.70
13	A:doors	31.0	a	0.20	0.10	0.10	0.10	0.10	0.10
			s	0.10	0.12	0.15	0.20	0.25	0.30
14	A:barriers	124.8	a	0.13	0.16	0.16	0.18	0.19	0.19
			s	0.10	0.12	0.15	0.20	0.25	0.30
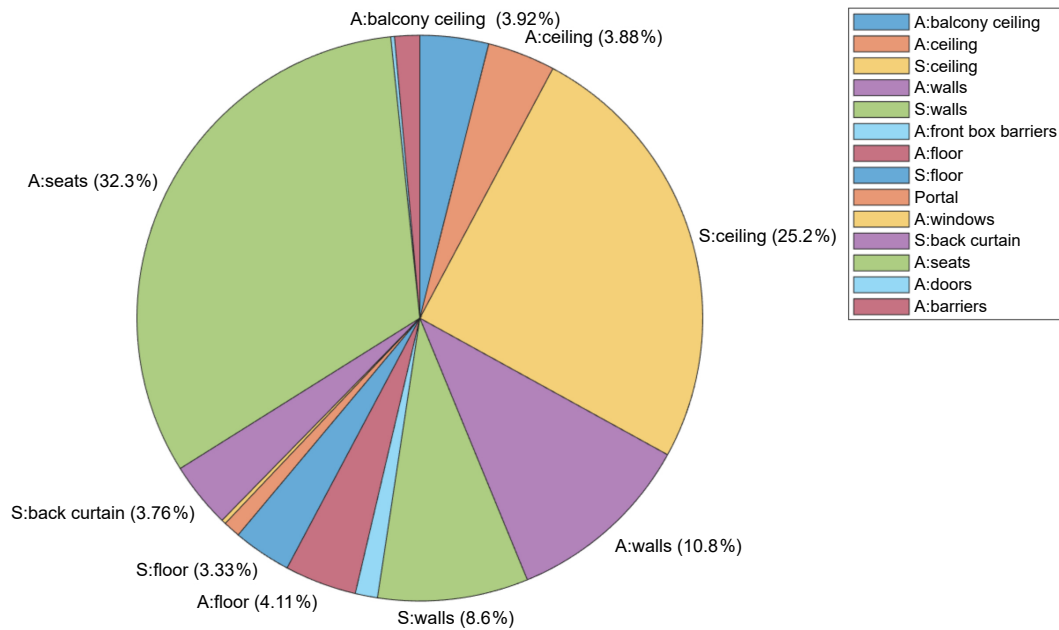
Fig. 4. Sound absorption (average values for 500 Hz and 1000 Hz) of each group of surfaces in the total sound absorption.

in the acoustic absorption of individual surfaces was conducted based on the results of numerical calculations performed using the calibrated model of the theatre hall. During the calculations, the input parameters were modified, assuming a variable value of the sound absorption coefficient. The impact of these changes in sound absorption was assessed based on the average values of the acoustic parameters for the entire hall. The effect of the sound scattering coefficient on the acoustic parameters of the hall was neglected in this study. Previous studies on the sensitivity of concert hall parameters to global changes in the sound scattering coefficient have shown that this influence is small and can be neglected (PILCH, 2024). On the other hand, local analyses of the changes in acoustic parameters after the installation of sound diffusers on the rear wall of the sub-balcony niche showed that the addition of sound diffusion causes locally significant changes in acoustic parameters (KAMISIŃSKI, 2012).

The adopted testing method involved altering the sound absorption coefficient of one surface or group of surfaces at a time. These values were modified within a range of ±50 % of their base values from the calibrated model, in increments of 10 %. For materials with high sound absorption coefficients, it was not possible to reduce their value by 50 %, so their absorption coefficient was increased to a maximum of 1.00 only. To present the results in a comparative format, we converted the change in the sound absorption coefficient into a change in acoustic absorption. Thus, the sensitivity of the hall's acoustic parameters was evaluated based on the change in absorption of each specific surface or surface group.

3.4.2. Analysed output parameters

The sensitivity analysis conducted focused on selected interior acoustic parameters. The EDT and reverberation time T_{20} were chosen to evaluate reverberation. Reverberation time is an indicator for evaluating the acoustics of an interior in the context of its intended function. EDT was included because it is sensitive to changes in the levels of early sound reflections in the room. Additionally, the analysis examined the sensitivity of changes in the clarity index C_{50}. The C_{50} index allows for evaluation of how the early reflection energy contributes to the total energy reaching the audience. Thus, it is important to analyse the impact of the placement of materials with different sound-absorbing properties in the theatre hall. Furthermore, the analysis also included the early sound strength G_{80}, which is essential for assessing the level of early sound reflections in a space and influences the perception of proximity to the sound source.

The choice of these parameters enabled a full and comprehensive analysis of the hall's acoustic characteristics. The analysis was carried out using the average values of the parameters for the 500 Hz and 1000 Hz octave bands, or in the case of the C_{50m} index – the average values across the 500 Hz–2000 Hz frequency range.

4. Analysis of results

4.1. Regression analysis and sensitivity indexes

Sensitivity analysis was conducted using linear regression analysis. This method allows for a quantitative evaluation of how changes in the sound absorption of individual surfaces affect the acoustic parameters. In this analysis, the values of the theatre's acoustic parameters were related to the JND. This approach enables the standardisation of the effects of changes in sound absorption, allowing for direct comparison of their impact on the hall's acoustic parameters. The JND values for the parameters analysed are shown in Table 6.

Table 6. JND for the analysed acoustic parameters.

No.	Parameter	JND
1	T_{20}	5 % of T_{20}
2	EDT	5 % of EDT
3	C_{50}	1 dB
4	G_{80}	1 dB

In order to analyse sensitivity, SRC was utilized. The sensitivity index, denoted by $S_{\text{param},i}$, was determined for the parameters T_{20m}, EDT_m, C_{50m}, and G_{80m} related to the JND for each i-th surface (see Table 5). Standardisation ensures that the value of the sensitivity index $S_{\text{param},i}$ reflects both changes in sound absorption and the area of particular groups of surfaces. The index was calculated in accordance with the following equation:

$$S_{\text{param},i} = b_{1,i} \frac{\sigma_{\Delta Ai}}{\sigma_{\Delta \text{JNDparam},i}} \tag{3}$$

where the standard deviation of the change in sound absorption, denoted here by $\sigma_{\Delta Ai}$ is calculated for each i-th group of surfaces. The change in the sound absorption $DA_i = A_{i,k} - A_{i,0}$ is defined as the difference between the sound absorption for the k-th value $A_{i,k}$ and the initial value $A_{i,0}$. The index k refers to the change in sound absorption The standard deviation of the change in acoustic parameter values related to JND ($DJND_{\text{param},i}$) is denoted by $\sigma_{DJND\text{param},i}$. The change in the acoustic parameters related to JND: $DJND_{\text{param},i} = \text{JND}_{\text{param},i,k}(A_{i,k}) - \text{JND}_{\text{param},i,0}(A_{i,0})$ is defined as the difference between the parameter values determined for the k-th value of sound absorption and those determined for the initial value of $A_{i,0}$.

The regression coefficients $b_{0,i}$, $b_{1,i}$ were determined from the regression model $DJND_{\text{param},i} = f(DA_i)$ for each i-th group of surfaces as follows:

$$\Delta\text{JND}_{\text{param},i} = b_{0,i} + b_{1,i}\Delta A_i. \tag{4}$$

4.2. Quantitative analysis

An analysis was performed to evaluate the sensitivity indexes ($S_{\mathrm{param},i}$) of the various acoustic parameters related to the JND. Based on the $S_{\mathrm{param},i}$ values, the surfaces with the most significant impact on each parameter were identified. A standard significance level of 0.05 was employed. A linear relationship between the change in JND ($DJND_{\mathrm{param},i}$) and the change in sound absorption (DA_i) was considered significant when $p < 0.05$. The regression coefficients and p-values for the $DJND$ of the T_{20} and EDT parameters are presented in Tables 7 and 8.

Table 7. Regression coefficients and p-values for $DJND_{T_{20}}$.

No.	Surface group	S [m²]	b_1	b_0	p-value
1	A:balcony ceiling	341.8	0.000	−0.208	0.998
2	A:ceiling	338.2	−0.002	−0.377	0.450
3	S:ceiling	371.8	−0.018	−0.516	**0.000**
4	A:walls	939.8	−0.019	0.295	**0.000**
5	S:walls	750.0	−0.015	0.088	**0.000**
6	A:front box barriers	111.1	−0.001	−0.216	0.854
7	A:floor	358.2	−0.008	0.136	**0.007**
8	S:floor	422.4	−0.006	0.131	0.090
9	Portal	84.9	−0.034	0.315	**0.019**
10	A:windows	22.0	−0.046	0.083	0.238
11	S:back curtain	327.5	−0.034	0.187	**0.000**
12	A:seats	567.8	−0.003	0.056	**0.000**
13	A:doors	31.0	−0.036	0.355	0.510
14	A:barriers	124.8	−0.010	−0.012	0.180

Table 8. Regression coefficients and p-value for $DJND_{\mathrm{EDT}}$.

No.	Surface group	S [m²]	b_1	b_0	p-value
1	A:balcony ceiling	341.8	−0.001	−0.297	0.505
2	A:ceiling	338.2	−0.005	0.173	**0.033**
3	S:ceiling	371.8	−0.017	−0.039	**0.000**
4	A:walls	939.8	−0.015	−0.076	**0.000**
5	S:walls	750.0	−0.023	0.125	**0.000**
6	A:front box barriers	111.1	−0.031	0.140	**0.001**
7	A:floor	358.2	−0.005	−0.108	0.303
8	S:floor	422.4	−0.021	−0.092	**0.000**
9	Portal	84.9	−0.016	0.067	**0.000**
10	A:windows	22.0	0.004	0.113	0.282
11	S:back curtain	327.5	−0.028	−0.032	**0.000**
12	A:seats	567.8	−0.004	0.101	**0.000**
13	A:doors	31.0	−0.002	0.078	0.392
14	A:barriers	124.8	−0.014	−0.245	0.105

The analysis shows that, for many surface groups with small areas, the results obtained were not statistically significant ($p > 0.05$). This implies that, for these areas, there is no linear relationship between changes in a parameter and changes in its acoustic absorption. Figure 5 presents the sensitivity indexes $S_{\mathrm{param},i}$ that are statistically significant.
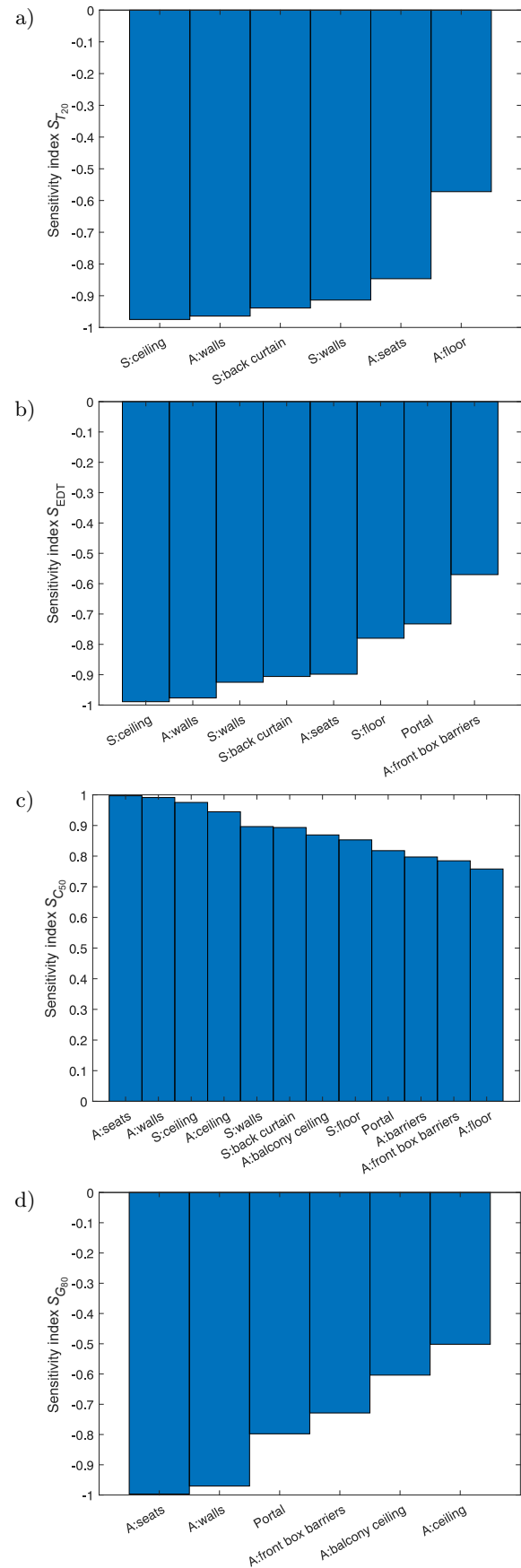
Fig. 5. Sensitivity indexes $S_{\mathrm{param},i}$ for the: a) T_{20}, b) EDT, c) C_{50}, and d) G_{80}, all relative to JND.

Negative values of the sensitivity indexes S_{EDT}, $S_{T_{20}}$, and $S_{G_{80}}$ (Figs. 5a, 5b, 5d) indicate that the corresponding parameter is negatively correlated with changes in the acoustic absorption ΔA_i of the i-th surface. Higher absolute values of the sensitivity indexes indicate that a change in the sound absorption of a specific surface has a greater impact on the associated acoustic parameter.

The obtained sensitivity indexes S_{EDT}, $S_{T_{20}}$ show that the reverberation time is sensitive to changes in the sound absorption of large surfaces on the stage and in the audience area, particularly the stage ceiling (Figs. 5a and 5b). This confirms the measured values of EDT, where higher values were observed near the stage. This indicates that the stage volume and the presence of sound-absorbing materials significantly affect the reverberation characteristics of the hall. The analysis of C_{50} revealed that this parameter is particularly sensitive to changes in the sound absorption of surfaces located in the audience area, which is related to providing early reflections from nearby surfaces (Fig. 5c). Results obtained for the early sound strength G_{80} show that this parameter is particularly sensitive to changes in seat absorption and the walls surrounding the audience (Fig. 5d). However, the sensitivity indexes $S_{\text{param},i}$ do not give any information on how much a change in the sound ab-

sorption of a given surface affects a specific acoustic parameter within the hall. The range of possible parameter changes is limited by the range of changes in the sound absorption coefficient of the given surface. Therefore, a qualitative analysis was carried out to determine the significance of potential changes in the acoustic parameters of the hall.

4.3. Qualitative analysis

A qualitative analysis was conducted in order to evaluate the significance of changes in acoustic parameters resulting from alterations in the sound absorption of a particular surface. Sensitivity analysis of the parameters related to JND was performed using unnormalized regression coefficients fitted to the model described in Eq. (4). In this particular analysis, the sensitivity index was represented by the unnormalized regression coefficient b_1. It was hypothesized that the coefficient b_0 would be zero, and that for an input $\Delta A_i = 0$, the change in output parameters would also be zero. An analysis was conducted to evaluate the maximum changes in the T_{20}, EDT, C_{50}, and G_{80} parameters relative to JND, in response to the modification of the absorption for individual values of A_i (see Figs. 6 and 7).

The most significant changes in $\text{JND}_{T_{20}}$ (Fig. 6a) were observed in the change of the sound absorption
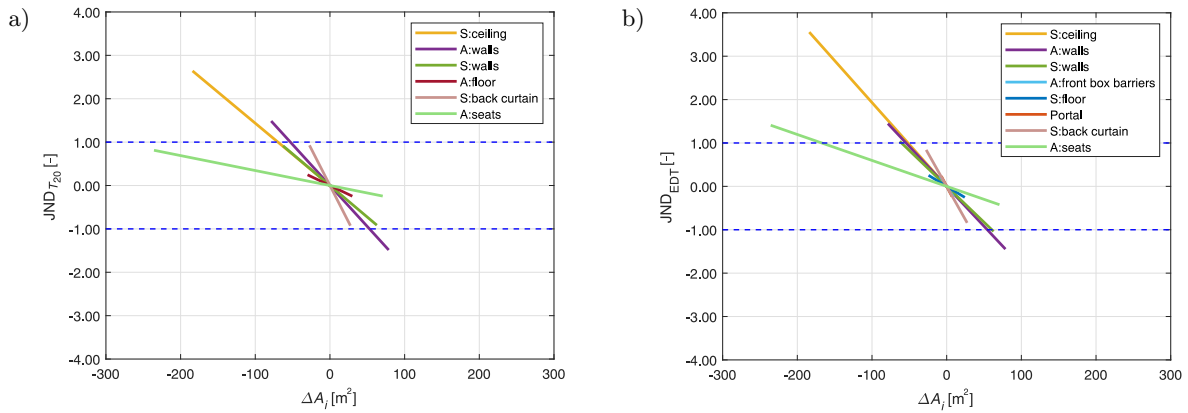

Fig. 6. Values of parameters related to the JND: a) T_{20} and b) EDT changes depending on the change in the sound absorption ΔA_i.
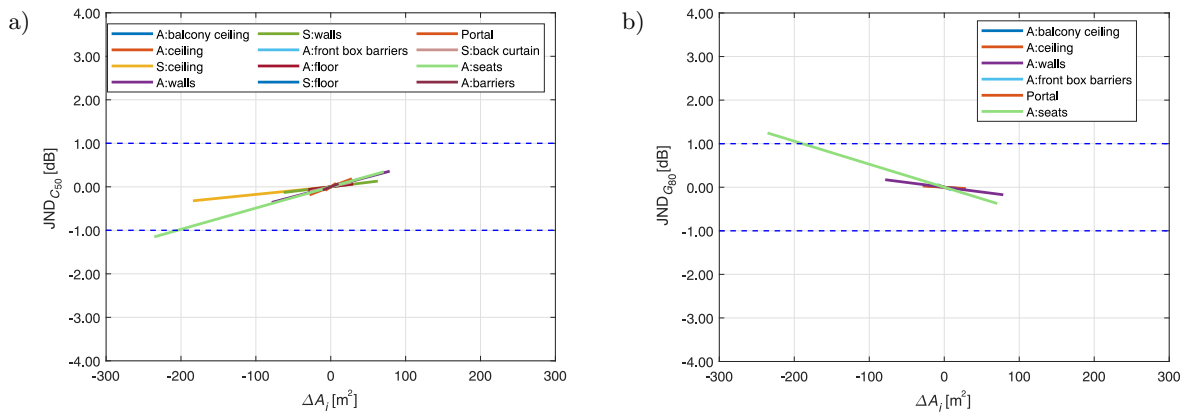

Fig. 7. Values of parameters related to the JND: a) C_{50} and b) G_{80} as a function of the change in the sound absorption ΔA_i.

of the auditorium walls and the ceiling above the stage. The range of potential JND changes for the stage ceiling is greatest, but only when absorption is reduced, leading to an increase in reverberation. This phenomenon can be attributed to the ceiling space above the stage being occupied by stage elements made of textile materials characterised by high sound absorption coefficients. Consequently, it is not feasible to achieve further increases in sound absorption. A similar effect is observed for the seating in the hall, where a substantial reduction in absorption results in an increase in reverberation time of approximately 1 JND. Conversely, modifying the sound absorption properties of the auditorium walls has the potential to either decrease or increase the reverberation time. A similar outcome can be achieved by adjusting the absorption of the rear stage curtain, although a change in the absorption coefficient of more than 50% is required for this adjustment to significantly impact the acoustic parameters. It is important to note that enhancing the reverberation time in this hall is not desirable, particularly on stage, as it can result in a discrepancy between reverberation conditions on stage and in the audience areas. In order to achieve equal reverberation between the stage and the auditorium, consideration should be given to increasing sound absorption on the stage by increasing the absorption of the stage walls and the rear curtain.

The JND results for the EDT parameter (Fig. 6b) are consistent with those observed for T_{20}. The only significant difference is that it is possible to significantly increase the EDT time by reducing the absorption of the audience seating, for example, by using wooden seats instead of upholstered ones. However, as previously indicated, increasing the reverberation in the hall is not desirable and would not be a reasonable approach in terms of spectator comfort.

Analysis of changes in JND for the C_{50} clarity index (see Fig. 7a) reveals that a radical reduction in seat absorption can result in a significant decrease of C_{50}, potentially leading to a noticeable reduction in speech intelligibility. It has been demonstrated that alterations in the sound absorption of other surfaces do not have a significant impact on the C_{50} value. It is noteworthy that the highest sensitivity of $JND_{C_{50}}$ is observed for the auditorium ceiling, auditorium walls, and seats, which is related to the steep slopes of the lines. Nevertheless, alterations in the sound absorption for these surfaces do not result in any significant changes in $JND_{C_{50}}$. This sensitivity is associated with early reflections that reach the audience area; therefore, modifying the absorptivity of these surfaces can control early reflections and adjust the balance between early and late arriving energy.

The maximum increase in $JND_{G_{80}}$ (see Fig. 7b) is observed in cases where there is a decrease in sound absorption of the seats. As was stated in the preceding paragraph, this phenomenon is associated with early reflections reaching the seating area. It has been demonstrated that reducing seat sound absorption can result in an increase the level of early reflection, thereby increasing G_{80}. However, an increase in G_{80} results in a reduction in C_{50}. Within the context of a theatre, the acoustic parameter that appears to be of the greatest importance is C_{50}, given its correlation with speech intelligibility. For all other surfaces, altering the sound absorption results in negligible alterations to G_{80}.

4.4. Sensitivity analysis of changes in sound absorption of the audience and stage area

The subsequent stage of the research was to analyse the sensitivity of the acoustic parameters to simultaneous changes in the sound absorption of walls and ceilings in both the stage and auditorium. For this analysis, surfaces were grouped into walls and ceilings. The walls on stage comprised the side walls and the curtain on the back wall. The term 'walls' in the context of the auditorium refers to all wall surfaces and the fronts of the boxes. Similarly, the term 'ceilings' refers to the main ceiling and the ceilings beneath the balconies. The objective of this analysis was to evaluate the impact of simultaneous increases or decreases in sound absorption in each area on the key acoustic parameters. A comprehensive description of the configurations analysed is provided in Table 9.

Table 9. Description of analysed modifications of sound absorption in the stage and audience areas.

| No.	Configuration	Description
1	S:↑w+c (−50% - 0%)	Increase the sound absorption coefficient of walls and ceilings on the stage from −50% to 0%
2	S:↑w (0% - 50%)	Increase the sound absorption coefficient of walls and ceilings on the stage from 0% to 50%
3	A:↑w+c (−50% - 50%)	Increase the sound absorption coefficient of walls and ceiling in the audience area from −50% to 50%
4	S:↑w (−50% - 50%), A:↓w (50% - −50%)	Increase the sound absorption coefficient of walls on the stage from −50% to 50% and decrease it in the audience area from 50% to −50%
5	S:↑c (−50% - 0%), A:↓c (50% - 0%)	Increase the sound absorption coefficient of ceilings on the stage from −50% to 0% and decrease it in the audience area from 50% to 0%
6	A:↓c (0% - −50%)	Decrease the sound absorption coefficient of ceilings in the audience area from 0% to −50%
7	S:↑w+c (−50% - 0%), A:↓w+c (50% - 0%)	Increase the sound absorption coefficient of walls and ceiling on the stage from −50% to 0% and decrease in the audience area from 50% to 0%
8	S:↑w (0% - 50%), A:↓w+c (0% - −50%)	Increase the sound absorption coefficient of walls on the stage from 0% to 50% and decrease it in the audience area from 0% to −50%
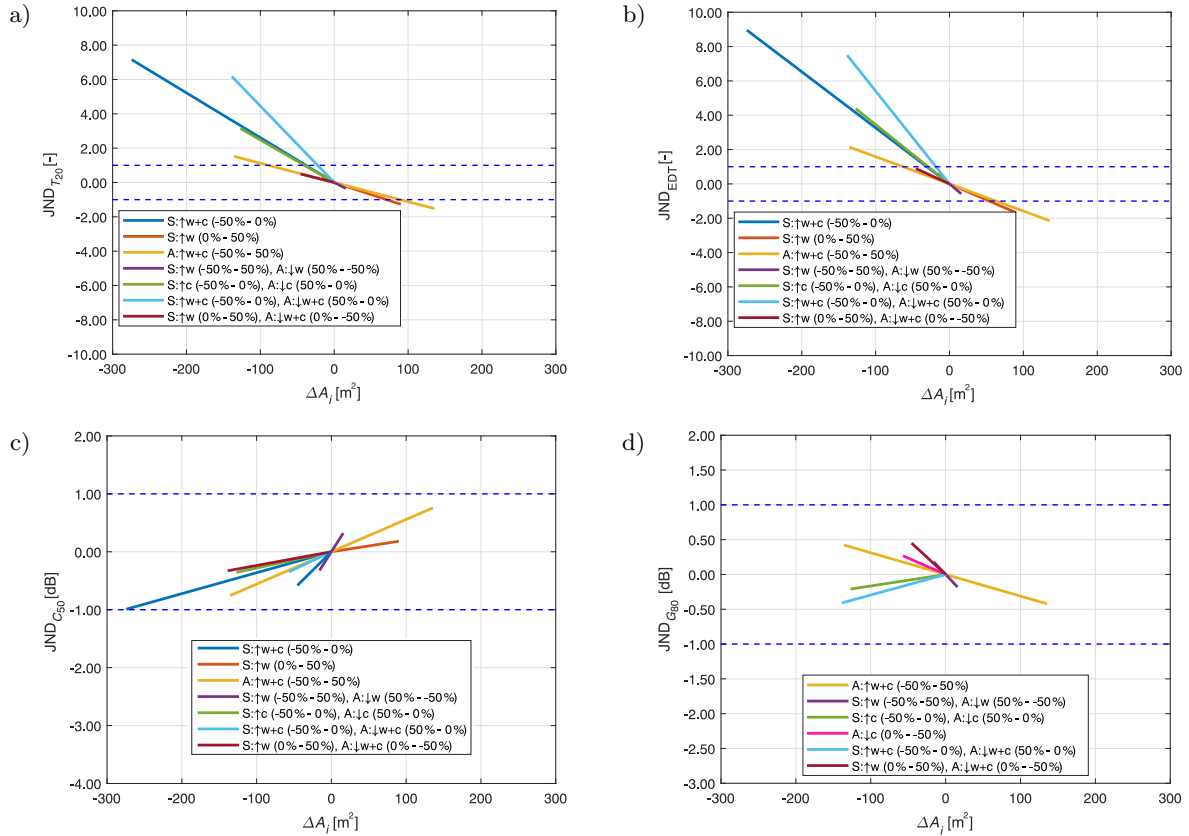
Fig. 8. Values of parameters related to the JND: a) T_{20}, b) EDT, c) C_{50}, and d) G_{80} depending on the change in the sound absorption ΔA_i.

Sensitivity analysis was performed to assess the impact of variations in acoustic parameters related to JND at specific locations, either on the stage or within the auditorium (configurations 1–3). Conversely, for configurations 4–8, the impact of sound absorption changes on the parameters was examined in both the stage and the auditorium areas (see Fig. 8). The arrows used in the descriptions are intended to indicate the direction of change in the sound absorption coefficient: an upward arrow indicates an increase in the sound absorption coefficient, while a downward arrow indicates a decrease. In the case of a stage ceiling characterised by a high sound absorption coefficient, the analysed changes in its absorption coefficient were carried out in the range from −50 % to 0 %.

The most significant changes in the T_{20} and EDT parameters relative to JND were observed for configurations 1, 5, and 7, where sound absorption on the stage walls and ceiling was reduced (see Figs. 8a and 8b). The increase in reverberation time in these configurations ranged from 4 to 10 JND. Conversely, a substantial decrease in reverberation time values, of approximately −2 JND, was observed when the sound absorption of the auditorium walls and ceiling or the stage walls was increased (configurations 2 and 3). In addition, it should be noted that increasing the sound absorption of the stage walls enables a significant re-

duction in reverberation time on stage and, consequently, throughout the entire hall, thereby achieving balanced reverberation conditions on stage as well as in the auditorium area.

The analysis of the C_{50} parameter (see Fig. 8c) demonstrates that a reduction in the sound absorption properties of the walls and ceiling in the stage area (configuration 1) can result in a substantial decline in speech intelligibility. Conversely, enhancing speech intelligibility can be accomplished by increasing sound absorption in the walls and ceiling of the audience area (configuration 3). Within the range of sound absorption variations examined, which not exceeded 50 %, the alterations did not rise above 1 JND. It is evident that in order to achieve a substantial modification in the C_{50} parameter, it is necessary for the change in the sound absorption coefficient to exceed 50 %.

The analysis of the G_{80} parameter demonstrates that the main factor in the alteration of early sound strength (G_{80}) is the modification of the sound absorption properties of the audience wall and ceiling (see Fig. 8d). For instance, alterations in sound absorption on the stage are not statistically significant (configuration 1), and there is no correlation between G_{80} and sound absorption. This finding serves to emphasise the negligible impact of alterations in sound absorption on the stage in terms of shaping the G_{80}

parameter. It is also important to note that even alterations in sound absorption exhibited by both surfaces do not result in a substantial change in G_{80}.

5. Conclusions

This paper presented the findings of a sensitivity analysis of selected acoustic parameters relevant to theatres halls. The surfaces and groups of surfaces that significantly influence the acoustic parameters of the hall were identified as the stage ceiling, the stage walls with curtains, the audience walls, and the seats.

The surfaces with the greatest influence on the EDT and T_{20} reverberation parameters were the stage walls and ceiling, where the majority of the sound-absorbing materials are located. The auditorium seating and side walls also had a significant impact on EDT changes. This finding indicates that EDT exhibits enhanced sensitivity compared to T_{20} in response to variations in early reflections from nearby surfaces due to modifications in sound absorption.

A substantial enhancement in speech intelligibility, characterized by the clarity index C_{50}, could be achieved by increasing the absorption of the auditorium walls and ceiling. Conversely, a significant reduction in C_{50} could be achieved by reducing the sound absorption of the seats, e.g., by removing upholstery. It is noteworthy that other surfaces, despite their relatively high sensitivity, do not have a significant effect on clarity.

In a similar manner, the value of early sound strength G_{80}, which is responsible for the perception of proximity to the sound source and also serves to assess the amplification of natural sound by the room, can be primarily influenced by reducing the absorption in the audience area, including seats, walls and ceilings.

The sensitivity analysis of the parameters also revealed that the clarity index C_{50} and the sound strength G_{80} were significantly less sensitive to changes in acoustic absorption compared to the reverberation parameters T_{20} and EDT. Increasing acoustic absorption on several surfaces did not result in a substantial improvement of these parameters. This phenomenon can be explained by the fact that there are already many existing absorbing elements in the hall, such as stage curtains or upholstered seats, which already introduce very high absorption into the hall. Consequently, the addition of absorption to the remaining surfaces results in a relatively small change in the total absorption and, therefore, does not result in a significant change in the acoustic characteristics. The potential for improving these parameters, particularly G_{80}, could be achieved by modifying early reflections, e.g., with appropriately placed reflective panels and scenery elements.

In conclusion, it is also worth noting that the research results obtained for the Maria Zankovetskay Theatre in Lviv can be generalized to other theatres with classical stage and auditorium layouts. This is especially applicable to theatres with a tower stage, which is characterized by a much larger volume than the auditorium area. In such theatres, the primary challenge in shaping the reverberation of the interior is to ensure a balance between the stage and auditorium areas. Conversely, when evaluating the parameters of speech intelligibility and the amplification of sound transmission within the auditorium, proper management of early sound reflections becomes important.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Authors' contribution

The authors contribution is as follows: Jarosław Rubacha – conceptualization, methodology, data analysis, writing the manuscript; Adam Pilch – development of a numerical model, numerical data analysis; Roman Kinasz – supervision, research organization, funding acquisition, manuscript review; Wojciech Binek – measurement and data analysis, verification of the numerical model; Tadeusz Kamisiński – supervision, results analysis, manuscript review; Mykhaylo Melnyk – measurement and data analysis. All authors reviewed and approved the final manuscript.

References

1. Barron M. (2009), *Auditorium Acoustics and Architectural Design*, 2nd ed., Spon Press.

2. Beranek L.L., Hidaka T. (1998), Sound absorption in concert halls by seats, occupied and unoccupied, and by the hall's interior surfaces, *Journal of the Acoustical Society of America*, **104**(6): 3169–3177, https://doi.org/10.1121/1.423957.

3. Binek W., Pilch A., Kamisiński T. (2022), Direct application of the diffusers' reflection patterns in geometrical acoustics simulations, *Applied Acoustics*, **198**: 108949, https://doi.org/10.1016/j.apacoust.2022.108949.

4. Borgonovo E. (2008), Sensitivity analysis of model output with input constraints: A generalized rationale for local methods, *Risk Analysis*, **28**(3): 667–680, https://doi.org/10.1111/j.1539-6924.2008.01052.x.

5. Borgonovo E., Plischke E. (2016), Sensitivity analysis: A review of recent advances, *European Journal of Operational Research*, **248**(3): 869–887, https://doi.org/10.1016/J.EJOR.2015.06.032.

6. Hamm N.A.S., Hall J.W., Anderson M.G. (2006), Variance-based sensitivity analysis of the probability of hydrologically induced slope instability, *Computers & Geosciences*, **32**(6): 803–817, https://doi.org/10.1016/J.CAGEO.2005.10.007.

7. Iooss B., Lemaître P. (2015), A review on global sensitivity analysis methods, *Uncertainty Management in Simulation-Optimization of Complex Systems. Algorithms and Applications*, pp. 101–122, https://doi.org/10.1007/978-1-4899-7547-8_5

8. International Organization for Standardization (2009), *Acoustics – Measurement of room acoustic parameters. Part 1: Performance spaces* (ISO Standard No. 3382-1:2009), https://www.iso.org/standard/40979.html.

9. Kamisiński T. (2010), Acoustic simulation and experimental studies of theatres and concert halls, *Acta Physica Polonica A*, **118**(1): 78–82, https://doi.org/10.12693/APhysPolA.118.78.

10. Kamisiński T. (2012), Correction of acoustics in historic opera theatres with the use of Schroeder diffuser, *Archives of Acoustics*, **37**(3): 349–354, https://doi.org/10.2478/v10168-012-0044-1.

11. Kamisiński T., Kinasz R., Rubacha J., Flach A. (2009), Acoustic studies of selected materials for the renovation of the interior of the Lviv Opera [in Polish], [in:] *Building Physics in Theory and Practice: XII Polish Scientific and Technical Conference*, pp. 143–144.

12. Manache G., Melching C.S. (2008), Identification of reliable regression- and correlation-based sensitivity measures for importance ranking of water-quality model parameters, *Environmental Modelling & Software*, **23**(5): 549–562, https://doi.org/10.1016/J.ENVSOFT.2007.08.001

13. Pianosi F. *et al.* (2016), Sensitivity analysis of environmental models: A systematic review with practical workflow, *Environmental Modelling & Software*, **79**: 214–232, https://doi.org/10.1016/J.ENVSOFT.2016.02.008.

14. Pilch A. (2020), Optimization-based method for the calibration of geometrical acoustic models, *Applied Acoustics*, **170**: 107495, https://doi.org/10.1016/j.apacoust.2020.107495.

15. Pilch A. (2021), Optimized diffusers for shoe-box shaped performance halls, *Applied Acoustics*, **178**: 108019, https://doi.org/10.1016/j.apacoust.2021.108019.

16. Pilch A. (2024), Sensitivity testing in room acoustic modeling on changes of input parameters on the example of selected concert halls, [in:] *Proceedings of the 10th Convention of the European Acoustics Association Forum Acusticum 2023*, pp. 111–114, https://doi.org/10.61782/fa.2023.0955.

17. Prodi N., Pompoli R. (2016), Acoustics in the restoration of Italian historical opera houses: A review, *Journal of Cultural Heritage*, **21**: 915–921, https://doi.org/10.1016/j.culher.2016.03.004.

18. Rubacha J. *et al.* (2019), Analysis of the acoustic parameters of the Maria Zankovetska Theatre in the Lviv before and after modernisation of the audience, [in:] *IOP Conference Series: Materials Science and Engineering*, **471**: 082035, https://doi.org/10.1088/1757-899X/471/8/082035.

19. Rubacha J., Pilch A., Zastawnik M. (2012), Measurements of the sound absorption coefficient of auditorium seats for various geometries of the samples, *Archives of Acoustics*, **37**(4): 483–488, https://doi.org/10.2478/v10168-012-0060-1.

20. Saltelli A., Annoni P. (2010), How to avoid a perfunctory sensitivity analysis, *Environmental Modelling Software*, **25**(12): 1508–1517, https://doi.org/10.1016/J.ENVSOFT.2010.04.012.

21. Saltelli A., Tarantola S., Campolongo F., Ratto M. (2004), *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*, Wiley.

22. Shtrepi L. (2019), Investigation on the diffusive surface modeling detail in geometrical acoustics based simulations, *The Journal of the Acoustical Society of America*, **145**(3): EL215–EL221, https://doi.org/10.1121/1.5092821.

23. Szeląg A., Kamisiński T., Lewińska M., Rubacha J., Pilch A. (2014), The characteristic of sound reflections from curved reflective panels, *Archives of Acoustics*, **39**(4): 549–558, https://doi.org/10.2478/aoa-2014-0059.

24. Szeląg A., Rubacha J., Pilch A., Majchrzak A., Kamisiński T. (2020), Reflective panels with Schroeder diffusers – A measure to broaden the effective frequency range of sound reflection from overhead stage canopies, *Applied Acoustics*, **157**: 107034, https://doi.org/10.1016/j.apacoust.2019.107034.

25. Tang Y., Reed P., Wagener T., van Werkhoven K. (2007), Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation, *Hydrology and Earth System Sciences*, **11**(2): 793–817, https://doi.org/10.5194/hess-11-793-2007.

Research Paper

Listening Effort in Reverberant Rooms: A Comparative Study of Subjective Perception and Objective Acoustic Metrics

Anna PASTUSIAK*[iD], Łukasz BŁASIŃSKI[iD], Jędrzej KOCIŃSKI[iD]

Department of Acoustics, Faculty of Physics and Astronomy, Adam Mickiewicz University
Poznań, Poland

*Corresponding Author e-mail: anna.pastusiak@amu.edu.pl

Modern room acoustics employs a variety of objective measures to characterize the acoustical properties of interiors. Despite these advancements, the relationship between these parameters and subjective assessments of room acoustics remains unclear. Subjective perception, particularly listening effort (LE), plays a critical role in how individuals experience acoustic environments, even when speech intelligibility (SI) is high. This study aims to bridge the gap between objective acoustic measures and subjective listening experiences. We conducted experiments in three rooms equipped with reverberation enhancement systems, resulting in nine different acoustic settings. Objective parameters, including reverberation time (RT), early decay time (EDT), clarity (C50), and the speech transmission index (STI), were measured. Additionally, subjective SI was assessed, and LE was rated on a 7-step Likert scale by 180 volunteers with normal hearing. The analysis revealed a non-linear relationship between LE and both RT20 and EDT ($R^2 = 0.6$), with an even weaker correlation for LE vs. C50 ($R^2 = 0.46$). The Pearson correlation coefficient for STI was 0.74, compared to 0.55 for SI. These findings indicate that the relationship between LE and objective parameters, as well as SI, is complex and not straightforward. Our results suggest the importance of incorporating LE into room acoustic design and evaluation. The disparity between objective measures and subjective experiences suggests that LE may be a crucial factor in accurately assessing acoustic environments. This approach sheds the light on a more holistic understanding of acoustic quality that prioritizes human perception.

Keywords: room acoustics; listening effort; speech intelligibility; reverberation time; early decay time; clarity (C50); speech transmission index.

1. Introduction

Processing spoken language, which begins with the extraction of key sensory information from a rapidly changing acoustic signal, requires a series of perceptual and cognitive analyses (STRAND *et al.*, 2018) that involve a certain amount of cognitive effort (PEELLE, 2018). The extent of this effort depends on several factors. Among the most significant are the presence and type of interfering signals, such as background noise or concurrent speech (FESTEN, PLOMP, 1990), the content and complexity of the speech being heard (JUST, CARPENTER, 1992), the degradation of the target signal (WILD *et al.*, 2012), the characteristics of a speaker (SCHMID, YENI-KOMSHIAN, 1999), and nat-

urally, those of the listener (MATTYS *et al.*, 2012). In general, it can be stated that when the target signal is distorted, disrupted in some way, or its reception is somehow limited, listeners must engage significantly more cognitive resources to extract useful information compared to an undistorted signal presented in favorable conditions and received by an unrestricted auditory apparatus (PEELLE, 2018), even when speech intelligibility (SI) remains unchanged (HOUBEN *et al.*, 2013).

In clinical audiology, this increased cognitive load is particularly relevant for individuals with hearing loss or auditory processing difficulties, as they may exert disproportionately greater effort to achieve comparable levels of speech understanding. In this context, lis-

tening effort (LE) becomes an important metric for assessing not only the performance of hearing aid but also the broader effectiveness of auditory rehabilitation strategies (OHLENFORST *et al.*, 2017).

Undoubtedly, this is why, in recent years, there has been growing interest in research on the cognitive load associated with speech perception, which is referred to as LE (e.g., PICHORA-FULLER *et al.*, 2016; LEMKE, BASSER, 2016). This phenomenon was first formally defined by the Cognition in Hearing Special Interest Group of the British Society of Audiology (MCGARRIGLE *et al.*, 2014) as 'the mental effort required to listen to and understand an auditory message'. An extension of this formulation is a more generic description describing it as the intentional allocation of mental resources to overcome obstacles in goal pursuit during a (listening) task (PICHORA-FULLER *et al.*, 2016). While this definition focuses on speech, it also accommodates other types of signals encountered in real-life conditions (SHINN-CUNNINGHAM, BEST, 2008). LE thus serves as a bridge between audiological diagnostics and environmental design. It provides insight not only into what a listener hears, but also how much effort they must put to comprehend it – highlighting the importance of considering both technological and architectural solutions in parallel (ZEKVELD *et al.*, 2010; MCGARRIGLE *et al.*, 2014).

Among the factors that can influence the amount of LE are the room acoustics parameters in which a speech signal is presented. One commonly used metric to describe such environments is reverberation time (RT). It is defined as the time it takes for the sound pressure level (SPL) of a specific sound source to decrease by 60 dB after being abruptly switched off (SABINE, 1922). Due to technical limitations (achieving a measurement range exceeding 60 dB is often not possible), RT20 and RT30 measures are typically used, beginning 5 dB below the steady-state energy level (PN-EN ISO, 2010). In such cases, RT20 is three times the time required for a 20 dB SPL decay, while RT30 is twice the time required for a 30 dB SPL drop.

It is commonly known that different event types require different RTs. For music with lyrical content, RTs of approximately 1 s are recommended (SAKAI *et al.*, 2000; ANDO, 2007), and for music without lyrics, longer RTs are generally preferred (KUHL, 1954). In theatres, suggested RT may be up to 1.6 s (MEYER, 1978). In auditoriums, big classrooms, and other spaced where speech remains the main signal, lower RT, starting from 0.5 s–0.7 s to 1 s are recommended to maintain proper SI (BRADLEY, 1985; EVEREST, 2001). Values within this range prevent reflected sounds from overlapping with the direct signal, a common issue associated with excessively long RTs. From the standpoint of LE, prolonged reverberation may not significantly impact intelligibility scores, but it can impose greater strain on listeners, especially over

longer periods or in cognitively demanding contexts. This is particularly relevant in educational or healthcare settings, where even small increases in LE may negatively affect comprehension, memory, and fatigue levels (HERRMANN, JOHNSRUDE, 2020).

Interestingly, the subjective perception of reverberation depends on the excitation and noise level, aligning more closely with the early decay time (EDT), which refers to the initial and most perceptually relevant portion of the decaying energy (KUTTRUFF, 2009), rather than with the RT itself (AHNERT, TENNHARDT, 2008).

Due to questions and doubts regarding the potential relationship between objective and subjective parameters that determine a room's acoustics, numerous studies have been conducted at the intersection of room acoustics, audiology, and cognitive science. In numerous studies (e.g., GIMÉNEZ *et al.*, 2014; KOCIŃSKI, OZIMEK, 2017; BLASINSKI, KOCIŃSKI, 2023), objective room parameters are compared with SI as determined by listening tests. This line of research has been particularly prominent in recent years, although the importance of such analyses has long been recognized. The methods for evaluating room acoustics developed by Beranek and Ando are fundamental for this field. BERANEK (2004) concentrated on technical parameters such as RT and SI. In contrast, ANDO (1998) introduced a more subjective approach, focusing on auditory preferences that consider factors such as the timing of early reflections and the width of the sound source.

Recent approaches increasingly incorporate LE scales, eye-tracking, and pupillometry to assess the mental exertion required in specific acoustic environments. These tools help reveal subtle deficits in comfort or usability that may go undetected through traditional intelligibility measures alone (WENDT *et al.*, 2018). Although these suggested methods were not originally designed to assess LE, they have proven valuable in analyzing room acoustic perception and indicate that even with excellent SI, excessive LE can lead to negative evaluations of room acoustics (VISENTIN *et al.*, 2018). In this way, LE serves as a more sensitive indicator of acoustic quality than intelligibility alone. It captures the hidden cognitive cost of seemingly 'good' communication – an essential consideration for audiologists, designers, and engineers alike. Assessing LE is therefore crucial in the field of room acoustics, as the ultimate goal is to ensure that human listeners feel comfortable in various enclosures, regardless of whether correct objective parameters or high SI scores. Consequently, after over 100 years of advancements in modern room acoustics, there is a growing emphasis on subjective evaluations, which can now be quantified using objective measures (e.g., effort scale).

Similar to trends in audiology research (OHLENFORST *et al.*, 2017; WANG *et al.*, 2018), there is a grow-

ing interest in assessing LE, especially in situations where SI is high. An example that demonstrates the relevance of this distinction is a conversation taking place in a highly reverberant room, where multiple speakers are talking from different directions. An individual with normal hearing is likely to understand most, if not all, of the spoken words, achieving an SI score close to 100 %. However, the level of concentration and cognitive effort required in this scenario is significantly greater compared to a conversation with a single interlocutor in a quiet setting (VISENTIN *et al.*, 2018).

Ultimately, decisions regarding the placement of equipment that affects room acoustics should consider the comfort of the intended users, defined as achieving minimal or no LE wherever possible.

High SI and low LE are absolutely crucial for ensuring the quality and effectiveness of communication and public address systems in any context.

While SI is a critical requirement, often governed by stringent safety standards and type-approval regulations, such as those for building announcement systems or aircraft communications. Similarly, minimizing LE is essential in challenging environments such as vehicles, mobile phone or headset usage, and conference settings, where poor audio quality can significantly hinder communication.

Therefore, rigorous testing, optimization, and validation are not just recommended – they are essential to guarantee both optimal user experience and, where necessary, compliance with certification standards. Failure to address these factors may compromise both usability and safety.

2. Materials and methods

2.1. Aim

In this study, we focus on speech perception effort as a critical factor in room acoustics. We present an analysis of the subjective characteristic of auditory effort associated with logatome perception, juxtaposed with objective acoustic parameters such as RT20, EDT, clarity (C50), and speech transmission index (STI).

C50, introduced by MARSHALL (1994), represents the logarithmic ratio of early-to-late arriving sound energy, where 'early' pertains to the initial 50 (or 80) milliseconds, and 'late' denotes the period following this (KOCIŃSKI, OZIMEK, 2017). The 50-millisecond threshold plays a crucial role in distinguishing beneficial reflections from detrimental ones and is instrumental in evaluating a room's suitability for speech perception.

The STI is derived from measurements of the modulation transfer function (HOUTGAST, STEENEKEN, 1973) and quantifies signal quality on a scale from 0

(poor intelligibility) to 1 (excellent intelligibility). For the rooms presented in this research, the STI is classified as 'fair' (for enclosure 1 and one setting in enclosure 3) and 'good' for the remaining settings (STI above 0.6). The study was carried out with the greatest possible care, respecting the principles of anonymity of respondents and adhering to the guidelines outlined in the Declaration of Helsinki. Given the scope of the study, ethics committee approval was not required.

2.2. Measurement setup

To create a range of different reverberant conditions, recordings from three enclosures equipped with reverberation enhancement systems (RES) (LOKKI, HIIPAKKA, 2001; BLASINSKI, KOCIŃSKI, 2023) were used in the listening tests. RES are recognized for their capacity to augment early reflections and modulate RT while preserving the room's intrinsic acoustic properties (BAKKER, GILLIAN, 2014) by employing digital signal processing technologies. Due to their greater controllability and precision, RES have become increasingly prevalent, and are gradually replacing traditional passive methods such as the use of absorbing panels (LOKKI, HIIPAKKA, 2001).

The Polish logatome test (described in Subsec. 2.3) was convolved with specific room impulse responses (RIRs) recorded from the three enclosures. A detailed description of the room acoustic data collection, equipment used, and methods for estimating the objective parameters determining the characteristics of the enclosures at different RES settings is provided by BLASINSKI, KOCIŃSKI (2023). Since these details are of secondary importance to the present research, they are omitted here. Sound samples were presented binaurally via Sennheiser HDA201 headphones connected to an SR46OH DOD preamplifier in a room compliant with American National Standard (1999). Prior to testing, the output sound-pressure levels of the headphones were calibrated using a Brüel & Kjaer 2203 level meter and a Brüel & Kjaer 4152 artificial ear. Throughout the 20-minute experiment, the speech level was maintained at 65 dB SPL at the eardrum. Each participant was presented with three phonetically balanced 50-element logatome lists convolved with RIRs from one of the three rooms under different RES settings. The logatomes were presented without any masking signal.

2.3. Listeners

Before the listening task, each of the 180 volunteers (60 per tested enclosure) underwent a hearing threshold examination conducted by a qualified specialist using a GSI 61 clinical audiometer and standard audiologic headphones (HDA200). The analysis of individual results, averaged across four frequencies (0.5 kHz, 1 kHz, 2 kHz, and 4 kHz), classified all par-

ticipants as having normal hearing according to World Health Organization (2021) criteria. None of the subjects reported any accompanying symptoms such as tinnitus or auditory hypersensitivity.

2.4. Test material

Describing and evaluating an enclosure's acoustic characteristics involves measurable acoustic parameters. Another source of information, as demonstrated in this study, is the assessment of SI using language-based tests. The results of subjective evaluations of speech transmission quality should, as far as possible, depend on the physical parameters of the communication channel being tested. Therefore, elimination of the information at the semantic level using lists of logatomes (nonsense words) (Brachmański, Dobrucki, 2021) is one of the recommended solutions, employed in this study. Logatomes are indeed more difficult for listeners, but also more reliable and robust, due to their low redundancy compared to, e.g., words, digits or sentences. Similar test material has been utilized in Polish studies, for example, by Brachmański (2021), who evaluated logatome-based SI transmitted through communication channels using the STI method.

Using nonsense words has the advantage of eliminating higher-level language processing that listeners use to understand words of degraded quality (Danhauer *et al.*, 1985). Consequently, the influence of cognitive association is limited, making hearing acuity more important than lexical prediction based on speech context or the participant's vocabulary. This approach results in decreased SI, measured as the percentage of correctly repeated words or sentences, because the amount of provided useful (meaningful) information is reduced (Stickney, Assmann, 2001). To avoid unnecessary listener fatigue that could impact speech perception while ensuring sufficient accuracy, lists of 50 or 100 logatomes are typically used (Howard, Angus, 2017). In this experiment, each RES setup was tested using a 50-element logatome list.

To ensure consistent assessment and minimize bias, all responses were evaluated by a single individual with expertise in SI research. Only accurately written logatomes (excluding spelling errors) were deemed correctly understood, employing binary word scoring as proposed by Kociński and Ozimek (2017).

2.5. Listening effort determination

In addition to the intelligibility test described above, all participants were asked to evaluate their perceived LE after each logatome list presentation. A 7-point Likert scale was used, ranging from 1 (no effort) to 7 (extreme effort), following methods similar to those used by Johnson *et al.* (2015).

3. Results

As was mentioned above, the most important and commonly used objective parameters used in room acoustics assessment are RT (RT30/RT20, EDT), C50, STI, and subjective SI. To investigate their influence on LE, we decided to compare LE scores with these objective measures. The calculated objective parameters and logatome intelligibility results are described in (Blasinski, Kociński, 2023). The averaged LE values, along with their standard deviations (in parentheses), are presented in Table 1.

Table 1. LE ratings in all room setups and across different RT20.

Enclosure	Setup	RT20 [s]	LE [1–7]
1	Setup 1	0.8 (0.3)	2.57 (1.18)
	Setup 2	1.1 (0.1)	2.90 (1.24)
	Setup 3	1.4 (0.1)	3.32 (1.45)
2	Setup 1	0.9 (0.0)	2.59 (1.18)
	Setup 2	1.2 (0.1)	3.18 (1.39)
	Setup 3	1.6 (0.1)	3.50 (1.50)
3	Setup 1	2.5 (0.1)	3.45 (1.48)
	Setup 2	3.1 (0.2)	3.43 (1.48)
	Setup 3	4.2 (0.2)	3.58 (1.58)

Figure 1 illustrates the relationship between RT (all nine values – three from each enclosure) and repor-
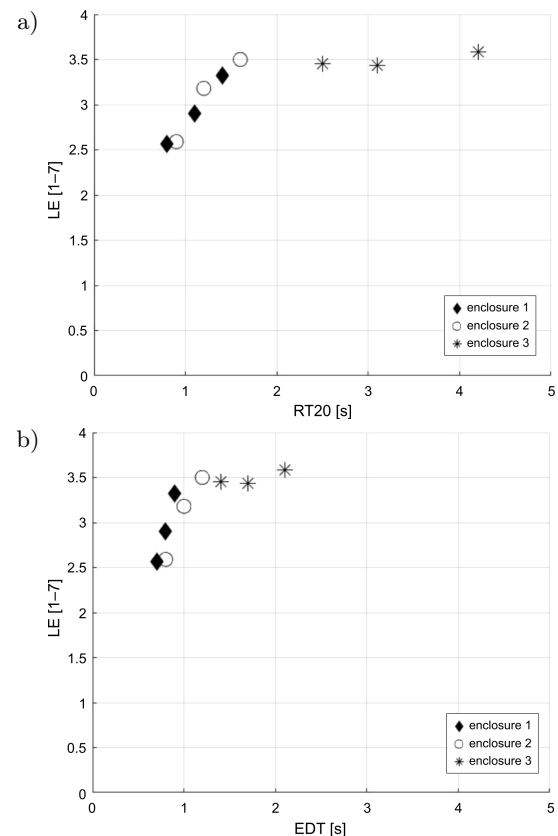

Fig. 1. LE as a function of (a) RT20 and (b) EDT.

ted LE. There is undoubtedly a trend where the reported LE increases with longer RT; however, a linear fit is not appropriate as $R^2 = 0.55$. A slightly stronger relationship is observed between LE and EDT, as indicated by an R^2 value of 0.56. Nevertheless, it still it is not sufficient to claim a linear relationship.

In this study, no significant correlation was found between LE and C50, as indicated by a Pearson correlation coefficient of 0.46 (Fig. 2). This suggests that variations in C50 do not reliably predict changes in LE.
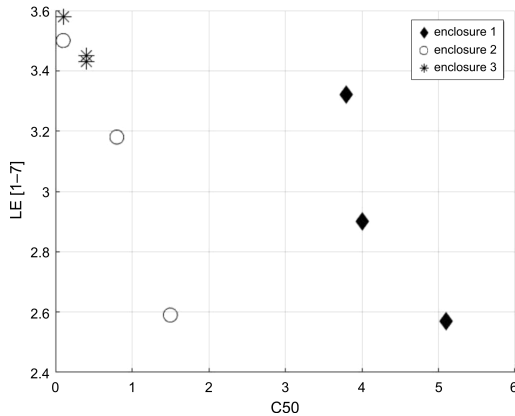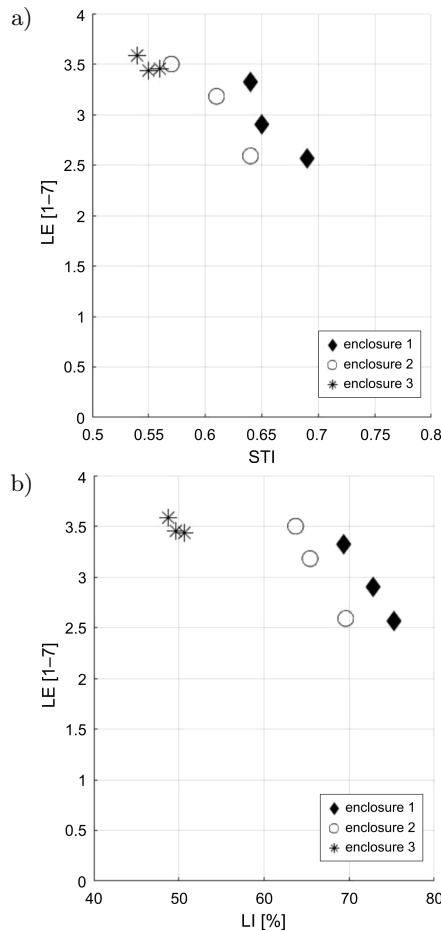

Fig. 2. LE as a function of C50.

Fig. 3. LE as a function of (a) STI and (b) LI.

In the final analysis, LE was compared with both objective and subjective measures (determined in tests with listeners) of SI.

Here, SI was defined as the percentage of correctly repeated logatomes from a 50-element list (as described in Subsec. 2.3). In contrast, the objective indicator used was the STI, which considers not only the acoustic characteristics of the enclosure but also the entire transmission channel (HOUTGAST, STEENEKEN, 1984). Figure 3 depicts LE as a function of objective intelligibility (STI) and measured logatome intelligibility (LI). It can be observed that for STI Pearson correlation coefficient is 0.74, while for LI it is only 0.55.

4. Discussion

This study compared objective room characteristics with the LE associated with understanding speech presented under conditions defined by these parameters. Given that:

- the logatom test is unaffected by the individual's lexicon knowledge (GIOVANNONE, THEODORE, 2021);
- the test items eliminate reliance on the context of the utterance;
- the entire test lasted less than 20 min, consistent with BRACHMAŃSKI and DOBRUCKI (2021), minimizing the effects of fatigue;
- all participants were of similar age (mean 23.5) and had normal hearing.

It is assumed that the only factor affecting SI is the characteristics of the propagation path (i.e., the enclosure) with potential influence of other non-auditory factors.

A slight trend of increasing mean LE values with higher RTs, along with a a correlation between higher LE and decreased LI, was observed. However, no statistical significance ($p > 0.05$) was found with R^2 values 0.56 for RT20 and 0.61 for LI. It is plausible that listeners overestimate their performance despite lower actual intelligibility levels. This tendency is particularly pronounced when using nonsense words, as listeners find it more challenging to assess their responses compared to semantically meaningful linguistic material. Another potential explanation for this lack of significance, and a limitation of this study, is that the 7-point scale used to measure LE may not be sensitive enough to detect differences across the range of RT20 values.

It is noteworthy that, although STI values were below 0.6 in environments with the longest RT20 (which typically corresponds to a 'fair' quality rating according to International Electrotechnical Commission (2020) standard), the LE scores did not significantly differ from those observed in environments with higher STI values. This suggests that, despite less favorable acoustic conditions, cognitive effort required was not

significantly influenced. While the limitations of using a 7-point scale to measure LE across a range of RT20 values are acknowledged, it is argued that eliminating subjective effort assessment from SI test batteries would be unjustified. Nonetheless, it is acknowledged that further research is needed to redefine the study protocol in this aspect.

These findings align with a broader body of research highlighting the complexity of LE and its multifactorial underpinnings (PICHORA-FULLER *et al.*, 2016; OHLENFORST *et al.*, 2017). Rather than aiming to redefine established models, this study contributes additional empirical data to an evolving and nuanced discussion within audiology and room acoustics. Such incremental research remains vital as the field continues to seek integrative frameworks that combine objective acoustical parameters with subjective listener experiences (MCGARRIGLE *et al.*, 2014; VISENTIN *et al.*, 2018).

For the three shortest RTs, reported effort falls slightly below a score of 3, whereas for all longer RTs, it edges slightly above 3 on a 7-point scale. This suggests that the effort required to comprehend logatomes remains relatively consistent across varying RTs.

Interestingly, LE correlates more strongly with the STI than with RT. However, due to the relatively small number of RES settings (nine in total), it would be premature to draw definitive conclusions. One possible reason behind this observation may be that STI is a more direct measure of intelligibility, encompassing a broader range of factors influencing speech signal comprehension. STI incorporates factors such as signal distortions and modulation, offering a more comprehensive assessment compared to the mere duration of sound persistence in a room after the source is switched off. It may be presumed that listeners subjectively assess their LE based on their overall SI, a metric better represented by STI than by RT.

Although statistical analysis reveals noticeable correlations between LE and various acoustic parameters, these relationships do not lend themselves easily to straightforward mathematical modeling. This is not entirely surprising, given the complex and inherently physical nature of LE. Still, the absence of a clearly defined functional relationship – such as that established between STI and LE or SI and LE – suggests that LE cannot yet be accurately predicted using objective parameters alone. This reinforces the idea that LE should be considered a separate and complementary descriptor when designing or adapting room acoustics. Its inclusion may help bridge the gap between objective acoustic indicators and the actual perceptual experiences of listeners.

While a linear fit can be applied to the relationship between declared LE and STI values, with Pearson's correlation coefficient reaching a significant R^2 value of 0.72, it cannot be assumed that cognitive effort in understanding speech in enclosures with varying acoustic characteristics, as defined by RES settings, can be estimated solely based on STI. Nevertheless, STI remains the measure most closely related to cognitive load, particularly LE. However, it is important to note that STI is not the only influencing factor. One should remember that until now it was assumed that one could rely on STI, with evaluations often limited to it (or other objective parameters). In some studies, additional measurements of SI with listeners were conducted (e.g., HODGSON, 2004). It turns out that the issue may be more complex, as indicated by results that are difficult to interpret clearly in terms of a specific functional relationship between predictors and LE. It should be noted that the relationship does not appear to be linear, but there is insufficient data to draw firm conclusions, necessitating further research. It is clear, however, that intelligibility alone is not sufficient, LE is also crucial. This effort can significantly influence the overall evaluation of a room. Consequently, a space might have good intelligibility and acoustic parameters indicating high quality, yet still be perceived poorly due to the high effort required from listeners.

Based on the obtained data, it is apparent that further research is essential. Although early attempts to consider the broad subjective experiences of listeners, beyond just physical measures and models, were made by BERANEK (2004) and ANDO (2007), it seems reasonable to address the issue of LE. This could involve developing specific tools and methods for its evaluation, incorporating a wider variety of system settings and possibly employing a different scale for rating LE (e.g., categorical or adaptive scales like ACALES (KRUEGER *et al.*, 2017)).

5. Conclusions

The presented findings reveal only a modest relationship between the objective acoustic parameters of the enclosures, as defined by the RES setting, and LE. Notably, there was lack of correlation with RT, commonly associated with SI but intricately linked to cognitive effort. Conversely, the strongest correlation coefficient was observed between LE and the STI, which may be attributed to the complexity inherent in this objective parameter. Over the past century, objective methods for room acoustics analysis have developed significantly. Numerous parameters have been established and are now considered standards widely used in measurements. However, it appears that the crucial role of the listener in the audience and the performer on stage in determining whether a room meets their acoustic expectations has been somewhat overlooked.

Given the ambiguous relationship between LE and standard acoustic parameters, it becomes evident that LE cannot be easily predicted or derived from existing objective metrics alone. This underscores its po-

tential as an independent and valuable indicator in comprehensive room acoustic assessment. Incorporating subjective measures such as LE into the design and adaptation processes enables the capture of perceptual aspects that objective measurements may overlook (Pichora-Fuller *et al.*, 2016; McGarrigle *et al.*, 2014). Ultimately, it is the human listener – not the abstract parameter set – who validates acoustic quality. Thus, including cognitive effort metrics provides a more representative evaluation of real-world listening conditions (Visentin *et al.*, 2018; International Organization for Standardization, 2018). This aligns with the broader shift in acoustic and auditory sciences, toward emphasizing not only what is measurable, but also what is perceptually meaningful (Oxenham, 2017; Rudner, *et al.*, 2012).

In the case of spoken performances, which this study focused on, it seems that parameters such as RT, C50, and STI do not predict the LE required to understand speech. This effort, however, may be critical in the overall evaluation of a room's acoustics, similar to how subjective preferences influence hearing aid users, which often determine their use despite objectively measured and adjusted parameters indicating improved auditory performance. Although the current data is limited, it is clear that intelligibility alone is insufficient – LE plays also a crucial role. This cognitive effort can significantly influence the overall evaluation of a room. Consequently, an enclosure may exhibit good intelligibility and favorable acoustic parameters indicating high quality, but still be perceived poorly due to the high effort listening it requires. Given the substantial impact of cognitive effort on listener experience, it is essential to consider metrics describing cognitive effort in the characterization of room acoustics, particularly for the presentation of speech signals.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Authors' contribution

Anna Pastusiak: conceptualization (equal), methodology (equal), data curation (equal), writing – original draft (lead). Łukasz Błasiński: conceptualization (equal), methodology (equal), data curation (equal), writing – review and editing (equal). Jędrzej Kociński: formal analysis (lead), writing – review and editing (equal), supervision (lead). All authors reviewed and approved the final manuscript.

Ethical approval

The study was conducted in accordance with the World Medical Association's Declaration of Helsinki. An informed consent was obtained from each participant. All data were anonymized prior to analysis.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Ahnert W., Tennhardt H.P. (2008), Acoustics for auditoriums and concert halls, [in:] *Handbook for Sound Engineers*, Ballou G.M. [Ed.], Focal Press, Waltham.

2. American National Standard (1999), *Maximum permissible ambient noise levels for audiometric test rooms* (Standard ANSI S3.1-1999 (R2003)).

3. Ando Y. (1998), *Architectural Acoustics: Blending Sound Sources, Sound Fields, and Listeners*, Springer New York, NY.

4. Ando Y. (2007), Concert hall acoustics based on subjective preference theory, [in:] *Springer Handbook of Acoustics*, Rossing T.D. [Ed.], pp. 351–386, Springer New York, NY.

5. Bakker R., Gillian S. (2014), The history of active acoustic enhancement systems, [in:] *Proceedings of the Institute of Acoustics*, **36**(Part 2).

6. Beranek L.L. (2004), *Concert Halls and Opera Houses: Music, Acoustics, and Architecture*, Springer.

7. Blasinski L., Kociński J. (2023), Perception of reverberation length in rooms with active acoustics enhancement systems, [in:] *Proceedings of the 10th Convention of the European Acoustics Association Forum Acusticum 2023*, pp. 1627–1634, https://www.doi.org/10.61782/fa.2023.0235.

8. Brachmański S. (2021), Test material used to assess speech quality in Poland, [in:] *Acoustics, Acoustoelectronics and Electrical Engineering*, Witos F. [Ed.], pp. 65–79, Wydawnictwo Politechniki Śląskiej, Gliwice.

9. Brachmański S., Dobrucki A. (2021), Impact of the level of noise and echo on the reaction time of listeners in the perception of logatoms, *Vibrations in Physical Systems*, **32**(2): 2021215-1–2021215-8, https://doi.org/10.21008/j.0860-6897.2021.2.15.

10. BRADLEY J.S. (1985), Uniform derivation of optimum conditions for speech in rooms, [in:] *Building Research Note* (no. BRN-239), https://doi.org/10.4224/40000478.

11. DANHAUER J.L., DOYLE P.C., LUCKS L. (1985), Effects of noise on NST and NU 6 Stimuli, *Ear & Hearing*, **6**(5): 266–269, https://doi.org/10.1097/00003446-198509000-00008.

12. EVEREST F.A. (2001), *The Master Handbook of Acoustics*, 4th ed., p. 153, McGraw Hill, New York.

13. FESTEN J.M., PLOMP R. (1990), Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing, *The Acoustical Society of America Journal*, **88**(4): 1725–1736, https://doi.org/10.1121/1.400247.

14. GIMÉNEZ A., CIBRIÁN R.M., CERDÁ S., GIRÓN S., ZAMARREÑO T. (2014), Mismatches between objective parameters and measured perception assessment in room acoustics: A holistic approach, *Building and Environment*, **74**: 119–131, https://doi.org/10.1016/j.buildenv.2013.12.022.

15. GIOVANNONE N., THEODORE R.M. (2021), Individual differences in lexical contributions to speech perception, *Journal of Speech, Language, and Hearing Research*, **64**(3): 707–724, https://doi.org/10.1044/2020_JSLHR-20-00283.

16. HERRMANN B., JOHNSRUDE I.S. (2020), A model of listening engagement (MoLE), *Hearing Research*, **397**: 108016, https://doi.org/10.1016/j.heares.2020.108016.

17. HODGSON M. (2004), Prediction of speech intelligibility in rooms – A comparison of five methods, *The Journal of the Acoustical Society of America*, **116**(4 Supplement): 2638, https://doi.org/10.1121/1.4785523.

18. HOUBEN R., VAN DOORN-BIERMAN M., DRESCHLER W.A. (2013), Using response time to speech as a measure for listening effort, *International Journal of Audiology*, **52**(11): 753–761, https://doi.org/10.3109/14992027.2013.832415.

19. HOUTGAST T., STEENEKEN H.J.M. (1973), The modulation transfer function in room acoustics as a predictor of speech intelligibility, *The Journal of the Acoustical Society of America*, **54**(2): 557, https://doi.org/10.1121/1.1913632.

20. HOUTGAST T., STEENEKEN H.J.M. (1984), A multilanguage evaluation of the RASTI-method for estimating speech intelligibility, *Acustica*, **54**(4): 185–199.

21. HOWARD D.M., ANGUS J.A.S. (2017), *Acoustics and Psychoacoustics*, 5th ed., Routledge, Oxfordshire.

22. International Electrotechnical Commission (2020), *Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index* (Standard no. IEC 60268-16:2020).

23. International Organization for Standardization (2018), *Acoustics – Soundscape. Part 2: Data collection and reporting requirements* (ISO/TS Standard No. 12913-2:2018), https://www.iso.org/standard/75267.html.

24. JOHNSON J., XU J., COX R., PANDERGRAFT P. (2015), A comparison of two methods for measuring listening effort as part of an audiologic test battery, *The Journal of the Acoustical Society of America*, **24**(3): 419–431, https://doi.org/10.1044/2015_AJA-14-0058.

25. JUST M.A., CARPENTER P.A. (1992), A capacity theory of comprehension: Individual differences in working memory, *Psychological Review,* **99**(1): 122–149, https://doi.org/10.1037/0033-295X.99.1.122.

26. KOCIŃSKI J., OZIMEK E. (2017), Logatome and sentence recognition related to acoustic parameters of enclosures, *Archives of Acoustics*, **42**(3): 385–394, https://doi.org/10.1515/aoa-2017-0040.

27. KRUEGER M., SCHULTE M., BRAND T., HOLUBE I. (2017), Development of an adaptive scaling method for subjective listening effort, *The Journal of the Acoustical Society of America*, **141**(6): 4680–4693, https://doi.org/10.1121/1.4986938.

28. KUHL W. (1954), About trying to find the best reverberation time for big music studios [in German], *Acta Acustica United with Acustica*, **4**(5): 618–634.

29. KUTTRUFF M. (2009), *Room Acoustics*, 5th ed., Spon Press, London.

30. LEMKE U., BESSER J. (2016), Cognitive load and listening effort: Concepts and age-related considerations, *Ear & Hearing,* **37**(1): 77S–84S, https://doi.org/10.1097/AUD.0000000000000304.

31. LOKKI T., HIIPAKKA J. (2001), A time variant reverberation algorithm for reverberation enhancement systems, [in:] *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-01)*, pp. 28–32.

32. MARSHALL L.G. (1994), An acoustic measurement program for evaluating auditoriums based on the early/late sound energy ratio, *The Journal of Acoustical Society of America*, **96**(4): 2251–2261, http://doi.org/10.1121/1.410097.

33. MATTYS S.L., DAVIS M.H., BRADLOW A.R., SCOTT S.K. (2012), Speech recognition in adverse conditions: A review, *Language and Cognitive Processes*, **27**(7–8): 953–978, https://doi.org/10.1080/01690965.2012.705006.

34. MCGARRIGLE R. *et al.* (2014), Listening effort and fatigue: What exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group 'white paper', *International Journal of Audiology*, **53**(7): 433–440, https://doi.org/10.3109/14992027.2014.890296.

35. MEYER J. (1978), *Acoustics and the Performance of Music*, Frankfurt/Main: Verlag das Musikinstrument, Frankfurt/Main.

36. OHLENFORST B. *et al.* (2017), Effects of hearing impairment and hearing aid amplification on listening effort: A systematic review, *Ear & Hearing*: **38**(3), 267–281, https://doi.org/10.1097/AUD.0000000000000396.

37. OXENHAM A.J. (2017), How we hear: The perception and neural coding of sound, *Annual Review of Psychology*, **69**: 27–50, http://doi.org/10.1146/annurev-psych-122216-011635.

38. Peelle J.E. (2018), Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior, *Ear & Hearing*, **39**(2): 204–214, https://doi.org/10.1097/AUD.0000000000000494.

39. Pichora-Fuller M.K. *et al.* (2016), Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL), *Ear & Hearing*, **37**(1): 5S–27S, https://doi.org/10.1097/AUD.0000000000000312.

40. PN-EN ISO (2010), *Acoustics – Measurement of room acoustic parameters. Part 1: Reverberation time in ordinary rooms* [in Polish: *Akustyka – Pomiar parametrów akustycznych pomieszczeń. Część 1: Czas pogłosu w zwyczajnych pomieszczeniach*] (PN-EN ISO 3382-1:2010).

41. Rudner M., Lunner T., Behrens T., Thorén E.S., Rönnberg J. (2012), Working memory capacity may influence perceived effort during aided speech recognition in noise, *Journal of the American Academy of Audiology*, **23**(8): 577–589, http://doi.org/10.3766/jaaa.23.7.7.

42. Sabine W.C. (1922), *Collected Papers on Acoustics*, p. 279, Cambridge (MA), Harvard University Press.

43. Sakai H., Ando Y., Setoguchi H. (2000), Individual subjective preference of listeners for vocal music sources in relation to the subsequent reverberation time of sound fields, *Journal of Sound and Vibration*, **232**(1): 157–169, https://doi.org/10.1006/jsvi.1999.2691.

44. Shinn-Cunningham B.G., Best V. (2008), Selective attention in normal and impaired hearing, *rends in Amplification*, **12**(4): 283–299, https://doi.org/10.1177/1084713808325306.

45. Schmid P.M., Yeni-Komshian G.H. (1999), The effects of speaker accent and target predictability on perception of mispronunciations, *Journal of Speech, Language, and Hearing Research*, **42**(1): 56–64, https://doi.org/10.1044/jslhr.4201.56.

46. Stickney G.S, Assmann P.F. (2001), Acoustic and linguistic factors in the perception of bandpass-filtered speech, *The Journal of the Acoustical Society of America*, **109**(3): 1157–1165, https://doi.org/10.1121/1.1340643.

47. Strand J.F., Brown V.A., Merchant M.B., Brown H.E., Smith J. (2018), Measuring listening effort: Convergent validity, sensitivity, and links with cognitive and personality measures, *Journal of Speech, Language, and Hearing Research*, **61**(6): 1463–1486, https://doi.org/10.1044/2018_JSLHR-H-17-0257.

48. Visentin C., Prodi N., Cappelletti F., Torresin S., Gasparella A. (2018), Using listening effort assessment in the acoustical design of rooms for speech, *Building and Environment*, **136**: 38–53, https://doi.org/10.1016/j.buildenv.2018.03.020.

49. Wang Y. *et al.* (2018), Relations between self-reported daily-life fatigue, hearing status, and pupil dilation during a speech perception in noise task, *Ear & Hearing*, **39**(3): 573–582, https://doi.org/10.1097/aud.0000000000000512.

50. Wendt D., Koelewijn T., Książek P., Kramer S.E., Lunner T. (2018), Toward a more comprehensive understanding of the impact of masker type and signal-to-noise ratio on the pupillary response while performing a speech-in-noise test, *Hearing Research*, **369**: 67–78, https://doi.org/10.1016/j.heares.2018.05.006.

51. Wild C.J., Yusuf A., Wilson D.E., Peelle J.E., Davis M.H., Johnsrude I.S. (2012), Effortful listening: The processing of degraded speech depends critically on attention, *Journal of Neuroscience*, **32**(40): 14010–14021, https://doi.org/10.1523/JNEUROSCI.1528-12.2012.

52. World Health Organization (2021), World report on hearing, https://www.who.int/publications/i/item/9789240020481 (access: 08.05.2025).

53. Zekveld A.A., Kramer S.E., Festen J.M. (2010), Cognitive load during speech perception in noise: The influence of age, hearing loss, and cognition on the pupil response, *Ear & Hearing*, **32**(4): 498–510, https://doi.org/10.1097/AUD.0b013e31820512bb.

Research Paper

Comparative Study of the Acoustic Efficiency of Prototype Sound Absorbing Panels Used in the Railway Track

Grzegorz KLEKOT[(1)], Mariusz WĄDOŁOWSKI[(1)*],
Cezary KRAŚKIEWICZ[(2)], Artur ZBICIAK[(2)]

[(1)] *Faculty of Automotive and Construction Machinery Engineering, Warsaw University of Technology*
Warsaw, Poland

[(2)] *Faculty of Civil Engineering, Warsaw University of Technology*
Warsaw, Poland

*Corresponding Author e-mail: mariusz.wadolowski@pw.edu.pl

The article presents the results of comparative studies concerning the efficiency of systems aimed at minimising the acoustic nuisance of noise generated by the railway vehicle movement. The issue of noise in railway traffic is a significant challenge, affecting both human health and the quality of life in the vicinity of railway lines. Prototype sound absorbing panels with varied surface geometry, a rubber slab, and ballast layer (stone aggregate, grain size 31.5/50 mm) were examined. Experiments were conducted in a reverberation chamber, analysing the response to broadband noise excitation. The reverberation chamber allows for obtaining repeatable results, eliminating the influence of external sound sources. It enables the assessment of the sound absorption properties of various materials which makes it possible to determine their effectiveness in noise reduction. The research methodology included measurements of reverberation time in different frequency bands for an empty chamber and a chamber containing the tested materials. The obtained differences in reverberation times provide information on the influence of the tested material on the distribution of acoustic energy in individual frequency bands. The research results allow for a preliminary assessment of the effectiveness of the tested materials in the task of reducing railway line noise.

Keywords: railway traffic noise; noise reduction; acoustic insulation; sound-absorbing materials; vibration damping; reverberation chamber.

1. Introduction

Currently, transport is a fundamental element of the global economy, serving as a key factor enabling the worldwide flow of goods and services. Modern producers and consumers prefer just-in-time delivery instead of stockpiling, which allows for the reduction of storage costs and increases the efficiency of logistic processes. This requires transport to be reliable, punctual, and economically justified. Rail transport plays a significant role in the implementation of both freight and passenger transport over medium and long distances. The dynamic development of high-speed rail positions

it as a serious competitor to air transport over medium distances. The main advantages of rail transport include high load capacity, reliability, low accident rates, and minimal sensitivity to changing weather conditions. However, one of the key challenges remains the problem of noise emission, which negatively affects the natural environment and human health. Railway noise can lead to chronic stress, sleep disturbances, and prolonged exposure to high noise levels may increase the risk of hypertension and other cardiovascular problems (BASNER *et al.*, 2014; MÜNZEL *et al.*, 2017). Moreover, railway noise adversely affects fauna, disrupting the natural behavioural patterns of animals

and contributing to a decrease in biological diversity in areas adjacent to railway lines.

Studies conducted in Europe have shown that noise generated by rail transport is one of the most burdensome sources of transport noise (Licitra *et al.*, 2016; Zhang *et al.*, 2019). These studies were based on measurements of noise levels at various locations along railway lines and on the analysis of the impact of this noise on residents of neighbouring areas. Both field data and surveys on people's perception of noise nuisance were used, which allowed for a comprehensive assessment of its impact.

In response to this problem, legal regulations have been developed to determine acceptable noise levels for different categories of areas. Depending on their purpose, these areas have been classified into four groups: spa areas (places of particular health significance, requiring the highest level of acoustic protection), single-family residential areas (characterised by low population density, where a moderate noise level is required), multi-family residential areas (with higher building density, requiring more flexible acoustic norms), and urban areas (areas with high traffic intensity, where due to the nature of activities, higher noise levels are acceptable). The maximum permissible noise level for spa areas is 50 dB(A), while in urban areas, this value is 68 dB(A).

Therefore, the aim of this article is to conduct a comparative study of prototype sound-absorbing element-rubber slabs, ballast layers, and specially contoured concrete panels, and to assess their effectiveness in reducing railway noise. The novelty of our approach lies in applying an uncommon, yet straightforward measurement method based on the difference in reverberation time (RT30). By comparing how each material influences RT30 under identical test conditions, we can rapidly evaluate and contrast their noise-attenuating properties. This direct and simple use of reverberation-time differences is rarely encountered in railway noise analyses, making our study particularly valuable for future track design. In this way, we offer new insights into how material selection and surface geometry can be optimized to meet modern standards of acoustic performance.

2. Sources of noise and its propagation

When analysing the problem of reducing noise generated by railway traffic, various factors affecting its level must be considered (Cao *et al.*, 2018). The most important factors include train speed, the design of locomotives and carriages, the type and condition of the track, the insulating materials used, and atmospheric conditions. Aerodynamic noise becomes the dominant source at train speeds above 250 km/h because, as speed increases, the airflow around the train generates a significant amount of turbulence, leading to

more intensive generation of sound waves. The design of the locomotive's front, the carriage connections and the speed of travel contributed to its creation. Another significant source of noise is the locomotive's drive system. The main noise generating components include traction motors, gearboxes (Andrés *et al.*, 2021), and cooling systems. Noise levels can be reduced by using sound-absorbing shields, optimising gearbox design, and modernising cooling systems to operate more quietly and efficiently. Sound-reducing shields include rubber mats and mineral fibre panels. According to (Leštinský, Zvolenský, 2019), optimising the design of carriage bogies and braking systems can reduce noise levels by up to 15 dB. Modernisation of railway infrastructure, including rail grinding and the installation of vibration dampers, also contributes to reducing noise emissions (Zvolenský *et al.*, 2017). Vibration dampers are devices mounted on rails that help reduce rail vibrations. They absorb vibration energy, limiting its transmission to other parts of the track and the surroundings, consequently reducing generated noise. Kraśkiewicz *et al.* (2024) investigated possible applications of rubber granulate SBR (styrene-butadiene rubber) produced from recycled waste tires as an elastic cover for prototype rail dampers. The authors performed laboratory tests, with a focus on their operational durability. Rail grinding removes surface irregularities, which lowers the generation of vibrations and wheel-rail friction, ultimately reducing the level of generated noise.

3. Noise reduction

There are many methods of reducing noise associated with railway traffic, which are based on changes in track construction, track geometry, and the materials used. An example of an effective method is the use of elastic track pads, which reduce vibrations transmitted through the tracks and thus reduce noise. Elastic track pads have been successfully implemented in projects in Germany and Switzerland (Němec *et al.*, 2020), where their implementation contributed to a significant improvement in acoustic comfort in urban areas. Elastic track pads are made of materials such as polyurethane or rubber mixtures, which provide appropriate elasticity, enabling effective vibration damping.

In the case of new railway lines, significant noise reduction, even up to 25 dB, can be achieved by building earth embankments on both sides of the tracks (Bunn, Zannin, 2016). However, constructing such embankments involves high costs and the need for a large amount of space, which may limit their use in densely built-up areas. Additionally, embankments require appropriate land development, which is not always possible in cities and urban areas. These embankments act as natural sound barriers, absorbing and dispersing sound waves, reducing their propagation to-

wards the surroundings. However, such infrastructure requires a significant amount of space, which often limits its application in urban areas (Sun *et al.*, 2019; 2020). In such cases, alternatives can be smaller acoustic barriers, green walls, or sound-absorbing screens, which can be more easily adapted to urban conditions. Examples of materials used in sound-absorbing screens include mineral fibre panels, concrete-rubber composites, and acrylic glass, which effectively reflect and disperse sound waves, limiting their propagation. An alternative approach involving trackbed sound absorptive panels for reducing wayside noise on slab track was examined by Glickman *et al.* (2011), who discuss both the benefits and practical considerations of using porous concrete elements in lieu of conventional ballasted track. There are also examples of combining multiple complementary track-related solutions, such as sound-absorbing panels, low acoustic barriers, and rail dampers. For instance, a study conducted on an LRT line in Athens (Vogiatzis, Vanhonacker, 2016) demonstrated that the simultaneous use of absorbing panels, low barriers, and rail dampers resulted in noise reduction reaching of up to 10 dB(A), significantly surpassing the initial goal of 6 dB(A). This 'multi-pronged' approach serves as an inspiring example for projects carried out in urban areas, where minimizing noise levels as much as possible is crucial while maintaining minimal interference with existing infrastructure.

4. Experimental study

The research was conducted to assess the effectiveness of prototype subgrade elements in reducing the acoustic nuisance of railway lines. Understanding the sound-absorbing properties of various structures is crucial for optimising noise reduction in railway infrastructure.

A reverberation chamber was used for the study, which allows for easy comparison of sound absorption by different materials and structural elements (Castiñeira-Ibáñez *et al.*, 2012). The diffuse, homogeneous acoustic field in the reverberation chamber enables repeatable results, leading to reliable comparisons in the assessment of materials (Szczepański *et al.*, 2023) and elements shaping the acoustic properties of the environment. The reverberation chamber eliminates the influence of external factors (reflections of sound waves from random objects, uncontrolled noise sources, etc.), allowing for repeatable measurement of the reverberation time RT30. The experimental setup enabled comparisons in the frequency domain (in ⅓ octave bands).

In the presented study, all tested samples occupied the same surface area 1.5 m × 1.3 m on the floor of the reverberation chamber, ensuring comparability of measurement results. Despite differences in shapes and ma-

terials (e.g., trapezoidal grooves, half-round grooves, rubber slabs, or ballast layers), each sample was arranged to cover an identical area. This approach allows us to focus on the intrinsic acoustic properties of each tested variant without bias due to uneven surface coverage. Consequently, we did not introduce detailed external references in this section, concentrating instead on our own measurement protocol. By measuring the RT30 in ⅓ octave bands for each sample and comparing it with the empty chamber baseline, we isolated the influence of material and geometry on sound absorption and dispersion, rather than variations in the sample size.

The tested sound absorbing panels (Fig. 1) were made of porous concrete, whose recipe was marked with the symbol '220/10' – number '220' refers to the volume of cement grout ($220\,\mathrm{dm^3}$), and number '10' indicates the percentage of sand in the crumb pile. The concrete recipe was elaborated within laboratory tests, and the surface grooving was designed using numerical simulations. All tested elements were produced in the laboratory of the Faculty of Civil Engineering at the Warsaw University of Technology (Kraśkiewicz *et al.*, 2024). A recent study by Zhao *et al.* (2014) likewise confirmed that porous sound-absorbing concrete slabs can effectively reduce railway noise by up to 4 dB at speeds around 200 km/h, highlighting the importance of careful aggregate gradation and fibber content selection. A similar approach to optimizing the mix design of porous, sound-absorbing blocks in urban train tunnels was proposed by Lee *et al.* (2016), emphasizing the role of lightweight aggregates and structural requirements for efficient noise mitigation.

Fig. 1. Sound absorbing panels with dimensions
500 mm × 500 mm × 100 mm.

Two types of panels, stone aggregate, and panel rubber were considered:

– panel 1 – porous concrete panels with trapezoidal grooves: 500 mm × 500 mm × 100 mm (Fig. 2);
– panel 2 – porous concrete panels with half-round grooves: 500 mm × 500 mm × 100 mm (Fig. 3);
– rubber: 1500 mm × 1300 mm × 150 mm (Fig. 4);
– ballast layer of 31.5/50 mm stone aggregate (track bed), covering an area of 1500 mm × 1300 mm × 150 mm (Fig. 5).
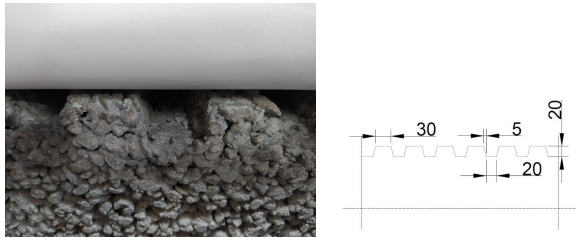
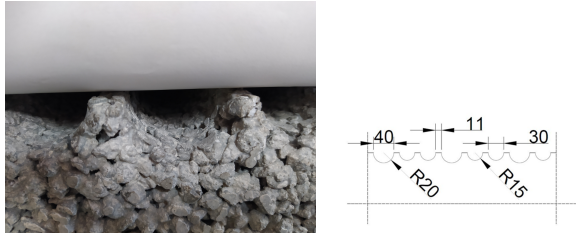Fig. 2. Panels type 1 (trapezoidal grooves).

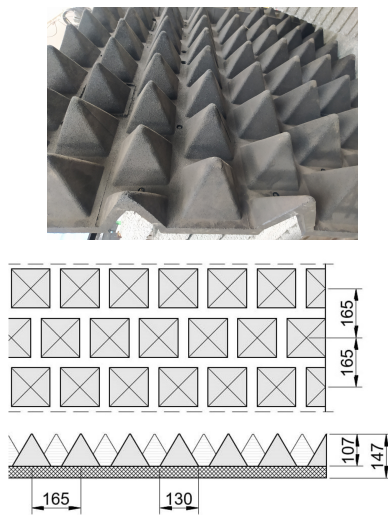Fig. 3. Panels type 2 (half-round grooves).

Fig. 4. Rubber panel.

Fig. 5. Ballast layer (stone aggregate,
grain size 31.5/50 mm).

Concrete slabs (sound absorbing panels) are characterised by high mass and stiffness, allowing them to effectively reduce vibrations and absorb acoustic energy. Rubber panels exhibit excellent sound-damping properties due to their elasticity, which enables the absorption of acoustic waves. Stone aggregate, owing to its varied shapes and porosity, can effectively scatter

and dampen sound waves. The selection of these elements is due to their wide applicability in railway track constructions and potential acoustic benefits. Concrete panels have a high mass, which can influence vibration reduction, rubber panels have sound-absorbing properties, and stone aggregate is commonly used in tracks due to its ability to dampen vibrations (FEDIUK *et al.*, 2021; SHAHIDAN *et al.*, 2017).

All measurements were conducted in the reverberation chamber; the reverberation time RT30 of the chamber with the tested samples was measured. Measuring and analytical equipment from Brüel & Kjær was used, consisting of an analyser and PULSE LabShop software. The chamber with the tested elements, the sound source, and the measurement microphones are shown in Fig. 6.

Fig. 6. Test setup in the reverberation chamber.

The test acoustic signal was generated by a white noise generator using B&K LabShop software, specifically the 'Reverberation Time' application, with an omnidirectional sound source BK 4292. White noise is characterised by a uniform distribution of energy across all frequency bands, allowing for a relatively comprehensive assessment of the acoustic properties of materials. The omnidirectional sound source enables even distribution of sound waves in the chamber, ensuring repeatable measurement results. Four BK 4189 microphones, arranged inside the chamber, recorded changes in acoustic pressure. The reverberation time was measured in standardised $1/3$ octave bands for random noise (so-called white noise) in the range from 50 Hz to 8 kHz. Changes in acoustic pressure were observed after switching off the sound source, and based on this, the RT30 was determined. The difference between the reverberation time of the empty reverberation chamber and the chamber with samples illustrates the effectiveness of the tested system, allowing for the assessment of how well a given material can absorb acoustic energy. A reduction in reverberation time indicates better sound-absorbing properties of the tested material. During the experiments, three series of five measurements were carried out with different positions of the sound source (the locations of the microphones and the sound source are schematically pre-

sented in Fig. 7). Each microphone recorded 11-second time histories. Based on them, reverberation times in $^1/_3$ octave bands were averaged. Then, the results for all series were averaged. Measurement results are presented in the form of the third-octave spectrum of the average reverberation time and the difference between the reverberation time for the empty chamber and with samples.
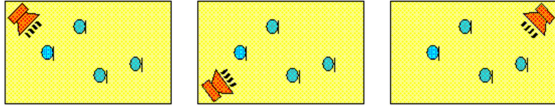

Fig. 7. Schematic of the measurement setup.

Third-octave spectra ($^1/_3$ octave bands) were obtained using the BK Connect analysis software. This research method allows for comparing the sound absorption by individual sound-absorbing structures.

The RT30 is determined based on a drop in the sound level of 30 dB. This parameter is primarily used in analyses of room acoustic properties (e.g., concert halls, recording studios, cinemas, etc.); nevertheless, the described methodology also allows for inferring the suitability of subgrade elements for reducing railway line noise.

5. Measurement results

The research results indicate that the differences in reverberation time between the empty chamber and the chamber with the tested samples provide significant information about the acoustic effectiveness of the materials tested. Figure 8 presents the reverberation time values measured in the empty reverberation chamber in third-octave bands. The longest reverberation time, approximately 4 s, was observed for the band with a centre frequency of 630 Hz.

Then, the reverberation times of the chamber in which the tested materials were placed were measured and the difference in reverberation time between the empty chamber and the chamber filled with the samples was calculated; the greater the ability of the tested material to absorb sound, the greater the difference. Which is crucial for applications in the context of railway infrastructure. These results can be used to optimise the selection of construction materials for more effective noise minimisation. Figure 14 presents a summary of the measurements.

The measurements were performed for the following types of samples: detailed measurement results are presented in Figs. 9–13. The graphs contain the reverberation time measured in the chamber with the tested samples and the path and graph of the calculated time difference between the empty chamber and the chamber with the samples.

The rubber panel were the most effective in attenuating sounds at low and medium frequencies (Fig. 14). The difference in reverberation time in the chamber with such panels for bands 125 Hz–400 Hz was up to 1 s. Concrete sound absorbing panels with specific surface configurations reduced the reverberation time to 1.9 s for frequency ranges 400 Hz–2000 Hz. Stone aggregate
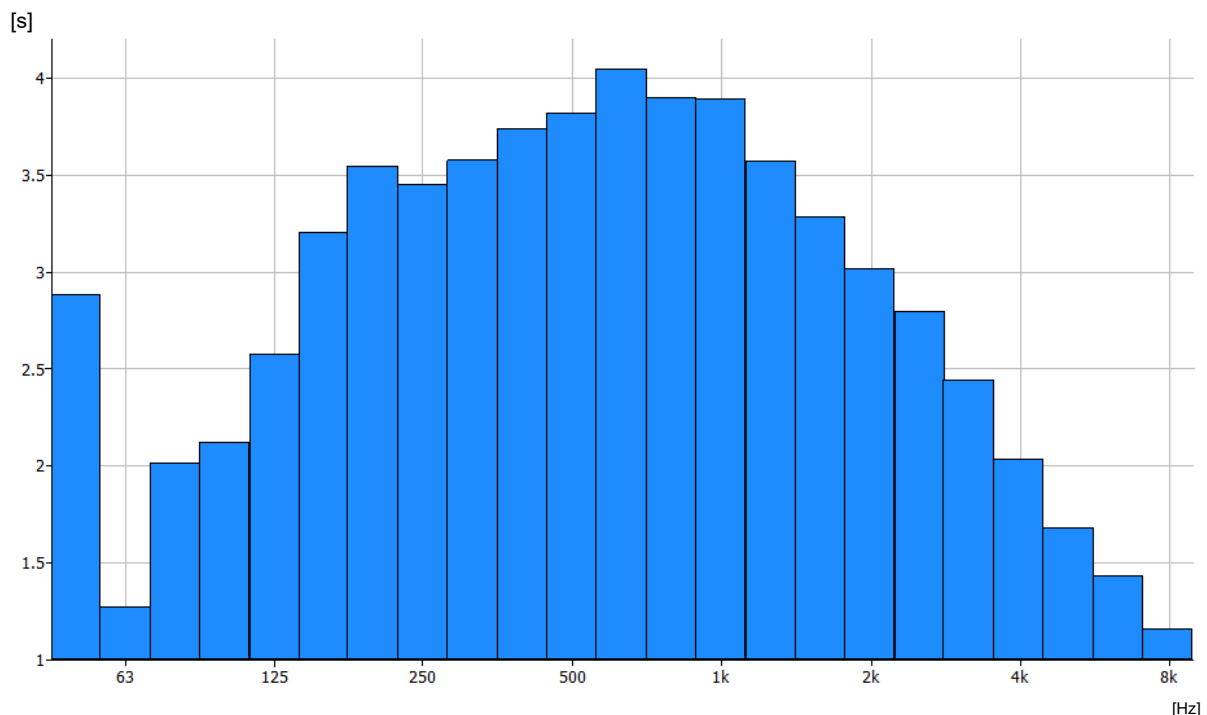
Fig. 8. Reverberation time of the empty chamber.
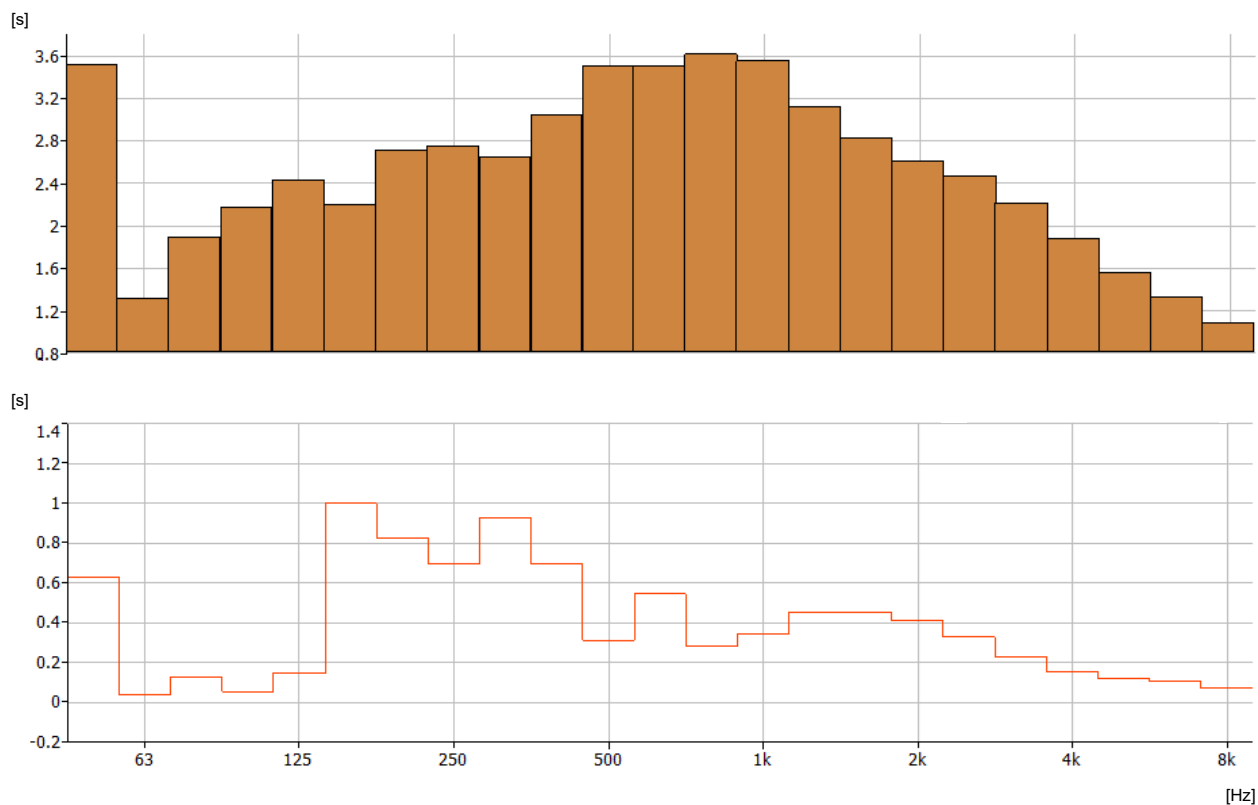
Fig. 9. Reverberation time – rubber, sample 1 (top); differences in reverberation time between an empty chamber and a chamber with tested samples (bottom).

Fig. 10. Reverberation time – panel type 1, sample 2 (top); differences between an empty chamber and a chamber with tested samples (bottom).

Fig. 11. Reverberation time – panels type 1 and 2, sample 3 (top); differences between an empty chamber and a chamber with tested samples (bottom).

Fig. 12. Reverberation time – stone aggregate, sample 4 (top); differences between an empty chamber and a chamber with tested samples (bottom).

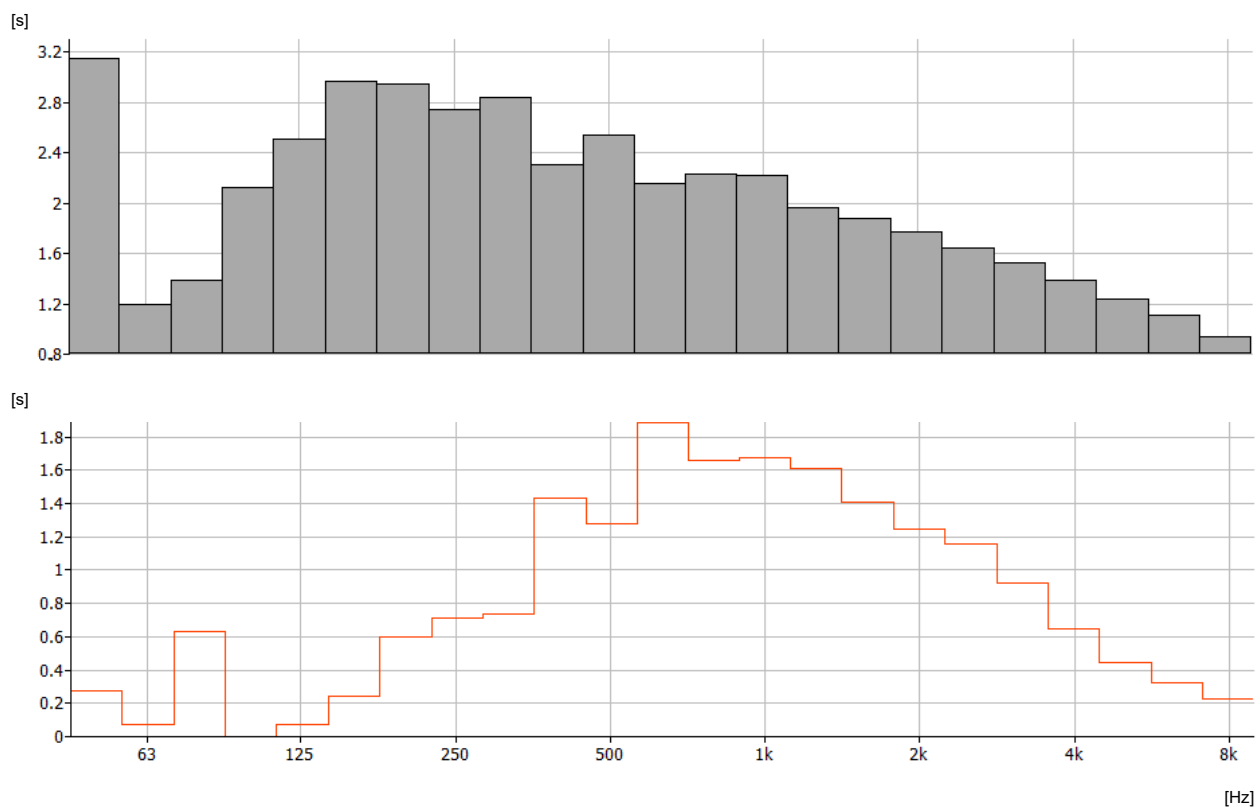Fig. 13. Reverberation time – panel type 2, sample 5 (top); differences between an empty chamber and a chamber with tested samples (bottom).
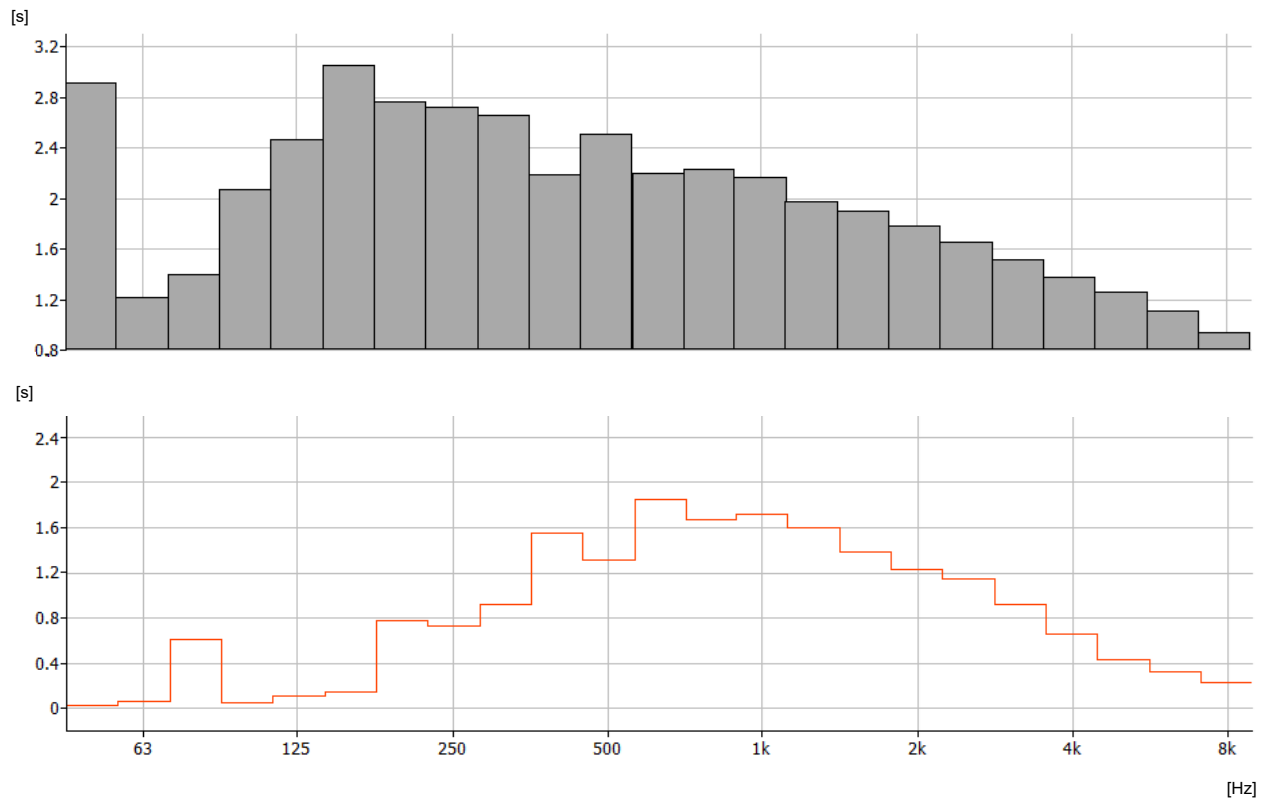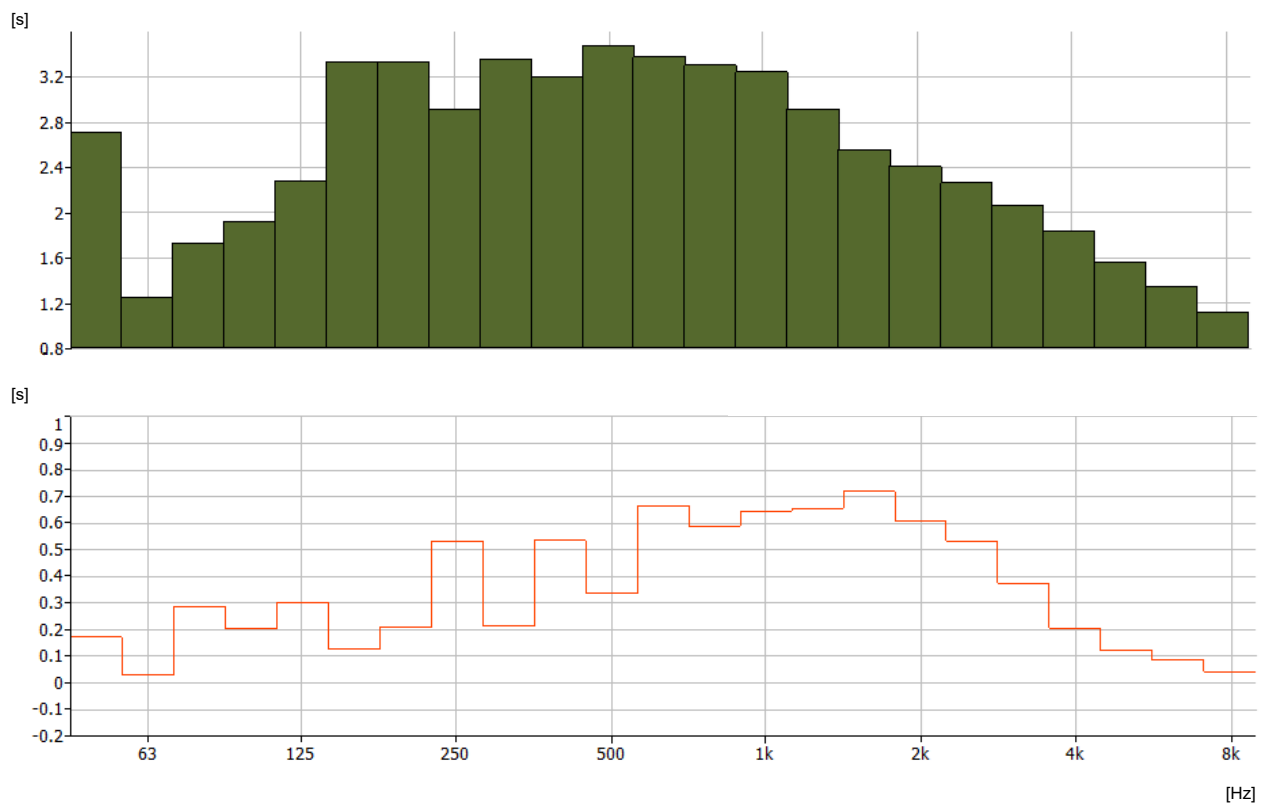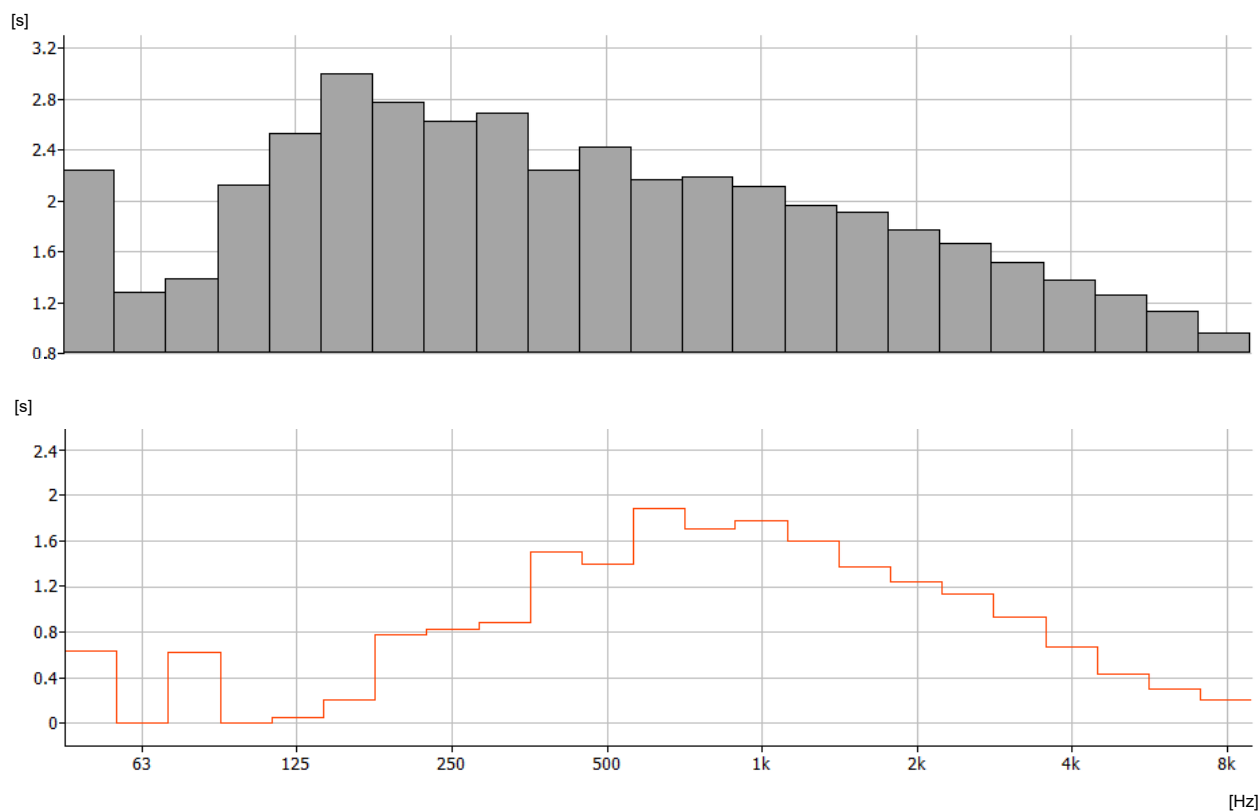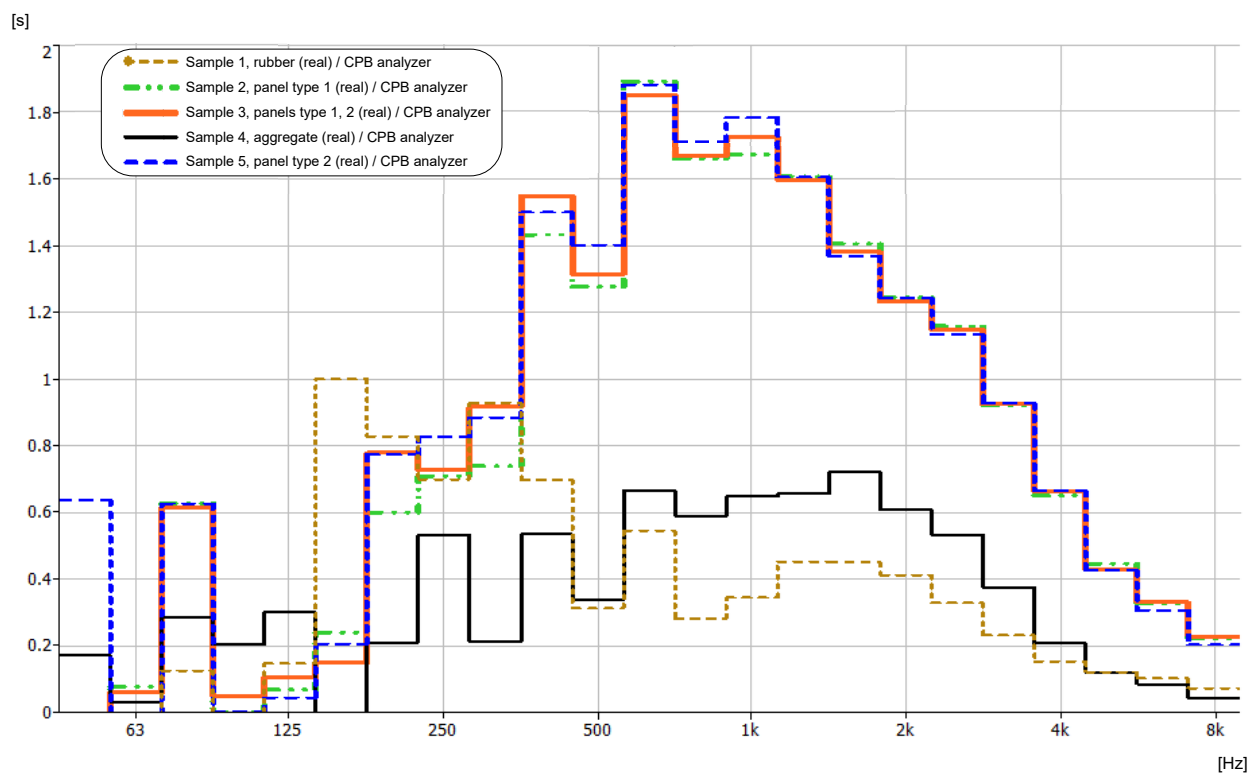

Fig. 14. Differences in reverberation time between an empty chamber and a chamber with tested samples in $^1/_3$ octave bands.

caused an average decrease in reverberation time of about 0.5 s across a wide band of tested frequencies.

The obtained results indicate that rubber panels were particularly effective in attenuating sounds in low and medium frequencies, while concrete panels are more effective in higher ranges. This analysis suggests that combining both materials could yield even better results, providing effective noise reduction over a wider frequency range. Combining these two materials may be particularly beneficial in applications where effective noise reduction is required under various acoustic conditions – for example, in places with high railway traffic intensity, where it is important to limit both low and high components of noise. For instance, railway stations located in residential areas could significantly benefit from such a solution, minimising both low-frequency vibrations and sharp high-frequency sounds generated by passing trains.

Although rubber panels demonstrated high sound attenuation in crucial frequency bands for human hearing, one must note that rubber-based solutions can be more costly compared to concrete or aggregate-based panels. Additionally, rubber may pose certain challenges in tunnel applications due to stricter fire safety regulations and limited heat resistance. Consequently, the decision to implement such materials must balance acoustic performance, economic feasibility, and safety requirements. For instance, in designing 'quiet' railway tunnels, where controlling reverberation is paramount, one might favour materials that effectively disperse acoustic waves while meeting fire protection standards. This highlights why alternative solutions – such as specifically shaped concrete panels or hybrid systems using aggregate – also deserve further investigation. They can offer a balanced compromise between acoustic efficiency, structural constraints, and compliance with tunnel safety regulations.

It is also crucial to note that noise reduction depends not only on the absorption of acoustic waves but equally on their dispersion. Materials with specific surface geometries can scatter energy effectively, reducing the perceptible noise level. In enclosed spaces such as tunnels, where sound reflections are intensified, geometry-induced dispersion can be as significant as direct absorption. Therefore, further research on various panel shapes (e.g., trapezoidal grooves, half-round hollows, or more complex fractal patterns) may provide solutions that combine high dispersion capacity with practical considerations such as fire safety or cost-effectiveness.

When selecting materials for full-scale application, economic factors play a vital role. Rubber-based solutions, although acoustically effective, can be more expensive to install and maintain than concrete panels. Additionally, their long-term durability in harsh conditions (e.g., high temperatures, tunnels) may raise further concerns. Hence, research aimed at balancing

performance, cost, and safety is necessary before deciding on the large-scale implementation of any chosen material.

6. Discussion

The results presented in the previous section indicate that rubber-based panels provided the highest level of sound attenuation in the low- and mid-frequency ranges, while the specially shaped concrete panels performed better in higher frequency bands. The ballast layer (stone aggregate), in turn, showed moderate effectiveness across a broad range of frequencies, though not as pronounced as rubber's performance in the low-mid domain.

From a practical standpoint, it is crucial to balance acoustic performance with cost, durability, and safety. Despite its strong noise reduction properties, rubber can be more expensive to produce and maintain and may pose challenges in high-temperature or fire-critical environments (e.g., tunnels). On the other hand, properly contoured concrete panels can effectively disperse acoustic waves and are often simpler to adapt to fire protection standards. The aggregate-based layer remains a standard solution in conventional track structures and can be further optimized by tailoring its geometry or combining it with other materials.

A key insight emerging from these tests is the importance of surface geometry in scattering acoustic energy. In enclosed spaces, such as railway tunnels, geometry-driven dispersion can be as significant as the direct absorption. Future efforts could therefore investigate hybrid solutions blending different materials with complementary frequency attenuation strengths, while also meeting cost and safety requirements.

Moreover, numerical simulations (for instance, the finite element method (FEM) analyses) could complement the experimental approach, allowing for refinement of material selection and panel geometry without the need for large-scale prototypes. Although the present study focused on a straightforward methodology based on reverberation time differences (RT30), integrating empirical data with simulation results may provide a more comprehensive foundation for designing next generation 'quiet' railway structures.

7. Conclusion

The conducted comparative studies in the reverberation chamber allowed for the analysis of the acoustic efficiency of various materials for use in railway infrastructure. The results of these studies may contribute to increasing acoustic comfort in the vicinity of tracks and support the optimisation of sleeper designs and other track elements. Identifying the most effective noise-damping materials is fundamentally impor-

tant for reducing the acoustic nuisance generated by railway traffic.

The use of a reverberation chamber and the measurement of the difference in reverberation time between the empty chamber and the chamber with samples constitutes an innovative and practical approach, especially in the context of research on new noise-damping materials. This allows for obtaining precise and repeatable results that are independent of variable external conditions. This approach enables the accurate determination of how effectively materials absorb sounds in different frequency bands. Consequently, researchers can better understand which materials have the greatest potential for practical application in terms of acoustic efficiency. Let us also note a certain convergence of conclusions from the studies conducted in the reverberation chamber and the previously presented results of the bench experiment (Kraśkiewicz *et al.*, 2024).

Further research could include a long-term assessment of the durability of the tested materials, particularly regarding their resistance to atmospheric factors and mechanical wear. Undoubtedly, it is necessary to analyse the implementation costs of individual materials in actual infrastructure projects and assess their impact on the total maintenance costs of tracks. Such data would be invaluable when making decisions by designers and managers of railway infrastructure.

As an additional step in future research, numerical methods (such as FEM) could be employed to complement the experimental findings. However, these analyses exceed the scope of the current study.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Authors' contribution

Grzegorz Klekot and Cezary Kraśkiewicz conceptualized the study; Mariusz Wądołowski wrote the original draft; Mariusz Wądołowski and Artur Zbiciak performed the analysis. All authors contributed to the interpretation of the data and reviewed and approved the final version of the manuscript.

References

1. Andrés V.T., Martínez-Casas J., Carballeira J., Denia F.D., Thompson D.J. (2021), Railway rolling noise mitigation through optimal track design, [in:] *Proceedings of the YIC 2021 – VI ECCOMAS Young Investigators Conference*, https://doi.org/10.4995/YIC2021.2021.12583.

2. Basner M. *et al.* (2014), Auditory and non-auditory effects of noise on health, *The Lancet*, **383**(9925): 1325–1332, https://doi.org/10.1016/S0140-6736(13)61613-X.

3. Bunn F., Zannin P.H.T. (2016), Assessment of railway noise in an urban setting, *Applied Acoustics*, **104**: 16–23, https://doi.org/10.1016/j.apacoust.2015.10.025.

4. Cao L., Fu Q., Si Y., Ding B., Yu J. (2018), Porous materials for sound absorption, *Composites Communications*, **10**: 25–35, https://doi.org/10.1016/j.coco.2018.05.001.

5. Castiñeira-Ibáñez S., Rubio C., Romero-García V., Sánchez-Pérez J.V., García-Raffi L.M. (2012), Design, manufacture and characterization of an acoustic barrier made of multi-phenomena cylindrical scatterers arranged in a fractal-based geometry, *Archives of Acoustics*, **37**(4): 455–462, https://doi.org/10.2478/v10168-012-0057-9.

6. Fediuk R., Amran M., Vatin N., Vasilev Y., Lesovik V., Ozbakkaloglu T. (2021), Acoustic properties of innovative concretes: A review, *Materials*, **14**(2): 398, https://doi.org/10.3390/ma14020398.

7. Glickman G.M., Bensing S.J., Carman R. (2011), An examination of trackbed sound absorptive panels for minimizing wayside noise from rail transit, [in:] *NoiseCon Conference*.

8. Kraśkiewicz C. *et al.* (2024), Field experiment as a tool to verify the effectiveness of prototype track structure components aimed at reducing railway noise nuisance, *Archives of Acoustics*, **49**(1): 61–71, https://doi.org/10.24425/aoa.2024.148770.

9. Lee H.-J., Oh S.-T., Lee D.-J. (2016), An optimal mix design of sound absorbing block on concrete ballast in urban train tunnel, *Journal of Korean Tunnelling and Underground Space Association*, **18**(1): 75–82, https://doi.org/10.9711/ktaj.2016.18.1.075.

10. Leštinský L., Zvolenský P. (2019), New methods of noise reduction in railway carriages, *Transportation Research Procedia*, **40**: 778–783, https://doi.org/10.1016/j.trpro.2019.07.110.

11. Licitra G., Fredianelli L., Petri D., Vigotti M.A. (2016), Annoyance evaluation due to overall railway noise and vibration in Pisa urban areas, *Science of the Total Environment*, **568**: 1315–1325, https://doi.org/10.1016/j.scitotenv.2015.11.071.

12. Münzel T. *et al.* (2017), Effects of noise on vascular function, oxidative stress, and inflammation: Mechanistic insight from studies in mice, *European Heart Journal*, **38**(37): 2838–2849, https://doi.org/10.1093/eurheartj/ehx081.

13. NĚMEC M., DANIHELOVÁ A., GERGEL' T., GEJDOŠ M., ONDREJKA V., DANIHELOVÁ Z. (2020), Measurement and prediction of railway noise case study from Slovakia, *International Journal of Environmental Research and Public Health*, **17**(10): 3616, https://doi.org/10.3390/ijerph17103616.

14. SHAHIDAN S., RAMZI HANNAN N.I.R., MD MAAROF M.Z., LEMAN A.S., SENIN M.S. (2017), A comprehensive review on the effectiveness of existing noise barriers commonly used in the railway industry, [in:] *MATEC Web of Conferences*, **87**: 01007, https://doi.org/10.1051/matecconf/20178701007.

15. SUN W., LIU L., YUAN H., SU Q. (2019), Influence of top shape on noise reduction effect of high-speed railway noise barrier, [in:] *IOP Conference Series: Materials Science and Engineering*, **493**(1): 012043, https://doi.org/10.1088/1757-899X/493/1/012043.

16. SUN Y., YANG L., ZHENG J. (2020), Emission control areas: More or fewer?, *Transportation Research Part D: Transport and Environment*, **84**: 102349, https://doi.org/10.1016/j.trd.2020.102349.

17. SZCZEPAŃSKI G., PODLEŚNA M., MORZYŃSKI L., WŁUDARCZYK A. (2023), Investigation of the acoustic properties of a metamaterial with a multi-ring structure, *Archives of Acoustics*, **48**(4): 497–507, https://doi.org/10.24425/aoa.2023.146814.

18. VOGIATZIS K., VANHONACKER P. (2016), Noise reduction in urban LRT networks by combining track based solutions, *Science of the Total Environment*, **568**: 1344–1354, https://doi.org/10.1016/j.scitotenv.2015.05.060.

19. ZHANG X., THOMPSON D., QUARANTA E., SQUICCIARINI G. (2019), An engineering model for the prediction of the sound radiation from a railway track, *Journal of Sound and Vibration*, **461**: 114921, https://doi.org/10.1016/j.jsv.2019.114921.

20. ZHAO C., WANG P., WANG L., LIU D. (2014), Reducing railway noise with porous sound-absorbing concrete slabs, *Advances in Materials Science and Engineering*, **2014**: 206549, https://doi.org/10.1155/2014/206549.

21. ZVOLENSKÝ P., GRENČÍK J., PULTZNEROVÁ A., KAŠIAR L. (2017), Research of noise emission sources in railway transport and effective ways of their reduction, [in:] *MATEC Web of Conferences*, **107**: 00073, https://doi.org/10.1051/matecconf/201710700073.

Research Paper

Comparison of Plate Vibration and Structural Sound Reduction Using Square-Based Sensor–Actuator Piezoelectric Hybrids with Different Shapes of Sensor Part

Roman TROJANOWSKI[iD]

AGH University of Krakow
Kraków, Poland

e-mail: roman.cz.trojanowski@agh.edu.pl

This work is a continuation of the author's previous research on modeling a piezoelectric sensor–actuator hybrid. It presents the results of vibration and structural sound reduction for a plate with attached piezoelectric elements. The models consist of a steel plate with two piezoelectric actuators attached on one side and a hemispherical air volume on the other side. One of the actuators is used to excite the plate's vibration and has the same shape and size in all models. The second actuator is used for vibration and structural sound reduction and varies between a standard square-based full actuator and a sensor–actuator hybrid with different sizes and shapes of the sensor component (either square- or disc-based). Harmonic analyses were performed for the first four mode shapes (skipping the third mode since it is a square plate). Optimization was performed using internal ANSYS functions, with the objective of minimizing the sum of displacement vectors at a number of nodes corresponding to either the full plate or the sensor placed on the said plate.

Keywords: AVC; FEM; plate vibration; sensors.

1. Introduction

During the late 19th century Pierre and Jacques Curie conducted research on various crystals and their properties. They presented their findings in (CURIE, CURIE, 1880), which was the first research paper dealing with the material property now known as piezoelectricity. Today, piezoelectric materials are widely used due to their advantageous properties, such as low power consumption, fast response time and high sensitivity, light weight, flexible sizing, and a wide frequency response range. These materials can be found in microphones, speakers, watches, printers, engines and, of course, in active vibration and sound systems, and more recently, even in energy harvesting applications (MAZUR *et al.*, 2023; WRONA *et al.*, 2022; ISAAC *et al.*, 2022; CHOI *et al.*, 2019). Typically, piezoelectric elements are embedded in or bonded to the surface of a structure to actively control the system's vibration and/or the resulting structural sound (WICIAK,

2008; TROJANOWSKI, WICIAK, 2012). Depending on the piezoelectric effect used, these elements work as actuators or sensors. Piezoelectric materials used for vibration control may consist of single-layer piezoelectric ceramics, multilayer piezoelectric ceramics or composite materials. The foundational work of FULLER *et al.* (1996) and HANSEN and SNYDER (1996) provided much of the necessary theoretical basis for reducing the structural vibrations of simple structures, such as beams and plates using lead zirconate titanate (PZT) elements. Reviews of the modeling of actively controlled piezoelectric smart structures are presented by GUPTA *et al.* (2011) and LI *et al.* (2023).

An important issue regarding the analytical approach to the optimal placement of symmetrical and asymmetrical PZT structures is discussed by BARBONI *et al.* (2000), BRAŃSKI and LIPIŃSKI (2011), BRAŃSKI (2013), and BRAŃSKI and KURAS (2022). The best results were obtained when the actuator was placed at the point of maximum beam bending moment under

the given boundary conditions. There are also studies that numerically investigated the optimal location of PZT actuators (Augustyn *et al.*, 2014; Fawade, Fawade, 2016; Kozień, Ścisło, 2015). In works by Dimitriadis *et al.* (1991), Sekouri *et al.* (2004), Sullivan *et al.* (1996), and Wiciak, Trojanowski (2014b) authors demonstrated the advantages of using distributed actuators of various shapes for active control. Some of the author's previous work focused on materials with step-changing properties (Wiciak, Trojanowski, 2014a). These studies primarily analyzed the efficiency of piezoelectric actuators with different shapes (circular, rectangular, triangular), as well as the effect of step changes in the actuators' material parameters. These analyses showed that even small reductions in vibration levels at particular resonant frequencies could lead to significant changes in the maximum levels of sound pressure radiated by the plate. These reductions in vibration levels were achieved by adjusting the material properties of a particular actuator regions and appropriately increasing the voltage applied to them (Wiciak, Trojanowski, 2015; Trojanowski, Wiciak, 2022). From this research, a new concept emerged: replacing part of the piezoelectric actuator area could create a sensor–actuator hybrid capable of comparable vibration reduction. Numerical models presented in (Trojanowski, Wiciak, 2022) show that such a hybrid is able to achieve similar vibration reduction levels to those of standard (full) actuator, although a higher voltage input is required.

This article is a continuation of the work presented in (Trojanowski, Wiciak, 2022; Trojanowski, 2024), and it presents numerical analyses of a piezoelectric sensor–actuator hybrid composed of a square-based outer section and disc-based inner section. The study includes results of both vibration reduction and sound pressure level reduction.

2. Materials and methods

The object of interest are the vibrations of plates and the resulting sound pressure radiated into the surroundings. For this purpose, a model of a steel plate clamped on all sides, with piezoelectric elements attached to one surface and a hemispherical air domain on the other is considered. Since we continue the previous works (Trojanowski, Wiciak, 2022; Trojanowski, 2024), the dimensions of the modeled objects as well as the material parameters used in models remain the same. Building upon the second iteration of the piezoelectric sensor–actuator hybrid (Figs. 1a, 1b) introduced in (Trojanowski, Wiciak, 2022), this study introduces a modification to the shape of the sensor component (Figs. 1c, 1d).

Changing the shape of the inner part of the sensor–actuator should allow for easier production of physical
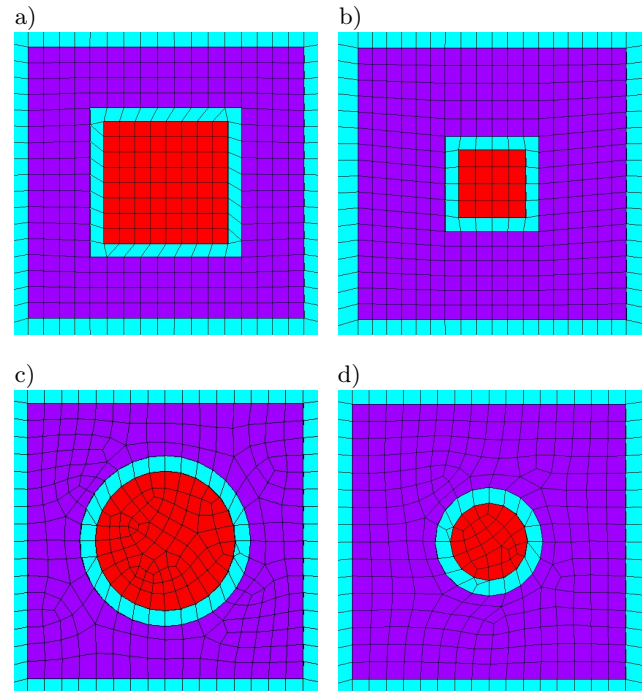
Fig. 1. Concept of a sensor–actuator: a) and b) – square-based sensor–actuators; c) and d) – disc-based sensor–actuators; red – sensor part, violet – actuator part, blue – modeled plate.

prototypes, as eliminating sharp edges makes it easier to cut out parts of piezoelectric elements.

The base of the model is a square steel plate with a side length of 400 mm and a thickness of 2 mm (Fig. 2a). A hemispherical air domain with a radius of 1 m is attached to the plate to simulate the acoustic field near the plate (Fig. 2b). The outer boundary of
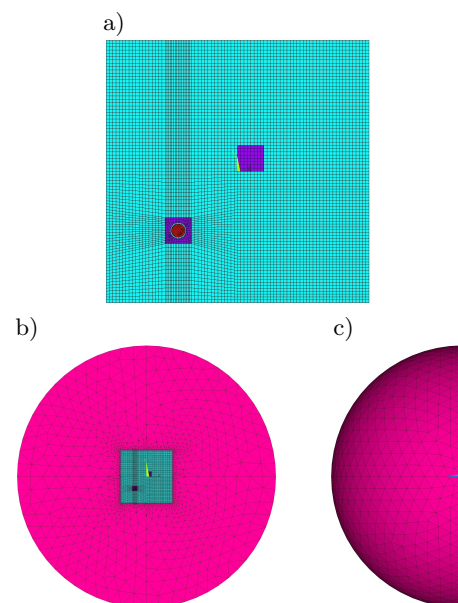
Fig. 2. a) Modeled plate; b) and c) modeled plate with surrounding air.

the air domain is treated as an open space, meaning no reflections occur at the boundary. The vibrations of the plate are excited by a standard square-based piezoelectric actuator placed near the center of the plate. A second piezoelectric element is placed along the diagonal of the plate, approximately $1/4$ of its length, and is used for vibration reduction. Depending on the case, this second element may be a standard square-based piezo actuator or a square-based sensor–actuator hybrid with varying shapes and sizes of the sensor region.

The total base area of the modeled piezoelectric elements is $1600\,\text{mm}^2$, which corresponds to a side length of $20\,\text{mm}$ for the square actuator. All piezoelectric elements have a uniform thickness of $1\,\text{mm}$. In the hybrid configuration, this base area includes the gap between the sensor and actuator parts. The inner sensor part of the hybrid can be either square or disc-shaped and is defined in two sizes:

– for the hybrid with a larger sensor region, the sensor part area ranged from 0 to 0.45 of the side length of the entire element, while the actuator area was from 0.55 to the full side length of the whole element (for a disc-based sensor, this corresponds to 0 to 0.45 of the radius of a disc element with the same base area);

– for the hybrid with a smaller sensor part, the sensor part area ranged from 0 to 0.25 of the radius of the entire element, while the actuator area extended from 0.35 to the full radius of the whole element (for a disc-based sensor, this corresponds to 0 to 0.25 of the radius of a disc element with the same base area).

Table 1 shows the elements and properties used in the model.

Table 1. Parameters used in the model; E – Young's modulus [Pa]; ν – Poisson's ratio; ρ – density [kg/m^3]; c – speed of sound in air [m/s].

Structural element	Element in ANSYS library	Properties
Plate	SOLSH190	$E = 1.93 \cdot 10^{11}$ $\nu = 0.29$ $\rho = 7800$
Piezo elements	SOLID226	Properties of PZ28
Air	FLUID30	$\rho = 1.2$ $c = 343$

Harmonic analyses were performed for the 1st, 2nd, 4th, and 5th modes (Fig. 3). It can be seen that the actuator placed near the center of the plate performs very effectively for the 1st and 5th modes. Although the placement is not optimal for the 2nd and especially the 4th mode, it still allows to evaluate the sensor–actuator hybrid under different conditions. The placement of the second actuator/sensor–actuator hybrid was chosen to achieve fairly optimal results across all analyzed modes.
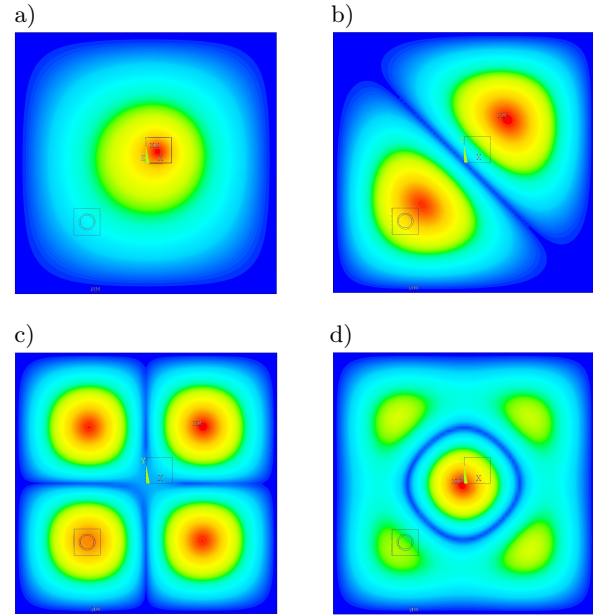
Fig. 3. Analyzed mode shapes: a) 1st (108.6 Hz); b) 2nd (222.1 Hz); c) 4th (330.0 Hz); d) 5th (401.9 Hz).

Figure 4 shows the distribution of sound pressure levels corresponding to the plate's 1st and 4th mode shapes. The distributions of sound pressure levels are visualized by cutting the hemispherical air domain in half through its center, perpendicular to the plate (in this case, it is cut so we can see the ZY-plane) and then rotating it by 30°. This approach allows visualization of the sound pressure levels above and to the side of the plate.
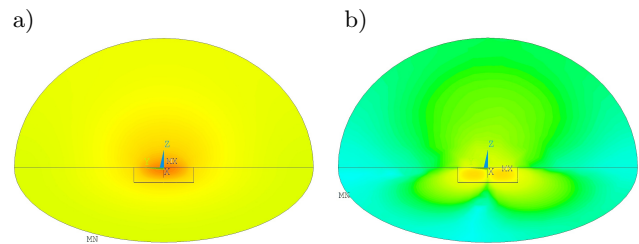
Fig. 4. Examples of sound pressure level distribution before vibration reduction: a) 1st mode; b) 4th mode; reference acoustic pressure is $20\,\mu\text{Pa}$.

The voltage applied to the actuators used to excite plates vibrations was $100\,\text{V}$ with a phase angle 0° in each case. To find the voltages for vibration reduction, an optimization procedure using ANSYS internal functions was performed. The number of steps for a single run of the optimization procedure was capped at 30 steps. For each run, a voltage range that could be applied to the disc-based element was specified, starting with $0\,\text{V}$–$600\,\text{V}$ for the first run. Upon completing the first run of the procedure, another run was initiated using the previous run's value as the starting point within a narrower voltage range. This process

was repeated until the voltage amplitude range was ±2.5 V around the staring value.

The phase angle of the voltage applied to the actuator was not treated as a variable for the analyses, as it was shown in previous studies that it should be either 0° or 180°, depending on the mode.

If any of the obtained vibration reduction values differed significantly from the other results, a manual check and tuning were performed.

The goal function used in the optimization procedure is given by

$$J_1 = \min \sum_{i=1}^{n} |\mathbf{x}_{\text{sum}}(i)|, \tag{1}$$

where min is the minimum value of the sum, $\mathbf{x}_{\text{sum}}(i)$ is the displacement vector sum at the i-th node, n is the total number of nodes considered in the calculation.

Three cases for sensor placement were analyzed, with the possible values of n defined as follows:

1) n corresponds to all nodes on the back of the plate (here the back of the plate is the side to which the piezoelectric elements are not attached). The actual number depends on the shape of the hybrid element as well as the ratio between its inner and outer parts (as these parameters influence the finite elements mesh). Typically, this number exceeds 7500 nodes. This is considered the best-case scenario;

2) n is equal to 25 nodes forming a 'virtual' sensor that covers a quarter of the size of the square-based actuator placed along the same diagonal as the piezoelectric actuators, but on the upper side of the plate (approximately 1/4 of the plate's length);

3) n is equal to at least 65 nodes forming a 'virtual' sensor corresponding in size to the sensor part of the sensor–actuator placed directly beneath it (the exact number changes depending on the sensor part's size).

3. Results

This section presents the results of the numerical analyses performed. Equation (2) defines how the vibration reduction levels were calculated:

$$L_{\text{red}} = 20\log_{10} \frac{\sum_{i=1}^{n} |\mathbf{x}_{1\text{sum}}(i)|}{\sum_{i=1}^{n} |\mathbf{x}_{2\text{sum}}(i)|}, \tag{2}$$

where $\mathbf{x}_{1\text{sum}}(i)$ is the displacement vector sum at the i-th node before the reduction, $\mathbf{x}_{2\text{sum}}(i)$ is the displacement vector sum at the i-th node after reduction, n is the number of nodes used (corresponding to the three cases mentioned earlier).

Equation (3) shows how the sound pressure level reduction was calculated. For calculations, a slice of the hemispherical air domain with a radius range from 0.4 m to 0.6 m was selected:

$$\text{SPL}_{\text{red}} = 20\log_{10} \frac{\sum_{j=1}^{m} |p_1(j)|}{\sum_{j=1}^{m} |p_2(j)|}, \tag{3}$$

where $p_{1(j)}$ is the sound pressure at the j-th node before reduction, $p_{2(j)}$ is the sound pressure at the j-th node after reduction, m is the number of nodes enclosed within the slice of the hemispherical air domain chosen (depending on the sensor part's shape and size, at least 992 nodes).

Equation (4) defines how the maximum sound pressure level reduction was calculated:

$$\text{SPL}_{\text{mred}} = \text{SPL}_{\text{max}\,1} - \text{SPL}_{\text{max}\,2}, \tag{4}$$

where $\text{SPL}_{\text{max}\,1}$ is the maximum sound pressure level at any node in the hemispherical air domain before reduction, $\text{SPL}_{\text{max}\,2}$ is the maximum sound pressure level in any node of the hemispherical air domain after reduction.

Equation 5 introduces the parameter called voltage efficiency calculated as:

$$UL = \frac{U}{L_{\text{red}}}, \tag{5}$$

where U is the amplitude of voltage applied to the actuator; L_{red} is the level of vibration reduction.

This parameter was introduced as a simplified method to compare the efficiency of different actuator shapes. Since all elements have the same area and are modeled using the same material parameters, this should help to compare the efficiency of piezoelectric elements used in terms of the voltage applied relative to the vibration reduction levels achieved.

Results presented in Tables 2, 3, and 4 showing vibration reduction (columns L_{red} and L_{ref}) as well as voltage efficiency (column UL) for the full actuator, sensor–actuator with larger square sensor and sensor–actuator with smaller square were previously reported in (Trojanowski, Wiciak, 2022). They are presented for comparison reasons and are highlighted with a grey background for clarity.

Table 2 presents the results for vibrations and sound pressure reduction levels achieved when using square-based actuators and sensor–actuators as well as using the entire back area of the plate as a sensor.

The obtained vibration reduction levels for the square-based sensor–actuator with a disc-shaped sensor part range from 26 dB to more than 43 dB. It can be seen that for the 1st mode, all sensor–actuator results are about 0.2 dB–0.3 dB lower than those of the full actuator. Similarly to the findings for the square-based

Table 2. Results obtained for square-based actuators using the entire back area of the plate as a sensor; U_a – amplitude of voltage applied to the actuator; ϕ_a – phase of the voltage applied to the actuator; L_{red} – vibration reduction; UL – voltage efficiency; $\mathrm{SPL}_{\mathrm{mred}}$ – maximum sound pressure levels reduction; $\mathrm{SPL}_{\mathrm{red}}$ – overall sound pressure level reduction.

Mode number	Type	U_a [V]	ϕ_a [°]	L_{red} [dB]	UL [V/dB]	$\mathrm{SPL}_{\mathrm{mred}}$ [dB]	$\mathrm{SPL}_{\mathrm{red}}$ [dB]
1	Actuator (full)	365.29	180.00	41.3	8.8	41.3	58.1
2		57.77	360.00	43.1	1.3	42.5	40.7
4		12.05	180.00	25.8	0.5	21.5	9.9
5		159.95	360.00	35.0	4.6	33.8	46.1
1	Sensor–actuator (larger square)	558.69	180.00	41.0	13.6	41.4	58.3
2		88.40	360.00	43.1	2.1	42.6	40.4
4		18.38	180.00	26.1	0.7	21.4	9.5
5		247.37	360.00	35.4	7.0	34.2	39.4
1	Sensor–actuator (smaller square)	424.45	180.00	41.0	10.4	41.4	58.7
2		67.14	360.00	43.1	1.6	42.5	40.2
4		14.04	180.00	26.2	0.5	21.4	9.6
5		186.95	360.00	35.4	5.3	34.0	44.4
1	Sensor–actuator (larger disc)	555.11	180.00	41.0	13.6	41.4	58.6
2		87.63	360.00	43.2	2.0	42.6	40.3
4		18.20	180.00	26.0	0.7	21.4	9.9
5		245.09	360.00	35.4	6.9	34.1	44.5
1	Sensor–actuator (smaller disc)	424.81	180.00	41.1	10.3	41.3	58.7
2		66.98	360.00	43.1	1.6	42.6	40.3
4		13.98	180.00	26.0	0.5	21.5	9.9
5		186.29	360.00	35.4	5.3	34.1	44.6

Table 3. Results obtained for square-based actuators when using a 'virtual' sensor on the diagonal in the upper side of the plate; L_{redf} – vibration reduction calculated for all nodes on the back side of the plate.

Mode number	Type	U_a [V]	ϕ_a [°]	L_{red} [dB]	L_{redf} [dB]	UL [V/dB]	$\mathrm{SPL}_{\mathrm{mred}}$ [dB]	$\mathrm{SPL}_{\mathrm{red}}$ [dB]
1	Actuator (full)	365.42	180.00	44.6	41.3	8.2	41.3	57.8
2		57.80	360.00	53.6	43.0	1.1	42.6	40.7
4		12.20	180.00	23.9	25.9	0.5	21.5	9.9
5		158.66	360.00	31.6	33.7	5.0	33.9	42.8
1	Sensor–actuator (larger square)	558.81	180.00	44.6	41.0	12.5	41.4	58.2
2		88.40	360.00	53.8	43.1	1.6	42.6	40.4
4		18.50	180.00	23.8	25.9	0.8	21.3	9.5
5		247.29	360.00	31.4	35.4	7.9	34.2	44.5
1	Sensor–actuator (smaller square)	424.48	180.00	44.6	41.0	9.5	41.4	58.7
2		67.14	360.00	53.9	43.1	1.2	42.5	40.2
4		14.04	180.00	23.8	26.2	0.6	21.4	9.6
5		186.81	360.00	31.5	35.4	5.9	34.1	44.4
1	Sensor–actuator (larger disc)	555.05	180.00	44.6	41.0	12.5	41.4	58.6
2		87.62	360.00	54.0	43.2	1.6	42.6	40.3
4		18.24	180.00	23.8	26.0	0.8	21.4	10.0
5		245.03	360.00	31.4	35.4	7.8	34.1	44.5
1	Sensor–actuator (smaller disc)	424.82	180.00	44.6	41.1	9.5	41.3	58.7
2		66.98	360.00	54.0	43.1	1.2	42.6	40.3
4		14.04	180.00	23.8	26.0	0.6	21.5	9.9
5		186.46	360.00	31.5	35.4	5.9	33.9	44.5

sensor part (TROJANOWSKI, WICIAK, 2022), there seems to be virtually no difference for the 2nd mode. For the 4th mode, all the sensor–actuators achieve vibration reduction levels around 0.2 dB–0.3 dB higher than the standard actuators, and this difference rises to about 0.4 dB for the last analyzed mode. None of these differences can be called significant, and there is almost no variation in vibration reduction level results when comparing hybrids with different sensor shapes and sizes.

As for the voltages applied to the sensor–actuator, since changes in the sensor part shape do not significantly change the required voltages, the findings presented in (TROJANOWSKI, WICIAK, 2022) are still valid. The full actuator requires less voltage to achieve the results shown, but reducing the sensor part size can somewhat mitigate this difference.

Table 3 presents the results of vibration and acoustic pressure levels reduction for square-based actuators and sensor–actuators in the scenario where the sensor

Table 4. Results obtained for square-based actuators using a 'virtual' sensor placed beneath the center of the square-based actuator.

Mode number	Type	U_a [V]	ϕ_a [°]	L_{red} [dB]	L_{redf} [dB]	UL [V/dB]	SPL_{mred} [dB]	SPL_{red} [dB]
1	Actuator (full)	363.40	180.00	25.4	39.2	14.3	40.7	45.1
2		57.76	360.00	41.9	43.1	1.4	42.5	40.7
4		12.27	180.00	31.4	25.7	0.4	21.4	10.0
5		161.03	360.00	33.4	34.4	4.8	32.1	40.3
1	Sensor–actuator (larger square)	557.24	180.00	26.0	40.2	21.4	41.2	50.5
2		88.45	360.00	42.8	43.1	2.1	42.6	40.4
4		18.36	180.00	32.2	26.1	0.6	21.4	9.5
5		247.48	360.00	35.1	35.4	7.1	34.2	44.5
1	Sensor–actuator (smaller square)	423.19	180.00	25.5	40.0	16.6	41.1	49.6
2		67.14	360.00	42.2	43.1	1.6	42.5	40.2
4		14.05	180.00	32.0	26.1	0.4	21.4	9.6
5		186.85	360.00	34.0	35.4	5.5	34.1	44.4
1	Sensor–actuator (larger disc)	553.37	180.00	26.0	40.0	21.3	41.1	49.5
2		87.64	360.00	42.8	43.2	2.0	42.6	40.3
4		18.21	180.00	32.3	26.0	0.6	21.4	9.9
5		245.01	360.00	35.0	35.4	7.0	34.1	44.5
1	Sensor–actuator (smaller disc)	423.48	180.00	25.5	40.1	16.6	41.0	49.0
2		66.98	360.00	42.1	43.1	1.6	42.6	40.3
4		13.98	180.00	32.0	26.0	0.4	21.5	9.9
5		186.08	360.00	33.9	35.4	5.5	34.2	44.5

is placed on the same diagonal as the actuator used for reduction, but located in the upper right quarter of the plate.

In this scenario, vibration reduction levels obtained range from nearly 24 dB (23.8 dB) for the 4th mode to 54 dB for the 2nd mode, regardless of the shape of the sensor part of the sensor–actuator hybrid. The largest difference between the full actuator and a sensor–actuator hybrid with a disc-shaped sensor part occurs in the 2nd mode and is about 0.4 dB, regardless the size of the sensor part of the hybrid. Again, there is almost no difference in performance among sensor–actuators, regardless of the size or shape of the sensor part.

As for the voltage efficiency, the same trend observed in (TROJANOWSKI, WICIAK, 2022) and in Table 2 is clearly apparent here. Using sensor–actuator hybrid results in higher values of UL. While changing the shape of the sensor part in the sensor–actuator hybrid does change the voltages that have to be applied to achieve vibration reduction, these changes are minor. Full actuators require less voltage to achieve similar vibration reduction levels due to their larger active area.

Table 4 presents the results for vibration and acoustic pressure levels reduction when using square-based actuators in a scenario where the sensor is placed beneath the center of the actuator (or sensor–actuator) and matches the size of the sensor part of the sensor–actuator.

In this scenario, vibration reduction levels for the disc-based hybrids range from around 25.5 dB for the 1st mode to almost 43 dB for the 2nd mode. For

this sensor placement, the sensor–actuator hybrids achieve slightly higher vibration reduction levels in every case (up to 1.6 dB). Again, changing the shape of the sensor part does not significantly influence the results.

As for the UL values, we can observe a similar trend to that observed in data in Tables 2 and 3. While the values differ depending on the vibration reduction levels achieved, the ratios between different actuator types are similar to those observed in previous cases.

Similarly as in (TROJANOWSKI, WICIAK, 2022; TROJANOWSKI, 2024), the next step in presenting the results is recalculating the vibration reduction levels for the best-case scenario – using the back of the plate as the sensor (L_{redf} column in Tables 3 and 4, and L_{red} column in Table 2). These results are shown in Fig. 5.

The first five bars for each mode represent the results from the first case (Table 2, discussed ealier).

The middle five bars for each mode show the results for the sensor placed on the diagonal, about one-quarter of the plate's length from the upper right side, recalculated to match the results for the full plate sensor scenario. It can be seen that these results are very similar to those of the first case. For the 1st mode, vibration reduction levels achieved using the sensor–actuator hybrids are about 0.2 dB–0.3 dB lower than those of the full actuator. There are almost no differences observed in the results for the 2nd mode. For the 4th mode, standard actuators achieve slightly lower reduction levels (up to 0.4 dB less). The biggest differences can be observed in the vibration reduction levels for the 5th mode. In this mode, sensor–actuator hybrids achieved results higher than those of the stan-
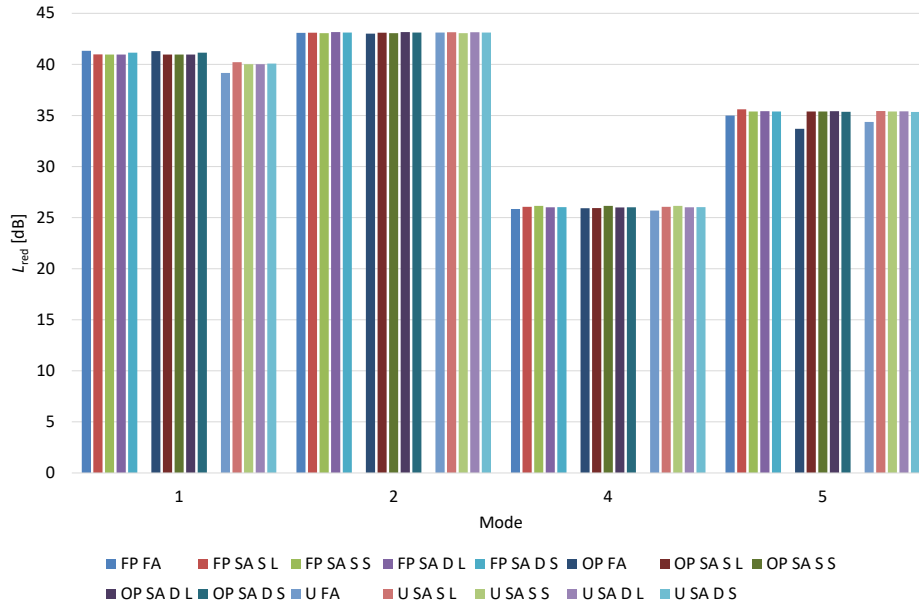
Fig. 5. Comparison of vibration reduction levels obtained for square-based actuators, recalculated for the full plate; FP – scenario using the whole back of the plate as a sensor; OP – sensor placed on the diagonal in the upper right quadrant of the plate; UP – sensor placed beneath the actuator (or sensor–actuator); FA – full actuator; SA S L – sensor–actuator with larger square-based sensor part; SA S S – sensor–actuator with smaller square-based sensor part, SA D L – sensor–actuator with larger disc-based sensor part; SA D S – sensor–actuator with smaller disc-based sensor part.

dard actuator by 1.7 dB. Differences between the recalculated results for the piezoelectric hybrid elements do not exceed 0.3 dB.

The last five bars represent the recalculated results for the scenario where the sensor corresponding in size and placement to the sensor part of the sensor–actuator. For the 1st mode, it can be seen that vibration reduction levels for this case are lower than in the other two scenarios – about 2.1 dB lower for the full actuator and 0.8 dB–1.0 dB lower for the sensor–actuator hybrids. This translates to the full actuator showing approximately 0.8 dB to 1.0 dB lower vibration reduction than the sensor–actuator hybrids in this scenario. For the 2nd and 4th modes, the obtained results are similar to those of the other two cases. For the 5th mode, it can be seen that results obtained for the sensor–actuator hybrids are higher by about 1.0 dB than the results for the full actuator. The differences in vibration reduction levels between the full actuators and sensor–actuator hybrids for the 1st and 5th modes in this scenario can be attributed to the presence of the 'working' part of the actuator occupying the same place as the sensor in the full actuator case. However, since this situation does not occur for the 2nd and 4th mode, and if we take into account the mode shapes shown in Figs. 2a and 3, it can be assumed that the actuator placement relative to the mode shapes impacts these differences.

The differences between different shapes and sizes of the sensor part of the sensor–actuator hybrids do not exceed 0.2 dB.

In (TROJANOWSKI, WICIAK, 2022), it was proposed that reducing the size of the sensor part in a sensor–actuator hybrid does not influence the vibration reduction levels obtained when using said hybrids. This observation seems to hold true even after changing the shape of the sensor part. Based on the results presented in this paper, we can add another assumption that changing the sensor part of a square-based sensor–actuator hybrid from square to disc-shaped also does not have a significant impact on the results of vibration reduction levels obtained. This is a positive finding, as it should be easier to produce physical prototypes of hybrids with a disc-shaped sensor part than with a square-shaped one.

The final step in presenting the results focuses on the acoustic pressure level reductions accompanying vibration reduction of the plates. These results are presented in Tables 2, 3, and 4 as well as in Fig. 6.

The leftmost five bars for each mode represent the sound pressure reduction when using the whole back area of the plate as a sensor. The obtained sound pressure reductions range from about 9.5 dB for the 4th mode up to almost 59 dB for the 1st mode. It can be seen that for the 1st mode, higher values of sound pressure reduction are obtained for the sensor–actuator hybrids than for the full actuator, despite the full actuator achieving about 0.3 dB greater vibration reduction. Although these differences are not significant (at most, around 0.6 dB), they occur across all hybrid configurations. For the 2nd and 4th modes, the differences in sound pressure level reduction among the modeled
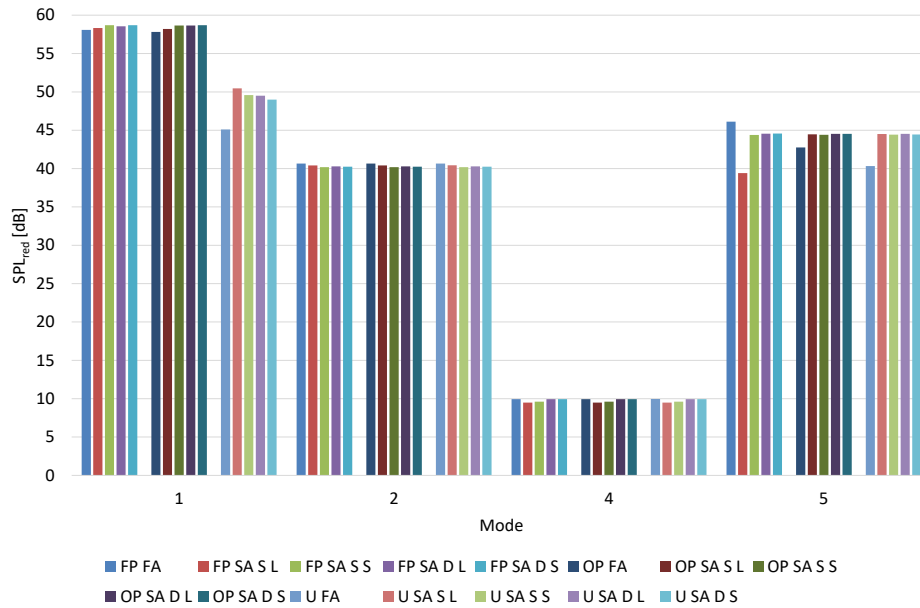
Fig. 6. Comparison of acoustic pressure levels reductions obtained for square-based actuators.

actuators are no greater than 0.5 dB. However, for the 5th mode, there are two interesting observations. First, the full actuator achieves the highest sound pressure level reduction, around 1.6 dB greater than most of the sensor–actuator hybrids. Second, the hybrid with the larger square-based sensor part achieves significantly lower sound pressure level reduction (around 5 dB) than the other hybrids, and around 6.7 dB less than the full actuator. This occurs at the same time when the vibration reduction levels corresponding to that sound pressure level reduction are the highest among all modeled actuators/sensor–actuator hybrids. This, coupled with the vibration reduction level data (Fig. 5), sound pressure level reduction results for other sensor placements and sizes (Fig. 6), the maximum sound pressure level reduction and the small differences in voltages applied to actuators to achieve these results (Tables 2–4), may indicate that this anomaly is the results of a specific vibration distribution pattern of the plate.

The middle five bars for each mode represent the sound pressure reduction accompanying the plate's vibration reduction when using a sensor placed on the diagonal opposite the actuator placement. It can be seen that for the 1st, 2nd, and 4th modes, these results are very similar to the results obtained from the previous scenario. The biggest difference can be seen in the 5th mode. In this scenario, all sensor–actuator hybrids achieve very similar values of sound pressure level reduction (with no anomaly as observed in the previous case) and these reductions are about 1.7 dB higher than those achieved using a full actuator.

The five bars on the right, for each mode, represent the sound pressure reduction accompanying the plate's vibration reduction when using a sensor corre-

sponding in size and placement to the sensor part of the sensor–actuator. Here, the sound pressure level reduction ranged from 9.5 dB for the 4th mode to 50.5 dB for the 1st mode. For the 1st mode, the small decrease in overall vibration level reduction obtained compared to previous scenarios (about 2.1 dB for the full actuator and 0.8 dB–1.0 dB for the sensor–actuators) results in a significant reduction in corresponding sound pressure level reduction (around 13 dB for the full actuator and around 8 dB–10 dB for the sensor actuator hybrids). For the 2nd and 4th modes, again, there are no significant changes in results. In the 5th mode, the sound pressure level reduction achieved using the full actuator was more than 4 dB lower than the one obtained using sensor–actuator hybrids.

It can be seen that changing the shape or size of the sensor part of the sensor–actuator hybrid does not have a significant impact on the sound pressure level reduction results.

Figure 7 shows examples of sound pressure level distributions before (left column) and after (right column) the plate's vibration reduction. The presented figure is for the first scenario (using full back area of the plate as a sensor) and compares the full actuator with a square-based sensor–actuator and a larger disc-based sensor part. It can be observed that the change in SPL_{red} values are not solely due to a decrease in sound pressure levels radiating from the plate, but also due to a change in the type of sound source that the plate represents. For example, in the 1st mode, before vibration reduction, the plate acts like a monopole sound source, while after reduction it behaves like a dipole. Similar changes can be seen in the 4th and 5th modes, where the sound pressure distribution also shift following the plate's vibration reduction. In a way, the pattern flips –
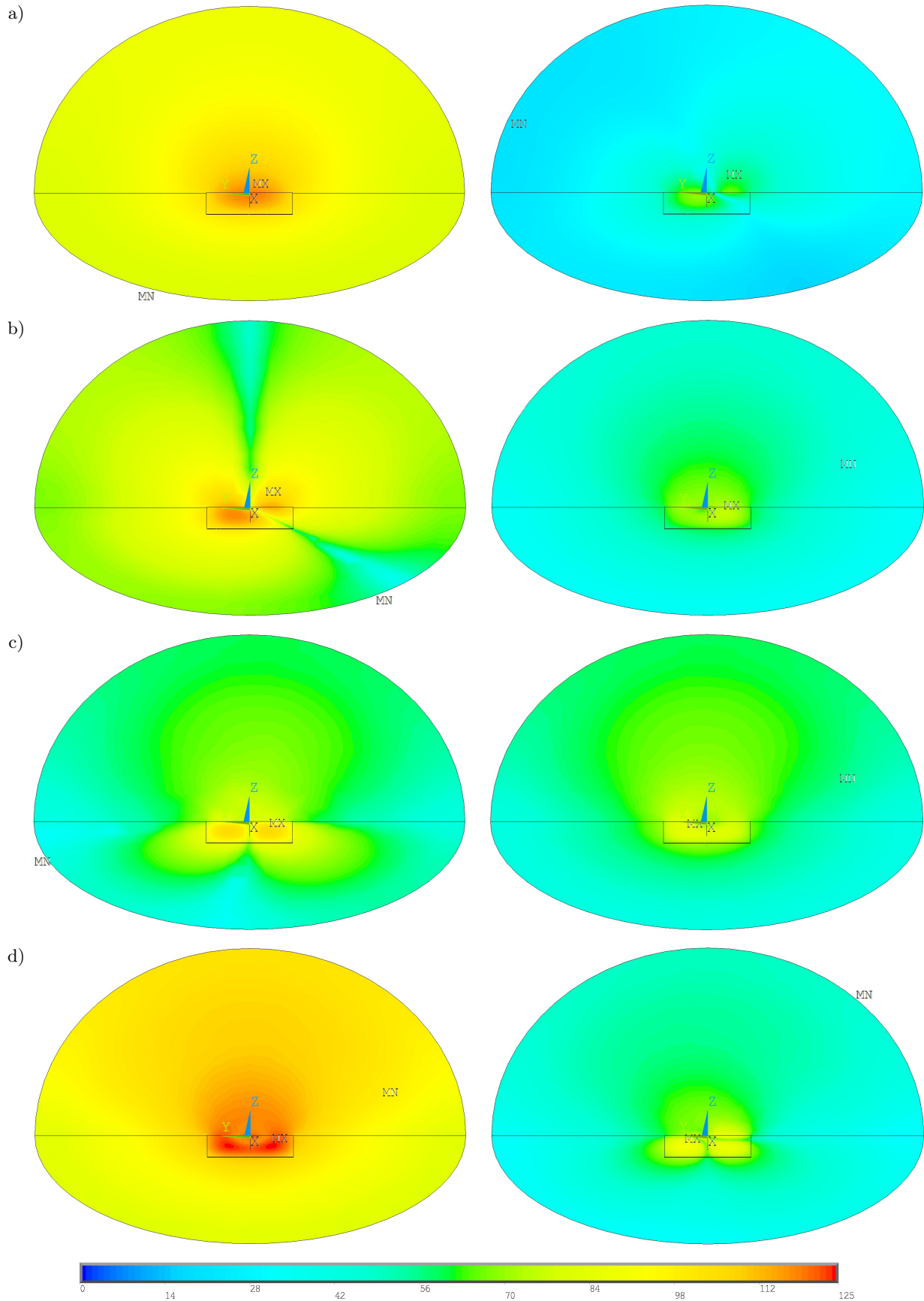
a)

b)

c)

d)

Fig. 7. Examples of sound pressure level distributions before (left column) and after (right column) plate vibration reduction for different modes: a) 1st mode; b) 2nd mode; c) 4th mode; d) 5th mode; reference acoustic pressure: 20 μPa.

for the 4th mode, the plate's vibration pattern initially resembles a quadrupole and transitions to a 'ring-like' pattern (with a sound pressure distribution similar to that of a monopole) after reduction. Conversely, in the 5th mode, this situation is reversed – it begins with a 'ring-like' pattern and transitions to a quadrupole. This change is partially responsible for the difference observed between SPL_{mred} (reduction of the maximum sound pressure levels) and SPL_{red} (reduction of sound pressure levels in a portion of the hemispherical volume) values, as shown in Tables 2–4. For example, in the 4th mode, SPL_{mred} exceeds 21 dB for every type of actuator, while SPL_{red} remains below 10 dB.

4. Conclusions

For the analyzed modes, changing the shape of the sensor part in a square-based sensor–actuator hybrid does not seem to significantly change its performance. The proposed square-based hybrid with a disc-shaped sensor part appears to be capable of achieving vibration reduction levels comparable to those of a standard actuator. In some cases, the results obtained for the hybrid were slightly higher than those for the full actuator, although the differences were not really significant. The sensor part appears to function effectively, and the results are fairly comparable in every scenario.

Similarly, square-based sensor–actuator hybrids demonstrate sound pressure level reductions comparable to those achieved by full actuators when taking into account sound vibration in the plate reduction. The downside of using sensor–actuators is the higher voltage required to achieve vibration reduction levels comparable to those of standard actuators. This is expected, as the actuator part of the sensor–actuator is smaller than a full actuator. Also, this can be partially mitigated by using a smaller sensor part.

The next step of this research will involve creating a physical prototype of the sensor–actuator based on the second iteration models, followed by physical experiment to confirm the numerical results. As for numerical modeling, next we will determine the optimal ratios between the sensor and actuator parts.

Conflict of interest

The author declare that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. AUGUSTYN E., KOZIEŃ M.S., PRĄCIK M. (2014), FEM analysis of active reduction of torsional vibrations of clamped-free beam by piezoelectric elements for separated mode, *Archives of Acoustics*, **39**(4): 639–644, https://doi.org/10.2478/aoa-2014-0069.

2. BARBONI R., MANNINI A., FANTINI E., GAUDENZI P. (2000), Optimal placement of PZT actuators for the control of beam dynamics, *Smart Materials and Structures*, **9**: 110–120, https://doi.org/10.1088/0964-1726/9/1/312.

3. BRAŃSKI A. (2013), Effectiveness analysis of the beam modes active vibration protection with different number of actuators, *Acta Physica Polonica A*, **123**(6): 1123–1127, https://doi.org/10.12693/APhysPolA.123.1123.

4. BRAŃSKI A., KURAS R. (2022), Asymmetrical PZT applied to active reduction of asymmetrically vibrating beam – Semi-analytical solution, *Archives of Acoustics*, **47**(4): 555–564, https://doi.org/10.24425/aoa.2022.142891.

5. BRAŃSKI A., LIPIŃSKI G. (2011), Analytical determination of the PZTs distribution in active beam vibration protection problem, *Acta Physica Polonica A*, **119**(6A): 936–941.

6. CHOI J., JUNG I., KANG Ch.-Y. (2019), A brief review of sound energy harvesting, *Nano Energy*, **56**: 169–183, https://doi.org/10.1016/j.nanoen.2018.11.036.

7. CURIE J., CURIE P. (1880), Development, via compression, of electric polarization in hemihedral crystals with inclined faces [in French], *Bulletin de la Societe de Minerologique de France*, **3**: 90–93, https://doi.org/10.3406/bulmi.1880.1564.

8. DIMITRIADIS E.K., FULLER C.R., ROGERS C.A. (1991), Piezoelectric actuators for distributed vibration excitation of thin plates, *Journal of Vibration and Acoustics*, **113**: 100–107, https://doi.org/10.1115/1.2930143.

9. FAWADE A.S., FAWADE S.S. (2016), Modeling and analysis of vibration controlled cantilever beam bounded by PZT patch, *International Journal of Engineering Inventions*, **5**(7): 7–14.

10. FULLER C.R., ELLIOTT S.J., NELSON P.A. (1996), *Active Control of Vibration*, Academic Press, London.

11. GUPTA V., SHARMA M., THAKUR N. (2011), Mathematical modelling of actively controlled piezo smart structures: A review, *Smart Structures and Systems*, **8**(3): 275–302, https://doi.org/10.12989/sss.2011.8.3.275.

12. HANSEN C.H., SNYDER S.D. (1996), *Active Control of Noise and Vibration*, CRC Press, London, https://doi.org/10.1201/9780367804831.

13. ISAAC C.W., WRONA S., PAWEŁCZYK M., KARIMI H.R. (2022), Modelling vibro-acoustic response of lightweight square aluminium panel influenced by sound source locations for active control, *Scientific Reports*, **12**(1): 10727, https://doi.org/10.1038/s41598-022-14951-y.

14. Kozień M.S., Ścisło Ł. (2015), Simulation of control algorithm for active reduction of transversal vibrations of beams by piezoelectric elements based on identification of bending moment, *Acta Physica Polonica A*, **128**(1A): 56–61, https://doi.org/10.12693/APhysPolA.128.A-56.

15. Li J., Zhang L., Li S., Mao Q., Mao Y. (2023), Active disturbance rejection control for piezoelectric smart structures: A review, *Machines*, **11**(2): 174, https://doi.org/10.3390/machines11020174.

16. Mazur K., Pawełczyk M., Karimi H.R. (2023), Global noise reduction in free-field conditions by using active noise-controlling casings, *Mechanical Systems and Signal Processing*, **197**: 110393, https://doi.org/10.1016/j.ymssp.2023.110393.

17. Sekouri E.M., Hu Y.R., Ngo A.D., (2004), Modelling of a circular plate with piezoelectric actuators, *Mechatronics*, **14**(9): 1007–1020, https://doi.org/10.1016/j.mechatronics.2004.04.003.

18. Sullivan J.M., Hubbard J.E. Jr., Burke S.E. (1996), Modeling approach for two-dimensional distributed transducers of arbitrary spatial distribution, *The Journal of the Acoustical Society of America*, **99**(5): 2965–2974, https://doi.org/10.1121/1.414861.

19. Trojanowski R. (2024), Piezoelectric disc based sensor-actuator hybrid in vibration reduction, *Vibrations in Physical Systems*, **35**(1): 2024117-1–2024117-9, https://doi.org/10.21008/j.0860-6897.2024.1.17.

20. Trojanowski R., Wiciak J. (2012), Structural noise reduction and its effects on plate vibrations, *Acta Physica Polonica A*, **121**(1A): A-148–A-151, https://doi.org/10.12693/APhysPolA.121.A-148.

21. Trojanowski R., Wiciak J. (2020), Impact of the size of the sensor part on sensor-actuator efficiency, *Journal of Theoretical and Applied Mechanics*, **58**(2): 391–401, https://doi.org/10.15632/jtam-pl/118948.

22. Trojanowski R., Wiciak J. (2022), Piezoelectric square based sensor-actuator hybrid in vibration reduction, *Vibrations in Physical Systems*, **33**(3): 2022303-1–2022303-9, https://doi.org/10.21008/j.0860-6897.2022.3.03.

23. Wiciak J. (2008), Sound radiation by set of L-jointed plates with four pairs of piezoelectric elements, *The European Physical Journal Special Topics*, **154**: 229–233, https://doi.org/10.1140/epjst/e2008-00551-0.

24. Wiciak J., Trojanowski R. (2014a), The effect of material composition of piezoelectric elements with chosen shapes on plate vibration reduction, *Acta Physica Polonica A*, **125**(4A): A-179–A-182, https://doi.org/10.12693/APhysPolA.125.A-179.

25. Wiciak J., Trojanowski R. (2015), Evaluation of the effect of a step change in piezo actuator structure on vibration reduction level in plates, *Archives of Acoustics*, **40**(1): 71–79, https://doi.org/10.1515/aoa-2015-0009.

26. Wiciak M., Trojanowski R. (2014b), Numerical analysis of the effectiveness of two-part piezoactuators in vibration reduction of plates, *Acta Physica Polonica A*, **125**(4A): A-183–A-189, https://doi.org/10.12693/APhysPolA.125.A-183.

27. Wrona S., Pawełczyk M., Cheng L. (2022), Sound transmission through a panel with a hybrid active and semi-active control system, *Journal of Sound and Vibration*, **536**: 117172, https://doi.org/10.1016/j.jsv.2022.117172.

Research Paper

Experimental and Computational Analysis on the Acoustic Performance of Side Outlet Muffler with Semi-Circular Baffles

Sandeep Kumar VISHWAKARMA, Suryappa Jayappa PAWAR*

Department of Applied Mechanics, Motilal Nehru National Institute of Technology Allahabad
Prayagraj, India

*Corresponding Author e-mail: sjpawar@mnnit.ac.in

In this study, the impacts of the inclusion of two semi-circular baffles and their orientations on the acoustic performance of the side outlet muffler have been investigated. The side outlet muffler has a circular simple expansion chamber with an axial inlet and a side outlet and two semi-circular baffles that have been placed inside the expansion chamber at different orientations. The axis of the outlet is at the right angle to the axis of the inlet. The acoustical investigation of the side outlet muffler with two semi-circular baffles is done using the plane wave analysis, the finite element method (FEM), and the two-load technique. Based on the orientations of the two semi-circular baffles, three different models of side outlet muffler with semi-circular baffles have been investigated. The plane wave analysis, FEM, and two-load method are applied to all models and it is found that analytical, computational, and experimental transmission loss (TL) are in good agreement. The analytical modelling successfully predicts the presence of semi-circular baffles in the form of peaks and troughs in the TL of side outlet muffler with semi-circular baffles before the cut-off frequency and thus proves its effectiveness. Among all the models, model_2 gives 42 % higher TL than model_1 and model_3 shows 16.20 % higher TL than model_2. Hence, model_3 proves to be the best design for the side outlet muffler with semi-circular baffles in the attenuation of noise. The model_3 is effective for 1030 Hz–1480 Hz, 1500 Hz–1570 Hz, and 1640 Hz–2400 Hz frequency sound waves. The TL curve, sound pressure contours for model_3, and the band power variation in the 1/3 octave band for all the models have also been presented.

Keywords: band power; finite element method; semi-circular baffles; side outlet muffler; sound pressure level; transmission loss; two-load method.

1. Introduction

Reactive mufflers (REDDY *et al.*, 2017) work by decreasing exhaust noise by adjusting the volume based on sound reflections (MUNDHE, DEORE, 2015). The plane wave effect (DAVIS *et al.*, 1954) has been used to analyze mufflers. MUNJAL (1975) suggested an enhanced 4-pole transfer matrix based on fluid dynamics. Some researchers (IGARASHI, 1958; JEONG *et al.*, 2015; XIANG *et al.*, 2016) also used the 4-pole transfer matrix to forecast the acoustic performance of mufflers. SULLIVAN and CROCKER (1978) established linked equations for the inner and outer pipes of a perforated muffler (SULLIVAN, 1979). The solution for the linked equations was reported by JAYARAMAN and YAM (1981).

As the solutions were based on the one-dimensional wave theory, without capturing the effect of higher-order modes of the wave. IH and LEE (1985; 1987) investigated a muffler with an expansion chamber at a higher-order mode to solve the shortcomings of the plane wave theory. ÅBOM (1990) created a transfer matrix for an expanded muffler with a circular portion considering the higher-order modes. MUNJAL (1987) used a numerical method (analytic method) to make the calculation procedure simple, but there was a limitation in the area ratio (only integer values were allowed). When the inlet and the outlet were perpendicular to each other, the analytical method was not able to analyze that muffler. IH (1992) proposed a numerical methodology for the evaluation of the acoustic char-

acteristics of two designs of the muffler. The mufflers with rectangular and circular cross-sections were taken into consideration. CHANG and CHIU (2014) suggested a simplified artificial neural network (ANN) model in combination with the boundary element method (BEM) (SEYBERT, CHENG, 1987; WANG, 1999), the finite element method (FEM) (ZHENG *et al.*, 2012), and the evolutionary technique to optimize the design of the rectangular-shaped mufflers. Simple baffles were also incorporated into this rectangular muffler. YI and LEE (1986) analyzed the excitation of the higher-order modes in a circular expansion chamber with the side inlet and side outlet muffler. MUNJAL (1997) derived the 4-pole parameters for the side inlet and side outlet elements and compared their performance with the extended inlet and extended outlet. KULKARNI and IN-GLE (2018) studied the consequence of the placement of the outlet section on the transmission loss (TL) (ZHAO, LI, 2022; GORAZD, 2021; MOHAMAD *et al.*, 2021) of the reactive muffler having two expansion chamber.

The TL is typically measured using 4-pole methods or 3-point (decomposition method), the two-load method, and the two-source method (TAO, SEYBERT, 2003). The popular computational softwares for evaluation of the TL of the mufflers are Actran and COMSOL Multiphysics (using FEM (MIMANI, MUN-JAL, 2012; YU *et al.*, 2016; NARAYANA, MUNJAL, 2005)), COUSTYX (using BEM (NARAYANA, MUN-JAL, 2005)), and Ricardo wave (using transfer matrix techniques (GUPTA, 2016a)). The baffles are used in the expansion chamber to suppress unwanted sound by directing the exhaust gas to travel a longer distance (DAS *et al.*, 2022). The usage of the baffles in the mufflers has been shown to improve the TL (DONE *et al.*, 2014; GUPTA, 2016b; LE ROY, 2011) by more than 50 %. LE ROY (2011) used harmonic BEM with LMS's Virtual Lab Acoustics module to explore the effect of interior baffles. HOROUB (2011) investigated the multiple connected expansion chambers and documented the impact of the tapered expansion chambers. The effect of the hole pattern on the TL has also been examined by GUPTA and TIWARI (2015). The flow of exhaust gases along with the sound wave through the muffler also influence the performance of the muffler. The mean flow of gases inside the muffler causes a general flattening of TL curve, where all the peaks and troughs were almost completely removed and there was a general decrease in TL (FAIRBROTHER, VARHOS, 2007; SIANO *et al.*, 2010; CHENG, WU, 1999). The other effect of the mean flow is the aerodynamic noise generated due to the turbulence of flow of the gases inside the muffler (LE ROY, 2011). The Mach number in the flow of exhaust gases of automotive engine is generally in between 0.1 and 0.3 (MUNJAL, 1975; FAN, GUO, 2016), and the aerodynamic noise is considered to be weak. Therefore, the aerodynamic noise may be disregarded (IH, LEE, 1987; MIMANI, MUNJAL, 2011).

The effect of mean flow on the performance of muffler is not considered.

The previously published research (DAS *et al.*, 2022; DONE *et al.*, 2014; GUPTA, 2016b; LE ROY, 2011) provides the motivation to use baffles in the side outlet muffler. These studies have shown that baffles can increase the TL of the muffler by a reasonable amount. The baffles are used to change the impedance and thus promoting the reflection of the sound waves (DAS *et al.*, 2022). This reflection of sound wave can cause the destructive interference with the incoming sound wave and hence reduces the intensity of sound wave. Additionally, baffles also deflect the sound waves to travel longer paths and hence dissipating the acoustic energy of the sound waves (GUPTA, 2016b). Based on the suppression of noise of different frequencies, the baffles can have the different shapes and the orientations inside the expansion chamber of mufflers. The reflection of sound waves depends upon the area discontinuity introduced due to the baffles (ELSAYED *et al.*, 2017), orientation of the baffles, and spacing between the baffles (DAS *et al.*, 2022). In this work, baffles are designed to decrease the cross-sectional area of the expansion chamber by 50 % at the area discontinuity. Since, the expansion chamber of side outlet muffler has a circular cross-section, the semi-circular shaped baffles are used in this study. Along with the area discontinuity due to the semi-circular baffles, the effect of their orientation inside the expansion chamber is also examined.

The impacts of the two semi-circular baffles and their orientation on the acoustical performance of the side outlet muffler have been investigated in this study. Based on the orientation of the semi-circular baffles, three models of the side outlet muffler with semi-circular baffles (SOMWB) have been investigated in this study. The models differ in the orientation of the two semi-circular baffles along the length of the simple expansion chamber of the side outlet muffler. The plane wave analysis, FEM, and two load method are used for the analytical modelling, computational modelling, and experimentation of all the models of the SOMWB, respectively. The analytical and computational results are compared with the experimental results and a fair agreement is found. The band power variation in the $1/3$ octave band for all the cases of the SOMWB and sound pressure level (SPL) contours at the desired frequencies for the best model of the SOMWB (model_3) have also been presented.

2. Analytical modelling

The analytical modelling of a plane traveling wave uses the continuity condition and mass conservation. The pressure for a one-dimensional traveling wave is the superposition of an incident wave and a reflected wave. The pressure equation contains separate terms for the incident and reflected waves. The modelling is

done for the sound wave travelling inside the SOMWB. The SOMWB is divided into A, B, C, D, E, F, and G regions and incident and reflected waves for the different regions have been shown in Fig. 1. The distance between the two semi-circular baffles, thickness of baffles, distance of the baffles from the ends of the expansion chamber, length of baffles, distance of the right side baffle from the centre of the outlet pipe, length of the expansion chamber, and diameter of the expansion chamber are represented by d, w, s, l, and t, a_x, and a_y, respectively. The A-region is from $x = -s - w$ to $x = -s$, B-region is from $x = -s$ to $x = 0$, C-region is from $x = 0$ to $x = w$, D-region is from $x = w$ to $x = w + d$, E-region is from $x = w + d$ to $x = 2w + d$, F-region is from $x = 2w + d$ to $x = 2w + d + t$, and G-region is from $x = 2w + d + t$ to $x = 2w + d + s$. The coefficients for incident and reflected wave in different regions are given by A^+, A^-, B^+, B^-, C^+, C^-, D^+, D^-, E^+, E^-, F^+, F^-, G^+, and G^-, respectively.

The pressure fields and the velocity components in the A-region, B-region, C-region, D-region, E-region, F-region, and G-region can be written as (Lee *et al.*, 2019; Vishwakarma, Pawar, 2022):

$$P_A = A^+ e^{-ik(x+s)} + A^- e^{ik(x+s+w)}, \tag{1}$$

$$u_A = \frac{A^+ e^{-ik(x+s)} - A^- e^{ik(x+s+w)}}{\frac{c}{S_A}}, \tag{2}$$

$$P_B = B^+ e^{-ikx} + B^- e^{ik(x+s)}, \tag{3}$$

$$u_B = \frac{B^+ e^{-ikx} - B^- e^{ik(x+s)}}{\frac{c}{S_B}}, \tag{4}$$

$$P_C = C^+ e^{-ikx} + C^- e^{ik(x-w)}, \tag{5}$$

$$u_C = \frac{C^+ e^{-ikx} - C^- e^{ik(x-w)}}{\frac{c}{S_C}}, \tag{6}$$

$$P_D = D^+ e^{-ik(x-w)} + D^- e^{ik(x-w-d)}, \tag{7}$$

$$u_D = \frac{D^+ e^{-ik(x-w)} - D^- e^{ik(x-w-d)}}{\frac{c}{S_D}}, \tag{8}$$

$$P_E = E^+ e^{-ik(x-w-d)} + E^- e^{ik(x-2w-d)}, \tag{9}$$

$$u_E = \frac{E^+ e^{-ik(x-w-d)} - E^- e^{ik(x-2w-d)}}{\frac{c}{S_E}}, \tag{10}$$

$$P_F = F^+ e^{-ik(x-2w-d)} + F^- e^{ik(x-2w-d-t)}, \tag{11}$$

$$u_F = \frac{F^+ e^{-ik(x-2w-d)} - F^- e^{ik(x-2w-d-t)}}{\frac{c}{S_F}}, \tag{12}$$

$$P_G = G^+ e^{-ik(x-2w-d-t)} + G^- e^{ik(x-2w-d-s)}, \tag{13}$$

$$u_G = \frac{G^+ e^{-ik(x-2w-d-t)} - G^- e^{ik(x-2w-d-s)}}{\frac{c}{S_G}}. \tag{14}$$

By continuity condition in the B-region and C-region, the pressure at the points just before the discontinuity and just after the discontinuity is equal, i.e., at $x = 0$:

$$P_B = P_C, \tag{15}$$

$$B^+ + B^- e^{iks} = C^+ + C^- e^{-ikw}. \tag{16}$$

By mass conservation condition, the mass of the particle is constant across all the junctions. Since the density of the medium is taken as constant, instead of mass one can equate the volume particle velocity at the discontinuities, i.e., at $x = 0$:

$$u_B = u_C, \tag{17}$$

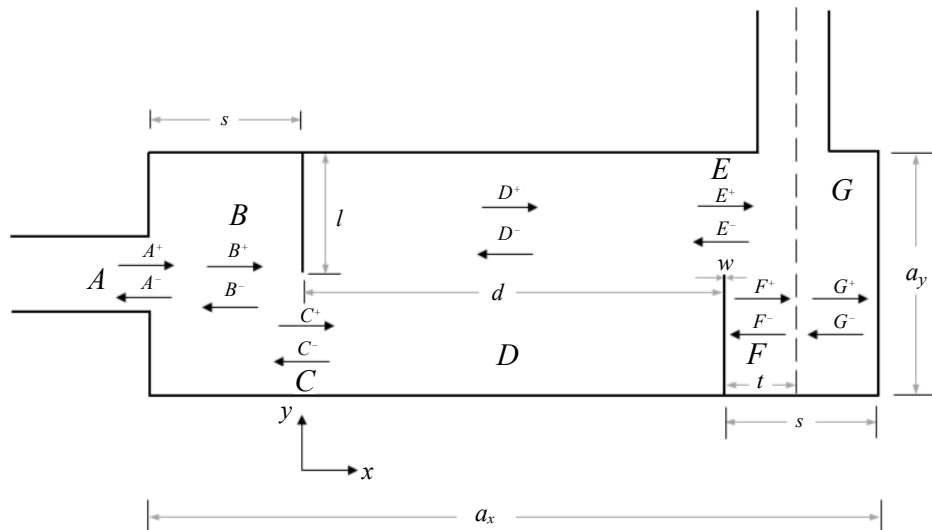$$B^+ - B^- e^{iks} = \left(C^+ - C^- e^{-ikw} \right) \frac{S_C}{S_B}. \tag{18}$$

Fig. 1. Different regions of side outlet muffler with semi-circular baffles.

From Eqs. (16) and (18), one can write the relation between coefficients in the B-region and C-region as

$$\begin{bmatrix} B^+ \\ B^- \end{bmatrix} = \begin{bmatrix} 0.5\left(1 + \dfrac{S_C}{S_B}\right) & -0.5\left(\dfrac{S_C}{S_B} - 1\right)e^{-ikw} \\ 0.5\left(1 - \dfrac{S_C}{S_B}\right)e^{-iks} & 0.5\left(1 + \dfrac{S_C}{S_B}\right)e^{-ik(w+s)} \end{bmatrix} \begin{bmatrix} C^+ \\ C^- \end{bmatrix}, \tag{19}$$

$$\begin{bmatrix} B^+ \\ B^- \end{bmatrix} = \begin{bmatrix} M_1 \end{bmatrix} \begin{bmatrix} C^+ \\ C^- \end{bmatrix}. \tag{20}$$

Similarly, relations among the other coefficient can be written as

$$\begin{bmatrix} A^+ \\ A^- \end{bmatrix} = \begin{bmatrix} 0.5\left(1 + \dfrac{S_B}{S_A}\right)e^{iks} & -0.5\left(\dfrac{S_B}{S_A} - 1\right) \\ 0.5\left(1 - \dfrac{S_B}{S_A}\right)e^{ik(s-w)} & 0.5\left(1 + \dfrac{S_B}{S_A}\right)e^{-ikw} \end{bmatrix} \begin{bmatrix} B^+ \\ B^- \end{bmatrix}, \tag{21}$$

$$\begin{bmatrix} A^+ \\ A^- \end{bmatrix} = \begin{bmatrix} M_2 \end{bmatrix} \begin{bmatrix} B^+ \\ B^- \end{bmatrix}, \tag{22}$$

$$\begin{bmatrix} C^+ \\ C^- \end{bmatrix} = \begin{bmatrix} 0.5\left(1 + \dfrac{S_D}{S_C}\right)e^{ikw} & -0.5\left(\dfrac{S_D}{S_C} - 1\right)e^{-ik(d+w)} \\ -0.5\left(1 - \dfrac{S_D}{S_C}\right)e^{2ikw} & 0.5\left(1 + \dfrac{S_D}{S_C}\right)e^{-ikd} \end{bmatrix} \begin{bmatrix} D^+ \\ D^- \end{bmatrix}, \tag{23}$$

$$\begin{bmatrix} C^+ \\ C^- \end{bmatrix} = \begin{bmatrix} M_3 \end{bmatrix} \begin{bmatrix} D^+ \\ D^- \end{bmatrix}, \tag{24}$$

$$\begin{bmatrix} D^+ \\ D^- \end{bmatrix} = \begin{bmatrix} 0.5\left(1 + \dfrac{S_E}{S_D}\right)e^{ikd} & -0.5\left(\dfrac{S_E}{S_D} - 1\right)e^{-ik(w-d)} \\ 0.5\left(1 - \dfrac{S_E}{S_D}\right) & 0.5\left(1 + \dfrac{S_E}{S_D}\right)e^{-ikw} \end{bmatrix} \begin{bmatrix} E^+ \\ E^- \end{bmatrix}, \tag{25}$$

$$\begin{bmatrix} D^+ \\ D^- \end{bmatrix} = \begin{bmatrix} M_4 \end{bmatrix} \begin{bmatrix} E^+ \\ E^- \end{bmatrix}, \tag{26}$$

$$\begin{bmatrix} E^+ \\ E^- \end{bmatrix} = \begin{bmatrix} 0.5\left(1 + \dfrac{S_F}{S_E}\right)e^{ikw} & -0.5\left(\dfrac{S_F}{S_E} - 1\right)e^{ik(w-t)} \\ 0.5\left(1 - \dfrac{S_F}{S_E}\right) & 0.5\left(1 + \dfrac{S_F}{S_E}\right)e^{-ikt} \end{bmatrix} \begin{bmatrix} F^+ \\ F^- \end{bmatrix}, \tag{27}$$

$$\begin{bmatrix} E^+ \\ E^- \end{bmatrix} = \begin{bmatrix} M_5 \end{bmatrix} \begin{bmatrix} F^+ \\ F^- \end{bmatrix}, \tag{28}$$

$$\begin{bmatrix} F^+ \\ F^- \end{bmatrix} = \begin{bmatrix} 0.5\left(1 + \dfrac{S_G}{S_F}\right)e^{ikt} & -0.5\left(\dfrac{S_G}{S_F} - 1\right)e^{ik(2t-s)} \\ 0.5\left(1 - \dfrac{S_G}{S_F}\right) & 0.5\left(1 + \dfrac{S_G}{S_F}\right)e^{ik(t-s)} \end{bmatrix} \begin{bmatrix} G^+ \\ G^- \end{bmatrix}, \tag{29}$$

$$\begin{bmatrix} F^+ \\ F^- \end{bmatrix} = \begin{bmatrix} M_6 \end{bmatrix} \begin{bmatrix} G^+ \\ G^- \end{bmatrix}, \tag{30}$$

where $[M_1]$, $[M_2]$, $[M_3]$, $[M_4]$, $[M_5]$, and $[M_6]$ are the coefficient matrices among different regions. The relation between the coefficient of A-region and G-region can be obtained by the multiplication of the coefficient matrices of different regions. The relation can be written as

$$\begin{bmatrix} A^+ \\ A^- \end{bmatrix} = [M_2][M_1][M_3][M_4][M_5][M_6]\begin{bmatrix} G^+ \\ G^- \end{bmatrix}, \quad (31)$$

$$\begin{bmatrix} A^+ \\ A^- \end{bmatrix} = [GM]\begin{bmatrix} G^+ \\ G^- \end{bmatrix}. \quad (32)$$

In G-region, the velocity component at $x = 2w+d+s$ becomes zero. Therefore, u_G can be written as

$$\frac{G^+ e^{-ik(x-2w-d-t)} - G^- e^{ik(x-2w-d-s)}}{\frac{c}{S_G}} = 0, \quad (33)$$

$$G^- = G^+ e^{-ik(s-t)}. \quad (34)$$

By using the relation between G^- and G^+ in Eq. (32), one can write

$$\begin{bmatrix} A^+ \\ A^- \end{bmatrix} = [GM]\begin{bmatrix} 1 \\ e^{-ik(s-t)} \end{bmatrix} G^+. \quad (35)$$

From Eq. (35), one can easily obtain the ratio of A^+ to G^+. Therefore the TL of the SOMWB can be given by (Vishwakarma, Pawar, 2021):

$$\text{TL} = 20\log_{10}\left(\frac{A^+}{G^+}\right). \quad (36)$$

3. Computational modelling

The FEA has been used for the computational modelling of SOMWB. The computational modelling of mufflers requires pressure and velocity equations (Lee *et al.*, 2019). The TL is evaluated with the help of the pressure and velocity equations. The governing equations in ANSYS (2022) are used for the acoustic modelling of muffler with assumptions such as absence of mean flow, no fluid-structure interaction, compressible and irrotational fluid medium, and an anechoic termination at the outlet.

In acoustic modelling, Navier–Stokes (NS) equations of fluid momentum and the continuity are used for the formulation of acoustic wave equations (ANSYS, 2022).

The continuity equation is

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho \mathbf{v}) + Q, \quad (37)$$

where ρ, t, $\mathbf{v}$, and Q are the density, time, velocity vector in the x-, y-, and z-directions, and mass source (kg/m^{3}t), respectively.

The Navier–Stokes equation is

$$\rho \frac{d\mathbf{v}}{dt} = -\nabla p + \nabla \cdot \overline{\overline{S}} + \rho \mathbf{b}, \quad (38)$$

where p, $\overline{\overline{S}}$, and $\mathbf{b}$ are pressure, viscous stress tensor, and body force, respectively.

The continuity and NS equations are linearized to obtain the acoustic wave equation. The resulting linearized forms are as follows:

$$\nabla \cdot \mathbf{v}_a = -\frac{1}{\rho c^2}\frac{\partial^2 p_a}{\partial t^2} + \frac{Q}{\rho}, \quad (39)$$

$$\frac{\partial \mathbf{v}_a}{\partial t} = -\frac{1}{\rho}\nabla p_a + \frac{4}{3}\frac{\mu}{\rho}\nabla\left(-\frac{1}{\rho c^2}\frac{\partial p_a}{\partial t} + \frac{Q}{\rho}\right), \quad (40)$$

where $\mathbf{v}_a$, p_a, c, and μ are the acoustic velocity, acoustic pressure, speed of sound, and dynamic viscosity of the medium, respectively.

From Eqs. (39) and (40) acoustic wave equation can be formulated as

$$\nabla \cdot \left(\frac{1}{\rho}\nabla p_a\right) - \frac{1}{\rho c^2}\frac{\partial^2 p_a}{\partial t^2} + \nabla \cdot \left[\frac{4}{3}\frac{\mu}{\rho}\nabla\left(\frac{1}{\rho c^2}\right)\frac{\partial p_a}{\partial t}\right]$$
$$= -\frac{\partial}{\partial t}\left(\frac{Q}{\rho}\right) + \nabla \cdot p\left[\frac{4}{3}\frac{\mu}{\rho}\nabla\left(\frac{Q}{\rho}\right)\right]. \quad (41)$$

The SPL is given by Eq. (42), where p_{rms} and p_{ref} are the root mean square of peak pressure in harmonic analysis and reference pressure $(2 \times 10^{-5}\,\text{Pa})$, respectively:

$$\text{SPL} = 10\log_{10}\left(\frac{p_{\text{rms}}^2}{p_{\text{ref}}^2}\right). \quad (42)$$

The TL is given by Eq. (43), where

$$P_{in} = \frac{p_{in}^2}{\rho c} \quad \text{and} \quad P_t = \frac{p_o^2}{\rho c}$$

are the incident sound power at the inlet and the transmitted sound power at the outlet, and p_{in} and p_o are the sound pressure at the inlet and at outlet, respectively:

$$\text{TL} = 10\log_{10}\left(\frac{P_{in}}{P_t}\right). \quad (43)$$

The radiation boundary condition is applied at the inlet and outlet of the muffler. The radiation boundary condition takes care of the anechoic termination. The surface velocity condition is applied at the inlet of the muffler to create the source of the disturbance in the medium. The walls of the domain and semi-circular baffles are set to the rigid wall to apply the Neumann boundary conditions. Ports are defined to calculate the TL of the muffler.

3.1. Side outlet muffler with semi-circular baffles (SOMWB)

The proposed SOMWB is the outcome of the inclusion of the semi-circular baffles and adjustments made in the position of the outlet in a simple expansion chamber (SEC) muffler (Tao, Seybert, 2003). The position of the outlet and orientation of two semi-circular baffles are shown in Figs. 2a–2c. There are three variations in the orientation of the two semi-circular baffles inside the SOM. Based on the variations, models such as model_1, model_2, and model_3 are analysed and shown in Fig. 2. The orientations of semi-circular baffles in model_1 and model_2 are the same and can be seen in *X-X* planes of Figs. 2a

and 2b. In model_1, the first semi-circular baffle is placed on the upper half of the expansion chamber, and the second semi-circular baffle is placed on the lower half of the expansion chamber. Whereas, in model_2, the first semi-circular baffle is placed on the lower half of the expansion chamber, and the second semi-circular baffle is placed on the upper half of the expansion chamber. On the other hand, model_3 has a different orientation of semi-circular baffles and is shown in Fig. 2c. The orientation of semi-circular baffles can be seen in the *X-X* plane of the model_3 of SOMWB. The first semi-circular baffle is placed on the left half and second semi-circular baffle is placed on the right half of the expansion chamber. The analysis is carried out in the harmonic acoustic module. The com-
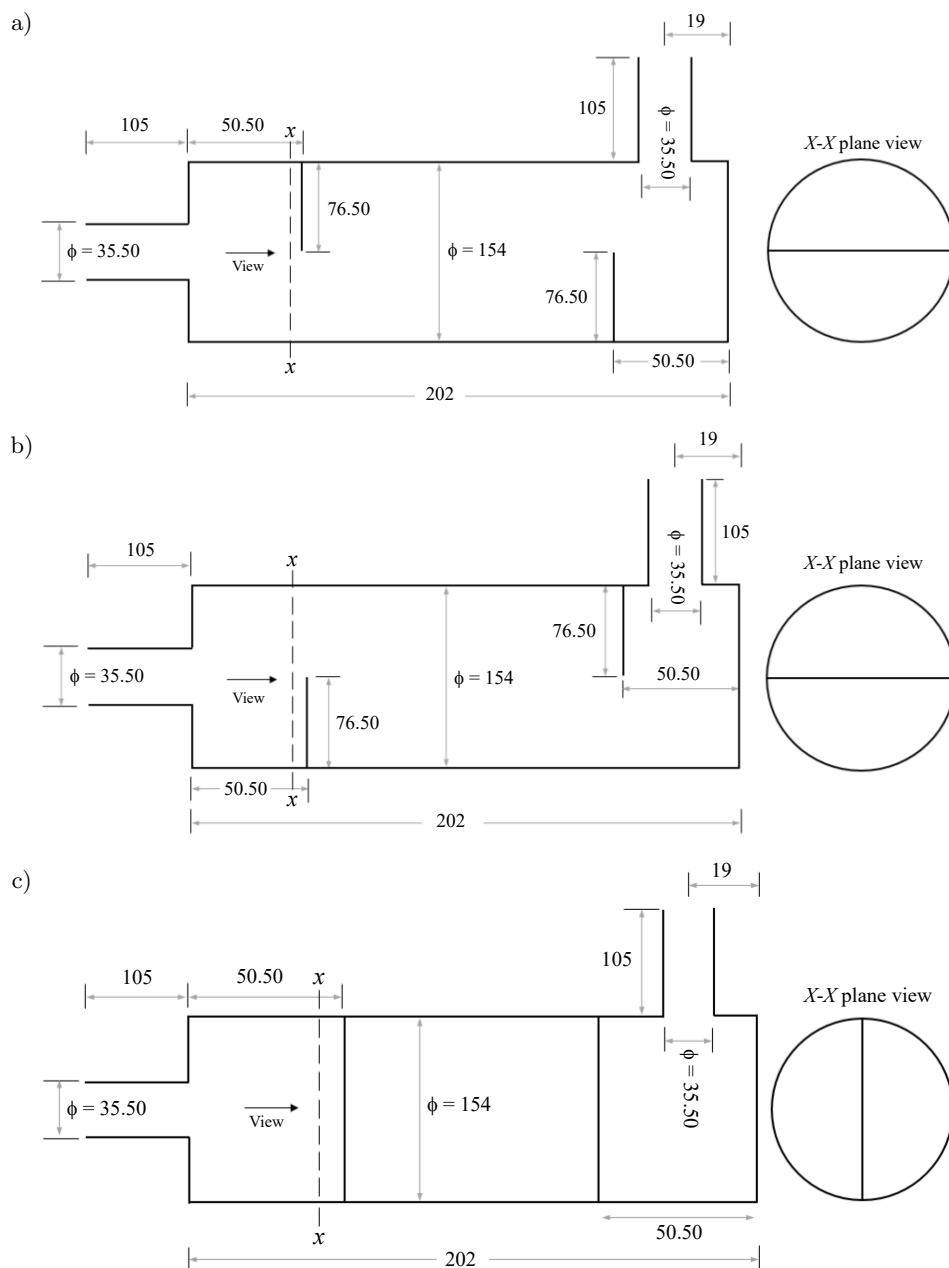

Fig. 2. Computational domains of (a) model_1, (b) model_2, and (c) model_3 of SOMWB (all dimensions are in mm).

putational domains of all the models of SOMWB have been discretized into FLUID 221 elements (10-noded acoustic element). FLUID 221 elements are three-dimensional tetrahedral elements. These elements are useful in modelling the medium for the sound wave propagation. The radiation condition is imposed on the inlet and outlet ports of the SOMWB. The disturbance in the medium is created by applying a surface velocity (equal to 10 m/s) at the inlet. The walls and semi-circular baffles of the SOMWB are set to rigid wall boundary conditions (Vishwakarma, Pawar, 2022). The inlet and outlet ports are defined to calculate the incoming and outgoing rate of sound energy. The calculations of the rate of sound energy at the ports directly determine the TL.

4. Experimental analysis

In this study, a two load method is used for the TL calculation of all models of SOMWB. This method comprises 4 microphones, among which 2 microphones

(1 and 2) are positioned along the inlet section and the other 2 microphones (3 and 4) are positioned along the outlet section to determine the progressive and reflective waves. The wave strengths of the decomposed waves are denoted by A, B, C, and D. The wave strengths and the positions of the microphones are shown in Fig. 3. Closed end termination and anechoic termination are the two loads applied in this study for the estimation of TL. Initially, the transfer functions such as $H_{1,\mathrm{ref}}$, $H_{2,\mathrm{ref}}$, $H_{3,\mathrm{ref}}$, and $H_{4,\mathrm{ref}}$ are measured for each load. After evaluation of the transfer functions, 4-pole parameters of the muffler are derived and thereafter the calculation of the TL is done. The 4-pole parameters are derived from the pressure and velocity equations of the wave. As 2 equations are not sufficient for the estimation of 4 pole parameters, 2 different loading conditions are used. For the first load, it is done by closing the outlet using a tube closed at one end, and is shown in Fig. 3a. For the second load, it is closed by using a tube filled with polyester felt (for anechoic termination) and is shown in Fig. 3b.
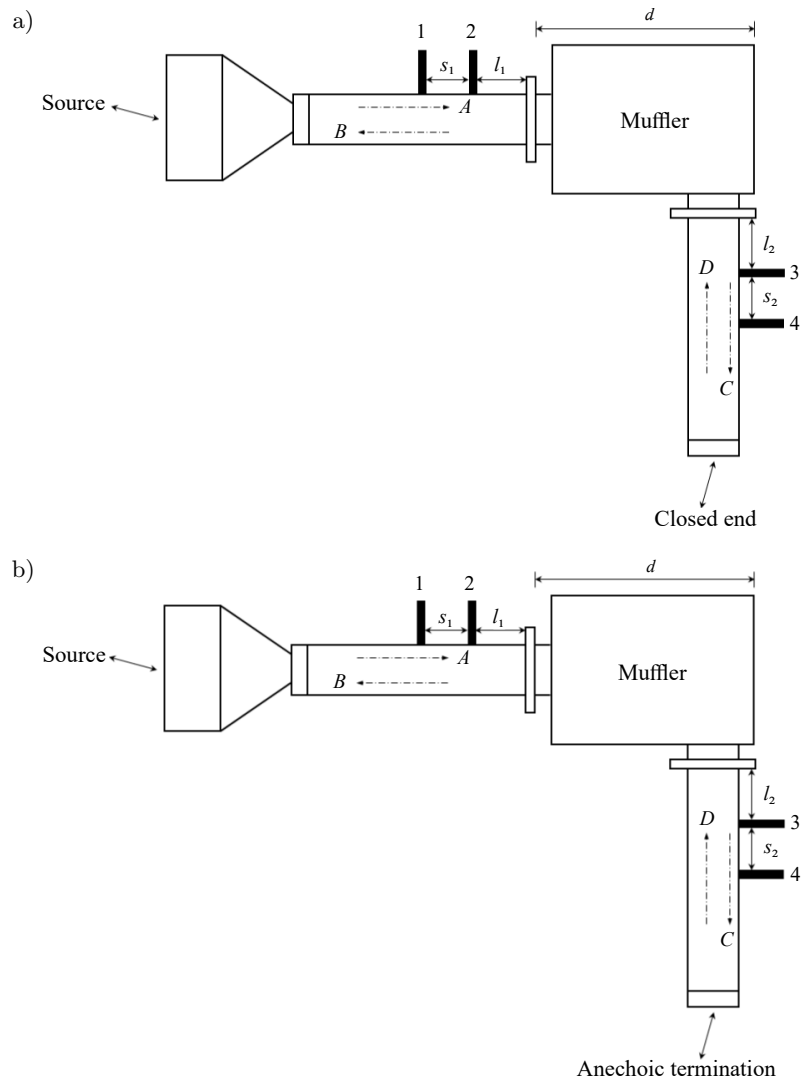

Fig. 3. Two-load setup with (a) closed-end termination and (b) an anechoic termination for estimation of TL of the muffler.

White noise is generated from the speaker and it ranges from 10 Hz to 5000 Hz. After the calibration of the microphones, readings were taken.

Wave decomposition data processing is the most popular method to measure the transfer functions to obtain the sound TL. The pressure relation for one-dimensional moving wave can be given by (Hua, Herrin, 2013):

$$P(x) = A * e^{-jkx} + B * e^{jkx}. \qquad (44)$$

The particle velocity for the 1-D moving wave can be given by (Hua, Herrin, 2013):

$$V(x) = (A * e^{-jkx} - B * e^{jkx})/(\rho c). \qquad (45)$$

By applying wave decomposition at both the upstream and downstream tubes, the wave strength A, B, C, and D can be determined as

$$A = j \frac{H_{1,\text{ref}} e^{-jkl_1} - H_{2,\text{ref}} e^{-jk(l_1 + S_1)}}{2 * \sin(kS_1)}, \qquad (46)$$

$$B = j \frac{H_{2,\text{ref}} e^{-j(l_1 + S_1)} - H_{1,\text{ref}} e^{-jkl_1}}{2 * \sin(kS_1)}, \qquad (47)$$

$$C = j \frac{H_{3,\text{ref}} e^{jk(l_2 + S_2)} - H_{4,\text{ref}} e^{jkl_2}}{2 * \sin(kS_2)}, \qquad (48)$$

$$D = j \frac{H_{4,\text{ref}} e^{jkl_2} - H_{3,\text{ref}} e^{jk(l_2 + S_2)}}{2 * \sin(kS_2)}, \qquad (49)$$

where k, x, ρ, and c are wave number, direction of wave propagation, density of medium, and velocity of sound in the medium, respectively.

For different loads, the pressure and particle velocity at two ends of the muffler are expressed as

$$p_0 = A + B, \qquad (50)$$

$$u_0 = (A - B)/\rho c, \qquad (51)$$

$$p_d = C e^{-jkd} + D e^{jkd}, \qquad (52)$$

$$u_d = (C e^{-jkd} - D e^{jkd})/\rho c, \qquad (53)$$

where d is the length of the muffler. The transfer matrix based on pressure and particle velocity can be written as

$$\mathbf{T} = \begin{bmatrix} \left[\dfrac{p_{0a} u_{db} - p_{0b} u_{da}}{p_{da} u_{db} - p_{db} u_{da}} \right] & \left[\dfrac{p_{0b} p_{da} - p_{0a} p_{db}}{p_{da} u_{db} - p_{db} u_{da}} \right] \\ \left[\dfrac{u_{0a} u_{db} - u_{0b} u_{da}}{p_{da} u_{db} - p_{db} u_{da}} \right] & \left[\dfrac{p_{da} u_{0b} - p_{db} u_{oa}}{p_{da} u_{db} - p_{db} u_{da}} \right] \end{bmatrix}, \qquad (54)$$

where the subscripts a (anechoic termination) and b (closed end) represent the loads. Then, the TL is expressed as

$$\text{TL} = 20 \log_{10} \left| \frac{1}{2} \left(T_{11} + \frac{T_{12}}{\rho c} + \rho c T_{21} + T_{22} \right) \right|. \qquad (55)$$

The impedance tube (Alfa Acoustic's tube) is used to evaluate the sound TL in accordance with ASTM E2611 standards. The experimental setup consists of several components such as impedance tubes, microphones, speaker, power amplifier, and a data acquisition system. The impedance tube is circular with an internal and external diameter of 34.90 mm and 40.90 mm. The microphones are pressure field microphones with a diameter of 0.63 mm. A loudspeaker (15 W, 6 Ohm, and 101.6 mm diameter) is used as the source of sound. All measurements are done at ambient conditions. The transfer function method is used in the measurement software. The environmental condition for impedance tube is set with the help of a software. The details of the parameter setup in the impedance tube during the experimentation is given in Table 1. The room temperature, relative humidity, and atmospheric pressure at the time of experimentation were found as 25.7 °C, 55 %, and 101.7 kPa, respectively.

In impedance tube, speed of sound is kept as 347.7 m/s and density of air is 1.183 kg/m³. Three design models of SOMWB have been fabricated using a 2 mm thick SS 409 and are shown in Fig. 4. The fabricated

Table 1. Impedance tube parameters during experimentation.

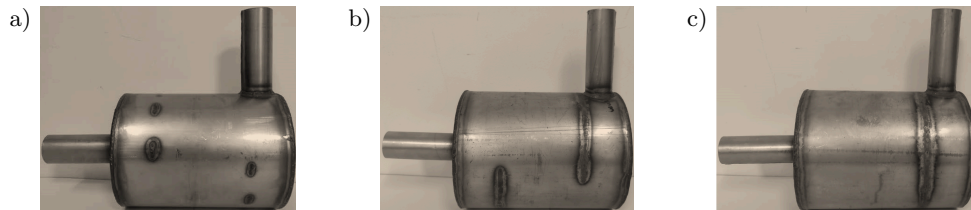| Impedance tube parameters	
Frequency range [Hz]	10–5000
Distance between microphones (S_1 and S_2) [mm]	30
Distance between muffler to upstream microphone (l_1) [mm]	90
Distance between muffler to downstream microphone (l_2) [mm]	240
Tube diameter [mm]	45

Fig. 4. Fabricated (a) model_1, (b) model_2, and (c) model_3 of the SOMWB.

Fig. 5. Experimental setup for TL measurement
of the fabricated mufflers.

muffler is then attached to the impedance tube setup
and is shown in Fig. 5, and henceforth, the TL of the
muffler is calculated.

5. Results and discussion

5.1. Transmission loss (TL)

The comparison of analytical TL, computational
TL, and experimental TL curves for model_1, model_2,
and model_3 of SOMWB has been shown in Fig. 6.
It can be seen that the orientation of the semi-circular

baffles inside the expansion chamber of SOMWB
has a serious impact on its TL. The analytical TL
curve is the same for the all the models of SOMWB
because the analytical modelling does not account
for the orientation of the semi-circular baffles as this
analysis is one-dimensional. The analytical modelling
of the SOMWB is compared with the previously
published work (VISHWAKARMA, PAWAR, 2024) on
the simple expansion chamber muffler with the side
outlet (SECMSO). The comparison indicates that the
analytical modelling of SOMWB successfully predicts
the presence of semi-circular baffles in the form of in-
creased TL. It also shows that the TL of the SOMWB
is greater than TL of SECMSO throughout the entire
frequency range except between 1740 Hz–1770 Hz.
Analytically, the TL of SOMWB has its maximum
value 38.60 dB at 880 Hz. There is an increment of
16 dB in the TL of SOMWB when compared with the
SECMSO. From Fig. 6, it can be depicted that
the maximum TL's of model_1, model_2, and model_3
obtained from computational modelling are 41 dB
(980 Hz), 63.71 dB (2194 Hz), and 75.47 dB (1520 Hz),
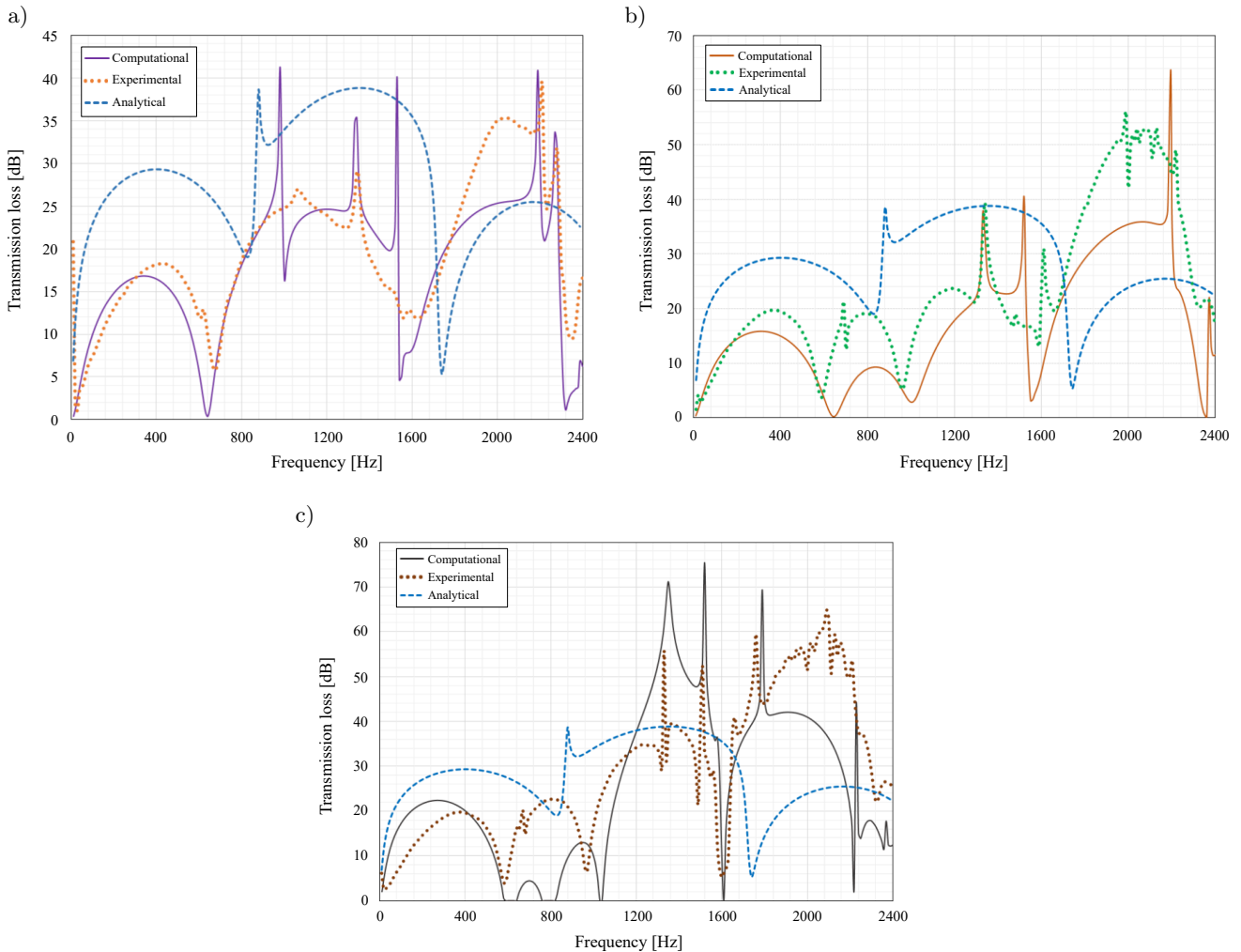respectively. The maximum TL of model_1, model_2,

Fig. 6. Comparison of analytical TL, computational TL, and experimental TL for: a) model_1; b) model_2; c) model_3.

and model_3 obtained from experiment are 39.27 dB (2210 Hz), 55.77 dB (1990 Hz), and 64.81 dB (2090 Hz), respectively. This is because of the presence of a multi-chambered expansion chamber in the SOMWB. The semi-circular baffles enhance the effectiveness of the muffler in the mid and high-frequency range (Elsayed *et al.*, 2017). The values of the TL from FEA and experimentation at peaks and troughs for all the models are listed in Tables 2 and 3, respectively. The order of frequencies in a single cell of the table is the frequency for the maximum TL followed by the frequency for the minimum TL. The percentage errors in the comparison of maximum TLs obtained from FEA and experimentation for model_1, model_2, and model_3 are 4.73 %, 12.46 %, and 14.12 %, respectively, and are acceptable. The comparison of the TL curves of SOMWB shows that the analytical modelling of SOMWB is able to capture all the peaks and trough in the TL of SOMWB up to the cut-off frequency. The cut-off frequency (Ih, Lee, 1985; 1987; Åbom, 1990) for all the models of SOMWB comes out to be 1311 Hz. The higher-order modes start to propagate above 1311 Hz. The propagation of the higher-order modes of sound waves can be seen in the FEA and experimental TL curve in the form of abrupt peaks. But the analytical TL curve does not have these peaks because it has the assumption of plane wave propagation only. From Fig. 6a, it can be observed that there are 2 peaks (340 Hz and 980 Hz) and 2 troughs (640 Hz and 1000 Hz) in the computational TL curve before the cut off frequency. The analytical TL curve also shows the 2 peaks (400 Hz and 880 Hz) and 2 troughs

(830 Hz and 920 Hz) in the TL with a slight shift in the frequencies for peaks and troughs. The values of TL in the analytical and computational curve are different due to the one- and three-dimensional analysis of the SOMWB, respectively. The experimental values are lower than the FEA values at specific frequencies because the incident sound pressure is completely absorbed by the end outlet in the computational investigation, but the experiment produces reflected sound pressure and other sounds. This is because FEA calculations are based on ideal boundary conditions. Limitations in the accuracy of microphones, manufacturing errors, experimental setup flaws, different environmental conditions, and leakage of sound waves through the connections are the other reasons for the difference in the experimental and FEA values. Although the analytical TL curve and computational TL curve deviate somewhat from the experimental TL curve, the general agreement is good, giving credible information for future investigations identifying the effect mechanism of complicated mufflers. From Table 3 it can be observed that the experimental TL curves for the model_1, model_2, and model_3 have the maximum TL as 39.27 dB, 55.77 dB, and 64.81 dB, respectively. The performance of model_2 is better (42 % higher TL) than model_1 because of the orientation of the second semi-circular baffle. In model_2, the second semi-circular baffle guides the sound wave to travel a longer path as compared to model_1. Similarly, model_3 shows better performance (16.21 % higher TL) than model_2. Therefore, model_3 proves to be the best design for SOMWB in the attenuation of

Table 2. TL values from FEA for all the models at different frequencies.

TL/Frequencies		340/640 [Hz]	980/1000 [Hz]	1340/1490 [Hz]	1530/1540 [Hz]	2190/2220 [Hz]	2270/2320 [Hz]
model_1	Max. TL [dB]	16.81	41.22	35.31	39.76	40.88	33.61
	Min. TL [dB]	0.40	16.45	19.80	4.63	20.96	1.19
TL/Frequencies		310/640 [Hz]	1330/1440 [Hz]	1520/1550 [Hz]	2060/2140 [Hz]	2194/2360 [Hz]	2370/2400 [Hz]
model_2	Max. TL [dB]	15.84	37.75	40.37	35.88	63.71	21.74
	Min. TL [dB]	0.28	22.70	3.14	35.40	0.17	11.57
TL/Frequencies		270/590 [Hz]	950/1030 [Hz]	1350/1470 [Hz]	1520/1570 [Hz]	1790/1830 [Hz]	1910/2220 [Hz]
model_3	Max. TL [dB]	22.41	13.03	71.22	75.47	69.41	42.08
	Min. TL [dB]	0	0	47.98	35.70	41.42	2.80

Table 3. TL values of all the fabricated models from the experiment at different frequencies.

TL/Frequencies		430/670 [Hz]	1070/1290 [Hz]	1340/1560 [Hz]	2040/2160 [Hz]	2210/2230 [Hz]	2280/2350 [Hz]
model_1	Max. TL [dB]	18.24	26.92	29.05	35.32	39.27	31.66
	Min. TL [dB]	5.78	22.32	11.86	33.49	24.70	9.39
TL/Frequencies		370/590 [Hz]	690/700 [Hz]	1340/1590 [Hz]	1610/1660 [Hz]	1990/2000 [Hz]	2360/2400 [Hz]
model_2	Max. TL [dB]	19.67	21.19	39.31	30.92	55.77	21.80
	Min. TL [dB]	3.70	12.50	13.05	19.44	42.10	16.91
TL/Frequencies		370/590 [Hz]	810/970 [Hz]	1330/1320 [Hz]	1510/1490 [Hz]	1760/1600 [Hz]	2090/2330 [Hz]
model_3	Max. TL [dB]	19.79	22.64	45.73	52.30	59.46	64.81
	Min. TL [dB]	4.07	7.35	29.28	21.58	5.29	22.43

sound waves in the 10 Hz–2400 Hz range. For model_1, the first, second, third, fourth, fifth, and sixth peaks in the experimental TL curve occur at 430 Hz, 1070 Hz, 1340 Hz, 2040 Hz, 2210 Hz, and 2280 Hz, respectively. The TL values at the respective peaks are 18.24 dB, 26.92 dB, 29.05 dB, 35.32 dB, 39.27 dB, and 31.66 dB. For the model_2, the first, second, third, fourth, fifth, and sixth peaks in the experimental TL curve occur at 370 Hz, 690 Hz, 1340 Hz, 1610 Hz, 1990 Hz, and 2360 Hz, respectively. The TL values at the respective peaks are 19.67 dB, 21.19 dB, 39.31 dB, 30.92 dB, 55.77 dB, and 21.80 dB. Similarly, the experimental TL curve for the model_3 has the first, second, third, fourth, fifth, and sixth peaks at 370 Hz, 810 Hz, 1330 Hz, 1510 Hz, 1760 Hz, and 2090 Hz, respectively. Experimental TL curves for all models indicate the effectiveness of the models in specific frequency ranges.

5.2. Band power

In the experimental setup, the band power of the sound wave detected by the microphones in the $^1/_3$ octave band is shown in Fig. 7. Band power in the $^1/_3$ octave bands helps in understanding the frequency distribution of noise. Mufflers are designed to attenuate specific frequencies of noise coming from internal combustion engines, so analyzing band power in specific frequency bands provides the effectiveness of mufflers in the suppression of noise across different frequency ranges. Microphone 1 and microphone 3 are chosen for capturing the band power of the progressive sound wave. The magnitude of the band power of the progressive wave detected by microphone 3 is having low value as compared to the microphone 1 for all the models. This is due to the attenuation provided by the muffler. There are fluctuations in the band power at the start. This is due to the weak signal from the loudspeaker at low frequencies, instability and insensitivity of the microphones in capturing
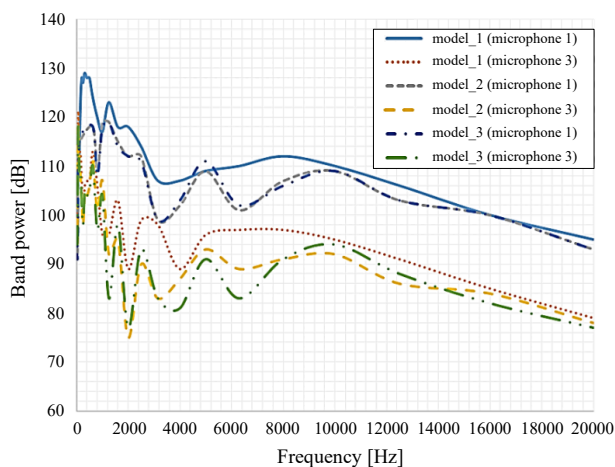
the sound wave at the initial stage. After the stability of the microphones, values given by them are within an acceptable range. The difference between the band powers detected by microphone 1 and microphone 3 shows the performance of all models of the SOMWB. The difference is maximum for model_3 of the SOMWB at 2000 Hz (octave middle frequency). The upper and lower frequencies for this band are 1782 Hz and 2245 Hz, respectively. This indicates that the design of the model_3 is suitable for the suppression of the noise in this frequency band. Band power curves for model_3 furthermore suggest other octave middle frequencies, such as 1000 Hz and 1600 Hz where model_3 is effective. In Fig. 6c, the experimental TL curve of the model_3 of SOMWB also indicates that model_3 is highly effective in 1030 Hz–1480 Hz (1000 Hz $^1/_3$ octave band), 1500 Hz–1570 Hz (1600 Hz $^1/_3$ octave band), and 1760 Hz–2400 Hz (1600 Hz $^1/_3$ octave band). The band power curves show that the difference between the band powers detected by microphone 1 and microphone 3 for model_2 of SOMWB is maximum at 1250 Hz and 2000 Hz (octave middle frequencies). Thus, it can be concluded from Fig. 6b and Fig. 7 that model_2 of SOMWB is effective in 1110 Hz–1260 Hz, 1310 Hz–1400 Hz, and 1690 Hz–2300 Hz. Similarly, from Fig. 6a and Fig. 7, it is concluded that model_1 of SOMWB is effective in 880 Hz–1360 Hz and 1800 Hz–2290 Hz.

5.3. Sound pressure level (SPL)

The SPLs can be presented in isolines form and contour form. The SPL isoline form provides information about the movement of the sound wave inside the SOMWB, whereas the contour form gives the intensity of the SPL inside the SOMWB. The SPL contours are presented for model_3 of the SOMWB at different frequencies and are shown in Fig. 8. The SPLs are presented at 270 Hz, 950 Hz, 1350 Hz, 1520 Hz, 1790 Hz, and 1910 Hz. Among the mentioned frequencies, 270 Hz and 950 Hz are below the cut-off frequency (1311 Hz) of the model_3 of SOMWB. Therefore, there is plain wave propagation of the sound wave, and can be seen in Figs. 8a and 8b. Other mentioned frequencies are above the cut-off frequency, hence higher order modes of sound wave start to propagate and the wave propagation inside the model_3 becomes non-planar. The SPL contours in Figs. 8c–f clearly show the non-planar movement of the sound wave. From Figs. 8a–f, it could be seen that the minimum SPL value is found at 1520 Hz. The SPL contour at 1520 Hz is shown in Fig. 8d. The maximum and minimum SPL at 1520 Hz are 163.52 dB and 53.69 dB respectively, which shows the effectiveness of model_3 of SOMWB. Therefore, at this frequency, noise reduction by the model_3 is maximum and is also supported by the computational TL curve data.
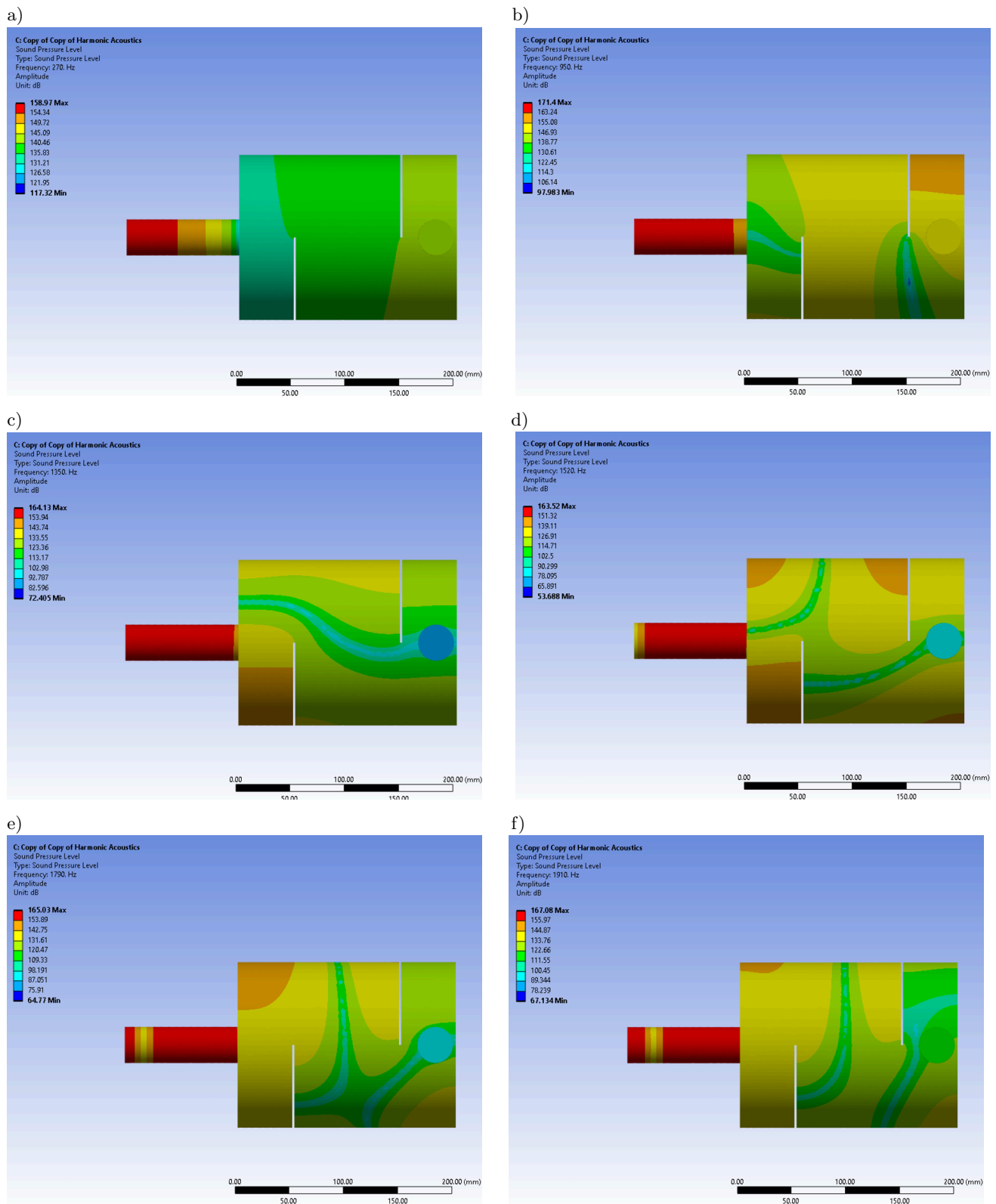

Fig. 7. Comparison of the band powers for model_1, model_2, and model_3 of SOMWB.

Fig. 8. SPLs at: a) 270 Hz; b) 950 Hz; c) 1350 Hz; d) 1520 Hz; e) 1790 Hz; f) 1910 Hz.

6. Conclusions

The analytical, computational, and experimental acoustical investigations of three models of SOMWB have been successfully done in this study. The analytical modelling of SOMWB is compared with previously published work on SECMSO and it is found that analytical modelling effectively predicts the presence of

semi-circular baffles inside the expansion chamber in the form of increased TL of SOMWB. It is found that there is an increment of 16 dB in the TL of SOMWB as compared to the TL of side outlet muffler without baffles. The computational and experimental TL curves for all the models of SOMWB are in fair agreement. The model_1, model_2, and model_3 of SOMWB show their maximum TL values at different frequencies in the range 10 Hz–2400 Hz. The comparison among the experimental TL curves of all the models concludes that model_2 gives 16.5 dB (42 %) higher TL than model_1. Whereas, model_3 shows 9.04 dB (16.20 %) higher TL than model_2. Through the band power analysis in the ¹/₃ octave band for all the models of SOMWB it is concluded that model_1 is effective in the range 880 Hz–1360 Hz and 1800 Hz–2290 Hz, model_2 is effective in the range 1110 Hz–1260 Hz, 1310 Hz–1400 Hz, and 1690 Hz–2300 Hz. Model_3 proves to be the best design of SOMWB and is effective for 1030 Hz–1480 Hz, 1500 Hz–1570 Hz, and 1640 Hz–2400 Hz frequency sound waves. The SPL contours of model_3 provides the information about the acoustic wave strength inside the SOMWB at different frequencies. The acoustic wave strength inside the model_3 is minimum at 1520 Hz and this indicates that model_3 can effectively suppress the sound wave of this frequency. The SPL also shows the excitation of higher order modes above the cut-off frequency in the form of non-planar movement of sound waves. This study indicates that the orientations of semi-circular baffles and the outlet on the SEC muffler have a significant impact on its performance. This study is beneficial in the design of the muffler with semi-circular baffles as the internal configuration.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Author contribution

First author conceptualized the study, done the analysis, and wrote the original draft. Second author performed the evaluation and contributed to data interpretation. All authors reviewed the results and approved the final version of the manuscript.

References

1. Åbom M. (1990), Derivation of four pole parameters including higher order mode effects for expansion chamber mufflers with extended inlet and outlet, *Journal of Sound and Vibration*, **137**(3): 403–418, https://doi.org/10.1016/0022-460X(90)90807-C.

2. ANSYS (2022), Mechanical APDL: 2022R1, Acoustics – Acoustic Fundamental, Chapter 8, Ansys workbench help documentation.

3. Chang Y.-C., Chiu M.-C. (2014), Optimization of rectangular multi-chamber plenums equipped with multiple extended tubes using the BEM, neural networks, and the genetic algorithm, *Journal of Mechanics*, **30**(6): 571–584, https://doi.org/10.1017/jmech.2014.66.

4. Cheng C.Y.R., Wu T.W. (1999), Exhaust muffler design and analysis using a boundary element method based computer program, SAE Technical Paper 1999-01-1661, https://doi.org/10.4271/1999-01-1661.

5. Das S. *et al.* (2022), A novel design for muffler chambers by incorporating baffle plate, *Applied Acoustics*, **197**: 108888, https://doi.org/10.1016/j.apacoust.2022.108888.

6. Davis D.D., Stokes G.M., Moore D., Stevens G.L. (1954), Theoretical and experimental investigation of mufflers with comments on engine-exhaust muffler design, *National Advisory Committee for Aeronautics*, Report 1192, https://ntrs.nasa.gov/api/citations/19930092208/downloads/19930092208.pdf.

7. Done V. *et al.* (2014), Muffler design for a refrigerator compressor, *International Compressor Engineering Conference*.

8. Elsayed A., Bastien C., Jones S., Christensen J., Medina H., Kassem H. (2017), Investigation of baffle configuration effect on the performance of exhaust mufflers, *Case Studies in Thermal Engineering*, **10**: 86–94, http://doi.org/10.1016/j.csite.2017.03.006.

9. Fairbrother R., Varhos E. (2007), Acoustic simulation of an automotive muffler with perforated baffles and pipes, SAE Technical Paper 2007-01-2206, https://doi.org/10.4271/2007-01-2206.

10. Fan W., Guo L.X. (2016), An investigation of acoustic attenuation performance of silencers with mean flow based on three-dimensional numerical simulation, *Shock and Vibration*, **2016**: 6797593, https://doi.org/10.1155/2016/6797593.

11. Gorazd Ł. (2021), Experimental determination of a reflective muffler scattering matrix for single-mode excitation, *Archives of Acoustics*, **46**(4): 667–675, https://doi.org/10.24425/aoa.2021.139643.

12. Gupta A.K. (2016a), Comparison of noise attenuation level by a convergent and divergent cylindrical duct with space constraints, *International Journal of Scientific Research in Science, Engineering and Technology*, **2**(2): 778–781.

13. GUPTA A.K. (2016b), Observation for transmission loss by applying multiple baffle plates on single expansion chamber: A simulation approach, *International Journal of Engineering Research and Modern Education*, **1**(1): 2455–4200, http://ijerme.rdmodernre search.com/wp-content/uploads/2016/04/28.pdf.

14. GUPTA A.K., TIWARI A. (2015), Measurement of sound transmission loss on straight and zigzag perforated concentric tube muffler with constant porosity, *International Journal on Emerging Technologies*, **6**(2): 35–40.

15. HOROUB M. (2011), Acoustic noise control using multiple expansion chambers, Ph.D. Theisis, King Fahd University of Petroleum and Minerals, Saudi Arabia.

16. HUA X., HERRIN D.W. (2013), Practical considerations when using the two-load method to determine the transmission loss of mufflers and silencers, *SAE International Journal of Passenger Cars – Mechanical Systems*, **6**(2): 1094–1101, http://doi.org/10.4271/2013-01-1881.

17. IGARASHI J. (1958), *Fundamentals of Acoustical Silencers (I): Theory and Experiment of Low Pass Filters*, Aeronautical Research Institute, University of Tokyo. Aeronautical Research Institute.

18. IH J.G., LEE B.H. (1985), Analysis of higher order mode effects in circular expansion chamber with mean flow, *The Journal of the Acoustical Society of America*, **77**(4): 1377–1388, https://doi.org/10.1121/1.392029.

19. IH J.-G., LEE B.-H. (1987), Theoretical prediction of the transmission loss for the circular reversing chamber mufflers, *Journal of Sound and Vibration*, **112**(2): 261–272, https://doi.org/10.1016/S0022-460X(87)80194-3.

20. IH J.-G. (1992), The reactive attenuation of rectangular plenum chamber, *Journal of Sound and Vibration*, **157**(1): 93–122, https://doi.org/10.1016/0022-460X(92)90569-J.

21. JAYARAMAN K., YAM K. (1981), Decoupling approach to modeling perforated tube muffler components, *The Journal of the Acoustical Society of America,* **69**(2): 390–396, https://doi.org/10.1121/1.385465.

22. JEONG U.-C., KIM J.-S., KIM Y.-D., OH J.-E. (2015), Noise reduction of the automobile multi-mode muffler using differential gap control and neural network control, [in:] *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, **230**(7): 928–941, https://doi.org/10.1177/0954407015597080.

23. KULKARNI M.V., INGLE R.B. (2018), Finite element analysis of double expansion chamber reactive muffler with side outlet, *International Journal of Research and Analytical Reviews,* **5**(4): 758–764.

24. LE ROY T.W. (2011), Muffler characterization with the implementation of the finite element method and experimental techniques, MSc. Thesis, Michigan Technological University, https://doi.org/10.37099/mtu.dc.etds/381.

25. LEE J.K., OH K.S., LEE J.W. (2019), Methods for evaluating in-duct noise attenuation performance in a muffler design problem, *Journal of Sound and Vibration*, **464**: 114982, https://doi.org/10.1016/j.jsv.2019.114982.

26. MIMANI A., MUNJAL M.L. (2011), Transverse plane wave analysis of short elliptical chamber mufflers: An analytical approach, *Journal of sound and vibration*, **330**(7): 1472–1489, https://doi.org/10.1016/j.jsv.2010.09.035.

27. MIMANI A., MUNJAL M.L. (2012), Acoustical behavior of single inlet and multiple outlet elliptical cylindrical chamber muffler, *Noise Control Engineering Journal*, **60**(5): 605–626, https://doi.org/10.3397/1.3701036.

28. MUNDHE V., DEORE E. (2015), Design and optimization of perforated muffler in an automobile exhaust system, *International Journal of Applied Research*, **1**(8): 390–395.

29. MUNJAL M.L. (1975), Velocity ratio-cum-transfer matrix method for the evaluation of a muffler with mean flow, *Journal of Sound and Vibration*, **39**(1): 105–119, https://doi.org/10.1016/S0022-460X(75)80211-2.

30. MUNJAL M.L. (1987), A simple numerical method for three-dimensional analysis of simple expansion chamber mufflers of rectangular as well as circular cross-section with a stationary medium, *Journal of Sound and Vibration*, **116**(1): 71–88, https://doi.org/10.1016/S0022-460X(87)81321-4.

31. MUNJAL M.L. (1997), Plane wave analysis of side inlet/outlet chamber mufflers with mean flow, *Applied Acoustics*, **52**(2): 165–175, https://doi.org/10.1016/S0003-682X(96)00053-9.

32. MOHAMAD B., KAROLY J., ZELENTSOV A., AMROUNE S. (2021), Investigation of perforated tube configuration effect on the performance of exhaust mufflers with mean flow based on three-dimensional analysis, *Archives of Acoustics*, **46**(3): 561–566, https://doi.org/10.24425/aoa.2021.138148.

33. NARAYANA T.S.S., MUNJAL M.L. (2005), Prediction and measurement of the four-pole parameters of a muffler including higher order mode effects, *Noise Control Engineering Journal*, **53**(6): 240–246, https://doi.org/10.3397/1.2839259.

34. REDDY K.P.K., FATIMA S., MOHANTY A.R. (2017), A new sound quality metric for the design of engine exhaust mufflers, *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, **232**(2): 254–263, https://doi.org/10.1177/0954407017696608.

35. SEYBERT A.F., CHENG C.Y.R. (1987), Application of the boundary element method to acoustic cavity response and muffler analysis, *Journal of Vibration, Acoustics, Stress, and Reliability in Design*, **109**(1): 15–21, https://doi.org/10.1115/1.3269388.

36. SIANO D., AURIEMMA F., BOZZA F. (2010), Pros and cons of using different numerical techniques for transmission loss evaluation of a small engine muffler, SAE Technical Paper 2010-32-0028, https://doi.org/10.4271/2010-32-0028.

37. SULLIVAN J.W. (1979), A method of modeling perforated tube muffler components. I. Theory, *The Journal of the Acoustical Society of America*, **66**(3): 772–778, https://doi.org/10.1121/1.383679.

38. SULLIVAN J.W., CROCKER M.J. (1978), Analysis of concentric tube resonators having unpartitioned cavities, *The Journal of the Acoustical Society of America*, **64**(1): 207–215, https://doi.org/10.1121/1.381963.

39. TAO Z., SEYBERT A.F. (2003), A review of current techniques for measuring muffler transmission loss, SAE Technical Paper 2003-01-1653, https://doi.org/10.4271/2003-01-1653.

40. VISHWAKARMA S.K., PAWAR S.J. (2021), Simulation studies on the transition from simple expansion chamber muffler to tapered expansion chamber muffler, [in:] *Advances in Fluid and Thermal Engineering. Lecture Notes in Mechanical Engineering*, pp. 389–398, https://doi.org/10.1007/978-981-16-0159-0_34.

41. VISHWAKARMA S.K., PAWAR S.J. (2022), Analytical and computational acoustic modeling of side outlet muffler and its extension in the modeling of tapered side outlet muffler, *Archives of Acoustics*, **47**(4): 491–499, https://doi.org/10.24425/aoa.2022.142893.

42. VISHWAKARMA S.K., PAWAR S.J. (2024), Analytical, computational, and experimental investigations on the impact of side outlet on the simple expansion chamber with axial outlet, *Noise & Vibration Worldwide*, **55**(9–10): 473–480, https://doi.org/10.1177/09574565241278721.

43. WANG C.-N. (1999), A numerical analysis for perforated muffler components with mean flow, *Journal of Vibration and Acoustics*, **121**(2): 231–236, https://doi.org/10.1115/1.2893969.

44. XIANG L., ZUO S., WU X., ZHANG J., LIU J. (2016), Acoustic behaviour analysis and optimal design of a multi-chamber reactive muffler, *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, **230**(13): 1862–1870, https://doi.org/10.1177/0954407016630112.

45. YI S.-I., LEE B.-H. (1986), Three-dimensional acoustic analysis of circular expansion chambers with a side inlet and a side outlet, *The Journal of the Acoustical Society of America*, **79**(5): 1299–1306, https://doi.org/10.1121/1.393709.

46. YU X., TONG Y., PAN J., SUN H., CHENG L. (2016), On the retrofitted design of a truck muffler with cascaded sub-chambers, *Noise Control Engineering Journal*, **64**(5): 602–607, https://doi.org/10.3397/1/376404.

47. ZHAO B., LI H. (2022), Analysis of the influencing factors of the acoustic performance of the muffler considering acoustic-structural coupling, *Archives of Acoustics*, **47**(4): 479–490, https://doi.org/10.24425/aoa.2022.142900.

48. ZHENG S., KANG Z., LIAN X. (2012), Acoustic simulation for exhaust silencer of vehicle engine with combined one/three-dimensional approach, *International Journal of Vehicle Design*, **60**(1/2): 57–70, http://doi.org/10.1504/IJVD.2012.049164.

Research Paper

The Impact of Training Strategies on Overfitting in Vowel Classification Using PS-HFCC Parametrization for Automatic Speech Recognition

Stanislaw GMYREK*, Urszula LIBAL, Robert HOSSA

Faculty of Electronics, Photonics and Microsystem, Department of Acoustics, Multimedia and Signal Processing
Wroclaw University of Science and Technology
Wrocław, Poland

*Corresponding Author e-mail: stanislaw.gmyrek@pwr.edu.pl

This paper investigates the overfitting problem in vowel classification task for automatic speech recognition (ASR). It utilizes a pitch synchronized human factor cepstral coefficients (PS-HFCC) as the parametrization method, which outperforms traditional methods like HFCC and mel-frequency cepstral coefficients (MFCC) in frame-level classification accuracy. While deep learning models are prevalent in contemporary ASR systems, they often lack explainability, a characteristic of classical classifiers. Therefore, this study examines overfitting phenomenon using a range of classifiers with well-understood properties. Specifically, it analyzes the impact of different training strategies on classifier performance, comparing the susceptibility to overfitting of several widely used classifiers, including the Gaussian mixture model (GMM), a standard approach in speech recognition. The analysis of training strategies considers various data splitting methods: random, speaker-based, and cluster-based. Our analysis of training strategies highlights the crucial role of data splitting methods: while random splitting is commonly used, it can lead to inflated accuracy due to overfitting. We demonstrate that speaker-independent splitting, where the classifier is trained on one set of speakers and tested on a separate, unseen set, is essential for robust evaluation and for accurately assessing generalization to new speakers. Potentially, the resulting insights may inform the future development and training of more reliable ASR systems.

Keywords: automatic speech recognition; vowel classification; classifier training strategy; pitch synchronized human factor cepstral coefficient; overfitting; robust parametrization; speaker grouping.

1. Introduction

The objective of speech recognition is to leverage machines, computers, and appropriate software to process speech signal and extract useful information for humans. This information can include the semantic content of speech, and the considered systems are referred to as automatic speech recognition (ASR) (KUNDEGORSKI *et al.*, 2014; UMA MAHESWARI *et al.*, 2020; CHERIFI, GUERTI, 2021), automatic voice recognition (AVR) systems for voice or speaker recognition (MACIEJKO, 2015), and automatic emotion recognition systems (AER) for emotional state recognition (NEDELJKOVIĆ *et al.*, 2020; PIĄTEK, KŁACZYŃSKI, 2021; STEFANOWSKA, ZIELIŃSKI, 2024). Speech recognition has been a highly researched topic in recent years and continues to develop intensively. It is inherently interdisciplinary, encompassing a multitude of fields, including acoustics, digital signal processing, mathematical statistics, machine learning, artificial intelligence, linguistics, semantics, and psychology (particularly the study of emotions). In order to optimize the efficacy of an ASR system, it is essential to consider a variety of factors influencing speech, as well as the operational conditions under which it is processed, during the system design phase. It is therefore necessary to distinguish between three main categories of ASR systems: speaker dependent (SD), designed for a single speaker, speaker independent (SI) dedicated to working with multiple speakers, and speaker adaptive (SA), in which parameters can be adjusted to fit the active speaker. In order to optimize the ASR per-

formance, it is also necessary to take into account other specific operational conditions. These include the need to work with continuous speech, single and isolated commands, and conversions to text discussions involving multiple speakers (MAKOWSKI, 2011).

Feature extraction is fundamental component of both traditional and modern ASR system architectures. Classical approaches, such as hidden Markov models with Gaussian mixture models (HMM-GMM) and hybrid HMM-deep neural network (HMM-DNN) rely on signal parameterization based on well-established methods such as spectral, cepstral, or time-frequency transformation methods. While end-to-end (E2E) architectures integrate, or partially integrate, feature extraction within DNNs, they still often utilize internal representations, such as acoustic feature vectors (in encoders) or spectrograms and mel-spectrograms (in convolutional neural networks – CNNs). Signal parameterization, therefore, remains a crucial step impacting the accuracy and efficiency of ASR systems. The ongoing search for robust parameterization methods is warranted to mitigate the negative influence of various factors related to the variability of acoustic speech features, which can detrimentally affect ASR performance. Additionally, improved signal representation can reduce the complexity requirements for recurrent neural networks (RNNs), CNNs in deep models, and E2E systems, addressing a key challenge in ASR system design: the reduction of computational complexity.

The Polish language contains six basic vowel sounds, which can be classified as either oral or nasal vowels. In this study, we focus only on the classification of six oral vowels in Polish speech: A /a/, E /ɛ/, I /i/, O /ɔ/, U or Ó /u/, Y /ɨ/, without the two nasal vowels: Ą /ɔ/ and Ę /ɛ/. The main cause of overfitting in vowel classification is the use of a small, homogeneous training dataset. When training data includes recordings from only a few speakers or environments, the model might overfit to these specific conditions and fail to generalize across different voices or settings. The other reasons for overfitting are overly complex models (e.g., DNN and E2E with many layers). In vowel classification, this can lead to overfitting, as the model may capture intricate details of the training data that do not generalize well.

2. Theoretical background of speech production

The Fant source-filter model assumes that the speech signal $s(n)$ can be described by the following relationship:

$$s(n) = u(n) \star v(n) \star r(n), \qquad (1)$$

where $\star$ is the convolution operator, n is the time index, and the component $u(n)$ denotes the excitation signal, $v(n)$ is the vocal tract, and $r(n)$ describes

emission of the signal through the speaker's mouth (RABINER, SCHAFER, 2010). For voiced speech, the excitation signal assumes a periodic form, a noisy character in voiceless speech, or a mixed model to describe plosive phonemes (QUATIERI, 2001). When the air from the lungs vibrates the vocal cords, the excitation takes the form:

$$u(n) = g(n) \star p(n) = \sum_{k=0}^{+\infty} g(nT_s - kT_0), \qquad (2)$$

where $g(n)$ is the shape of a single excitation pulse, $p(n)$ is a pulse train with a repetition period equal to the fundamental period T_0, which is related to the periodic opening and closing of the vocal cords, and T_s is the sampling interval.

3. Feature extraction

In general, speech is characterized by both high variability and randomness, therefore, its time signature is not an adequate representation of it. One of the key elements in the signal processing scheme of ASR systems is therefore the preprocessing and feature extraction stage. The main goal of the parameterization of the speech signal here is to represent the signal using a possibly small set of parameters that effectively extract its distinctive features relevant for further processing and analysis. The literature in this area is very extensive. In general, speech parameterization methods can be divided into two categories: solutions based on signal filtering, i.e., using linear predictive coding (LPC) analysis, and methods based on time-frequency transformations, usually short-time Fourier transform (STFT), and cepstral analysis (psychoacoustic model) of the signal. The latter are considered classic solutions.

The cepstral parameterization process results in a vector of cepstral coefficients, expressed as

$$c(t,m) = \sum_{j=1}^{J} Y_l(t,j) \cos\left(m\left(j - \frac{1}{2}\right)\frac{\pi}{J}\right); \quad m = 1, ..., M, \qquad (3)$$

where t is the index of the signal frame, m is the index of the cepstral coefficient, M is the number of coefficients, j is the index of the mel scale bin, J is the number of mel bands, and Y_l is the logarithm of the amplitude spectrum in the mel-frequency scale obtained at the output of a bank of perceptual filters.

Various cepstral parameterization solutions differ mainly in the way the perceptual filter bank coefficients are determined. A commonly used feature extraction method in speech recognition are mel-frequency cepstral coefficients (MFCC), introduced by DAVIS and MERMELSTEIN (1980). MFCCs are popular preprocessing method, not only in speech recognition (UPADHYAYA *et al.*, 2015), but also in multiple
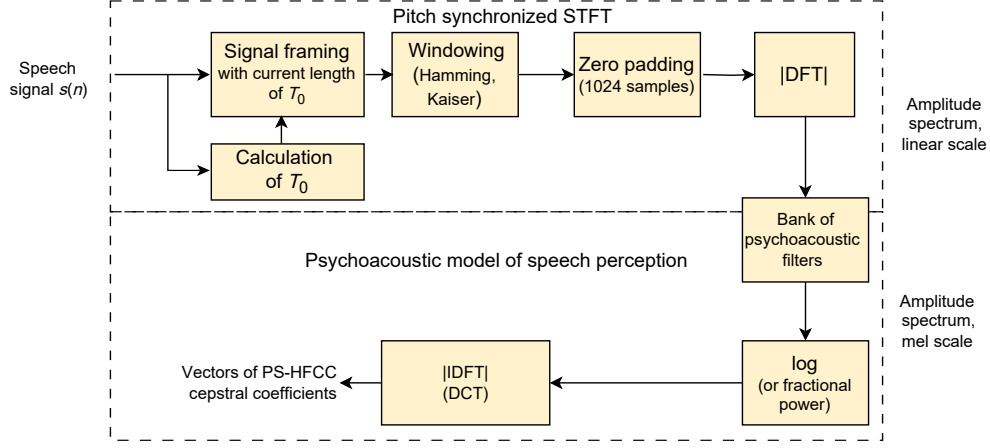
Fig. 1. PS-HFCC cepstral parameterization scheme.

other applications, such as honeybee sound analysis (Libal, Biernacki, 2024a; 2024b; 2024c). However, a more robust cepstral representation is obtained by human factor cepstral coefficient (HFCC) parameterization (Skowronski, Harris, 2004). Studies show that HFCCs perform better under noisy signal conditions and lead to improved classification results, i.e., lower errors during the single-frame recognition stage of the signal (Skowronski, Harris, 2004).

In order to make the classical HFCC parametrization method robust against the negative effects of excitation periodicity in voiced speech phonemes, the pitch synchronized human factor cepstral coefficient (PS-HFCC) parametrization is employed. This approach utilizes variable-length signal frame processing, as depricted in Fig. 1.

The result of the frequency analysis of a signal frame $s_w(n)$ containing a voiced fragment of speech can be described by the following relation (Gmyrek *et al.*, 2023):

$$S_w(\omega) = (S(\omega) \cdot G(\omega, T_0)) \star W(\omega), \qquad (4)$$

where $S_w(\omega)$ is the frame spectrum of the speech signal, $S(\omega)$ is the desired form of the spectrum with clearly visible formants, and $W(\omega)$ is responsible for the impact of the windowing operation needed to extract individual frames from the recorded signal. The negative impact of the fundamental frequency f_0 on HFCC coefficients was studied in detail in (Gmyrek *et al.*, 2023; 2024).

The PS-HFCC method makes it possible to compensate for the undesired effect of $G(\omega, T_0)$, which results in occurrence of amplitude spectrum ripples at multiples of the fundamental frequency f_0. In this case, the modification of the method consists of estimating the current value of the fundamental frequency f_0 and synchronizing the signal frame length with the fundamental period T_0. Details of this solution are described in (Gmyrek, Hossa, 2025a; 2025b). By applying the PS-HFCC method, the values of the variance

estimators of the cepstral coefficients decrease and, consequently, a higher concentration of areas with data representing the cepstral parameters of the elementary frames is observed. At the same time, this results in narrower multivariate probability density distributions of the data, which in turn translates into better classification results, i.e., a decrease in recognition errors at the level of individual signal frames levels.

The preprocessing of the Polish vowel recordings using PS-HFCC parametrization was performed using custom-prepared MATLAB scripts.

4. Database

The authors developed a proprietary speech corpus, comprising recordings from 37 adult male speakers, collected from various regions of Poland. For each speaker, 150 Polish words were recorded, with speech fragments containing vowels (six classes) from 43 words subsequently employed in the experiment. The sampling rate of the signals was 12 kHz. The original database was characterized by a low noise level with a signal-to-noise ratio (SNR) of 35 dB. The experiments, presented in this work, were conducted on both the original dataset and its noisy versions, with SNRs of 20 dB, 10 dB, and 5 dB (representing progressively higher noise level). This database is highly representative – it captures both inter-speaker and intra-speaker variability, as well as contextual and phonetic diversity. The preparation of this dataset was exceptionally labor-intensive and costly, involving semi-automatic signal segmentation and detailed phonetic labeling. All recordings were manually segmented and labeled, with six phonemes ('a', 'e', 'i', 'o', 'u', and 'y') chosen as the phonetic units for labeling process. The frame length was set to 30 ms, with a 10 ms shift. The YIN estimator (de Cheveigné, Kawahara, 2002) and its statistically improved version PYIN (Mauch, Dixon, 2014) were employed to estimate the current value of the fundamental period T_0. After signal preprocessing, we

obtained $N = 14$ cepstral coefficients for each frame. In our database, the numbers of frames for individual phonemes were: 12 208 for 'i', 9288 for 'y', 29 778 for 'e', 35 406 for 'a', 23 628 for 'o', and 8082 for 'u'. The total number of frames that contain vowel sounds was 118 390.

The authors acknowledge the limitations associated with analyzing only male speech recordings, so a comparable female speech corpus is currently under development. However, it is not expected to play a central role in the current analysis. The primary acoustic distinction between male and female voices lies in the higher fundamental frequency (f_0) typically observed in female speakers. In pitch-synchronous analysis, this leads to a greater number of pitch cycles within a single frame, which may facilitate improved signal averaging and potentially enhance recognition accuracy. Nonetheless, this factor has a limited impact on overfitting, which is primarily influenced by the chosen training and test data partitioning strategy.

5. Training and test sets

All classification methods based on the maximum likelihood lead to fit models to the training dataset. The real challenge is to construct the training and testing sets in a way that prevents overfitting. To trace the influence of the training process of the classifier, we conducted a series of tests using a database of vowel sounds from Polish speech. These tests involved dividing the data into training and test sets in various ways, with an effort to maintain the following proportions: two-thirds of data for the training set, and the remaining one-third for the test set. The training and testing sets were always kept disjoint.

For the purposes of this research, we employed three methods of splitting the data into training and testing sets: random split, speaker split, and cluster split, each of which is described further.

5.1. Method 1: random split

The first method for splitting the data into training and test sets is a random split, with exact proportions of two-thirds for the training set and one-third for the test set. Each recording frame was randomly assigned to one of these sets. The training and test sets remained disjoint. The random split was performed 10 times, and all classifiers were trained 10 times on the obtained sets to average the results.

5.2. Method 2: speaker split

The second method relies on dividing the speakers into two groups: one for training the classifiers and the second for testing them. The database contains recordings from 37 speakers. While the division into training and test sets was not exactly two-thirds to one-third, but quite close with 24 speakers in the training set and 13 speakers in the test set. The random choice of speakers for the two sets was also repeated 10 times, leading to 10 separate experiments to average the results. In this split, the training set was disjoint from the test set, and additionally, no recordings from the same speaker appeared in both sets at the same time.

In contrast to random split, the speaker split method for the training of the classifiers should prevent overfitting to some degree and provide more realistic (lower) performance results.

5.3. Method 3: cluster split

The third method involves K-means clustering. Three clusters were separated from the data by the K-means algorithm with $K = 3$. Each recording frame was allocated to one of the three clusters. The clusters produced by the K-means algorithm are groups of data points that share similar features and are spatially close to one another in the feature space. In K-means clustering, each cluster is defined by a centroid, which represents the center of the cluster and is calculated as the average of all data points within that cluster. The algorithm iteratively adjusts the positions of the centroids to minimize the sum of squared distances between the data points and their respective centroids, ensuring data points are grouped in a way that reduces intra-cluster variance.

To maintain the two-thirds to one-third ratio, in each experiment one cluster served as the test set, while the remaining two clusters were used as the training set. The clusters are formed by grouping similar feature vectors based on Euclidean distance. We expect that this data-splitting method will produce the worst performance for the classifiers.

6. Vowel classification

6.1. Classifiers

For the classification of vowels and the study of overfitting, we performed a series of classifications using the following classifiers: Gaussian mixture model (GMM), K-nearest neighbors (KNN), random forest (RF), support vector machines (SVM), and multi-layer perceptron (MLP). The classification experiments were performed in the Python programming language with the scikit-learn library.

In the context of speech recognition, the GMM (Reynolds, 2009; McLachlan, Peel, 2000) is a popular statistical approach used to model the distribution of acoustic features in speech. A GMM is a mixture of M multivariate, normal distributions, which together describe the distribution of input data, such as acoustic feature vectors (e.g., MFCC, HFCC or PS-HFCC) extracted from speech signals. GMM is particularly

useful in acoustic classification, as it allows for modeling the variability in speech across different speakers, acoustic conditions, and over time. The diagonal covariance matrices Σ_{ic} were determined based on the expectation-maximization (EM) algorithm:

$$p_c(o) = \sum_{i=1}^{M} w_{ic} \mathcal{N}(o, m_{ic}, \Sigma_{ic}), \qquad (5)$$

where w_{ic} denotes the weights and m_{ic} denotes the means for the mixture of i-th component and c-th phoneme (class). The EM algorithm was described in detail in (DEMPSTER *et al.*, 1977).

The KNN (COVER, HART, 1967; BISHOP, 2006) is a simple, instance-based algorithm that classifies data points based on the majority class of their nearest neighbors. It uses distance metrics (such like Euclidean distance) to find the closest K-neighbors and assigns the most common class among them. To investigate the overfitting, tests were conducted for $K = 1, 21$, and 51 neighbors. KNN can overfit when K is too small (e.g., $K = 1$), as it becomes sensitive to noise and outliers in the training data. This leads to a model that performs well on the training data but poorly on unseen data.

MLP, also called deep feedforward network (RU-MELHART *et al.*, 1986; GOODFELLOW *et al.*, 2016), is a type of neural network with multiple layers of neurons. It uses backpropagation to learn the weights, making it capable to learn complex patterns and nonlinear decision boundaries. MLPs are widely used for various classification tasks. MLP can overfit when the network is too deep (i.e., there are too many layers or neurons) or when training is conducted for too many epochs. Overfitting occurs when the network becomes too specialized to the training data, capturing noise and irrelevant patterns, leading to poor performance on new data. This is particularly likely if regularization techniques such as dropout or L2 regularization are not used. For our experiments, we utilized an MLP with one hidden layer consisting of 100 neurons and reclective linear unit (ReLu) activation function. The MLP was trained for 200 epochs.

RF (BREIMAN, 2001) is an ensemble learning method that builds multiple decision trees and combines their predictions to improve accuracy. Each tree is trained on a random subset of the data, and the final prediction is based on a majority vote from all trees. This approach reduces variance and improves accuracy compared to a single decision tree. Although RFs are less prone to overfitting compared to individual decision trees, they can still overfit if the trees are grown too deep or if the number of trees is too large. Overfitting can occur if the model becomes overly complex and captures noise in the data.

SVM (BISHOP, 2006; CORTES, VAPNIK, 1995) is a powerful classification algorithm that finds the hyperplane that best separates the data into different classes. It maximizes the margin between the closest data points (support vectors) from different classes. For non-linear data, SVM can use kernel functions (e.g., radial basis functions (RBFs) or polynomial) to map the data into higher-dimensional spaces where it can be linearly separated. In this study, we used RBFs as kernels. SVM can overfit when a very complex kernel (e.g., a high-degree polynomial) is used, or when the regularization parameter is set too high, causing the model to fit the training data too closely, including noise, at the expense of generalization.

6.2. Classification error analysis

The classification results were analyzed by using three error measures: accuracy, frame error rate (FER), and the confusion matrix.

Accuracy is the most general method for comparing different classification methods. It is defined as the fraction of correct predictions N_{correct} out of all predictions N:

$$\text{Acc} = \frac{N_{\text{correct}}}{N} \cdot 100\,\%. \qquad (6)$$

The higher the accuracy, the better the classification quality. On the other hand, the accuracy measure does not distinguish between the accuracy across individual classes, which can sometimes be crucial when analyzing classifiers. Other measures, such as the FER and the confusion matrix, are employed to address this limitation.

A confusion matrix is a performance evaluation tool commonly used in machine learning to evaluate the accuracy of classification models. It provides a summary of prediction results for a classification problem by comparing predicted labels with the actual labels for each class. Due to varying number of frames for each class, we present confusion matrices containing percentage results instead of numbers.

FER measure is traditionally used to assess the quality of speech recognition at the individual frame level and is defined for a class c as

$$\text{FER}(c) = \frac{N_{\text{err}}(c)}{N(c)} \cdot 100\,\%, \qquad (7)$$

where $N(c)$ is the total number of frames undergoing recognition and $N_{\text{err}}(c)$ is the number of unrecognized frames from class c. FER can also be calculated directly from the confusion matrix by taking the ratio of the sum of all values in a row of the confusion matrix, excluding the diagonal value, to the sum of all values in that row.

6.3. Classifier overfitting

The aim of this paper is to investigate the overfitting of vowel classification for Polish speech. Classifier overfitting (HASTIE *et al.*, 2001; KUHN, JOHNSON, 2013; NG, 2004) occurs when a model learns to

perform exceedingly well on the training data to the extent that it begins to memorize irrelevant details, noise, or peculiarities that are specific to that dataset. As a result, while the model achieves high accuracy on the training set, its performance deteriorates on unseen test data. In essence, the model fails to generalize well, being overly tailored to the particular examples it has encountered, rather than capturing the underlying patterns that could apply to new, unseen data.

7. Results

7.1. Comparison of accuracy with split methods

In this section, we analyze the performances of various classifiers using different train-test split methods: random, speaker and cluster. In this experiment, we used seven classifiers with different overfitting tendencies: GMM, KNN with $K = 1$, 21, and 51, a fully connected neural network of 100 perceptrons (MLP), RF with trees of 15 branches in depth, and SVM with RBF kernels. Results for our custom dataset with a signal-to-noise ratio (SNR) of 35 dB (indicating a low level of noise), as well as artificially noised versions of the dataset (with SNRs of 20 dB, 10 dB, and 5 dB, representing very high levels of noise) are presented in Fig. 2 and Tables 1, 2, and 3. The most important factor in overfitting analysis is the comparison of the first two methods, i.e., random split versus speaker split. The greater the accuracy between these two splits (assuming the random split yields higher accuracy compared to the speaker split), the stronger the overfitting effect. At higher noise levels (i.e., lower SNRs), all classifiers exhibit a small drop in accuracy, at most by just a few percentage points. Nonetheless, the PS-HFCC parametrization generally demonstrates robustness against significant noise in the recordings. As de-

Table 1. Accuracy [%] for random split (method 1).

Classifier	35 dB	20 dB	10 dB	5 dB
GMM	85.92	85.89	85.56	84.20
1NN	99.52	99.05	98.89	98.02
21NN	94.36	92.70	92.19	90.75
51NN	91.82	90.64	90.32	89.12
MLP	92.76	92.00	90.52	89.16
RF	92.59	90.94	90.10	88.71
SVM	94.03	92.93	91.90	90.53

Table 2. Accuracy [%] for speaker split (method 2).

Classifier	35 dB	20 dB	10 dB	5 dB
GMM	81.11	80.73	81.42	80.04
1NN	77.52	76.97	76.39	75.15
21NN	81.19	81.54	81.25	80.62
51NN	82.47	82.65	82.38	81.58
MLP	79.84	79.54	79.63	79.58
RF	82.84	82.97	82.84	82.12
SVM	82.72	83.24	83.02	82.44

Table 3. Accuracy [%] for cluster split (method 3).

Classifier	35 dB	20 dB	10 dB	5 dB
GMM	64.05	71.28	65.27	64.73
1NN	60.22	58.20	57.62	54.90
21NN	57.81	55.98	53.66	52.52
51NN	54.75	54.51	50.35	50.37
MLP	77.01	76.76	74.41	76.55
RF	56.50	56.41	56.04	57.53
SVM	81.44	78.05	76.77	73.82

picted in Fig. 2, classification results across various SNRs (original 35 dB, 20 dB, 10 dB, and 5 dB) remain consistent regardless of the train-test split method. This consistency underscores the noise resilience of PS-HFCC parametrization.
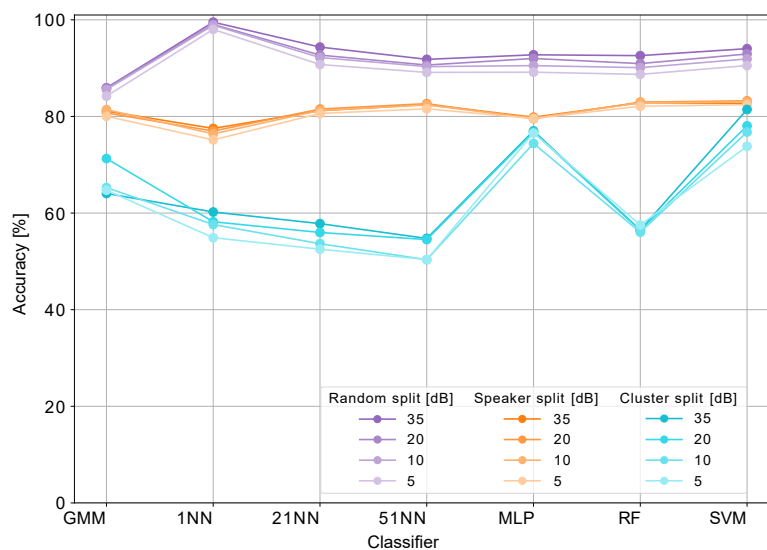

Fig. 2. Classification accuracy for different train-test splits and SNRs.

As expected, the KNN classifier with one neighbor ($K = 1$) exhibited overfitting, serving as a 'litmus test' for this behavior. The difference in accuracy between the random and speaker splits for 1NN was very high, at 22 %. This indicates that the 1NN algorithm fits the training data. Considering that the signal was divided into 30 ms frames with 10 ms shifts, it is highly probable that a neighboring frame, with 20 ms overlap, could be chosen as the nearest neighbor. The difference in accuracy between the random and speaker splits for 21NN dropped to 13.17 %, indicating better generalization of results due to the majority voting among the 21 neighbors. For 51NN, the difference further decreased to 9.35 %. Obviously, the more neighbors vote for the predicted class label, the better the generalization. At the same time, there are disproportions in the number of frames for individual classes (since speech naturally contains more instances of some vowels, e.g., 'a' and 'e'), which influences the final classification result.

Very similar results to 51NN were obtained for the RF classification, for which the difference in accuracy between the random and speaker splits was equal to 9.75 %. The RF model used in this experiment was an ensemble of 100 trees. These trees were allowed to grow to the maximum depth of 15. The RFs constructed by shallower trees led to relatively lower classification accuracy, especially for the speaker split. Using deeper trees can lead to much stronger overfitting effect, which is observed here, but at the same level to that of relatively well generalized 51NN classifier.

To a certain extent, comparable behavior was noticed for the MLP and SVM methods. The difference in accuracy between the random and speaker splits was 12.94 % and 11.31 %, respectively, representing average results of overfitting compared to the other classifiers. Interestingly, the performance of the two classifiers on the speaker and cluster splits was very close, with only negligible small drop in accuracy from 79.84 %

to 77.01 % for MLP, and from 82.72 % to 81.44 % for SVM – see Tables 2 and 3. This indicates that both methods handle clustered data (from K-means partitioning) significantly well, due to their nonlinear mappings. However, despite this strength, the overfitting effect remains evident in the case of the random split, for both MLP and SVM.

Among the many parameters by which speakers can be divided into groups, their personal vocal characteristics, including vocal tract parameters, are especially noteworthy. One such parameter is the fundamental frequency f_0, which can be taken as an indicator of vocal tract size, as it is closely related to the length of the speaker's vocal cords (Makowski, 2011). The lengths of the oral and pharyngeal parts of the vocal tract can be taken as a basis for grouping, as they directly affect the positioning of formants on the frequency axis (Naito *et al.*, 2002), as well as influence the parametrization coefficients. These coefficients aim to maximize the distance between the multidimensional probability distributions of the feature vectors in terms of the chosen distance measure. Partitioning can also be performed hierarchically, using multiple factor to distinguish speakers from one another. For example, clustering based on gender and speaking speed has been proposed in (Hazen, 2000). One of the more recent algorithms proposed in the literature is an approach based on adapting the weights of universal background model (UBM) proposed in (Hossa, Makowski, 2016). However, in the current study, clustering was performed numerically using the K-means method. As expected, the classification quality in this case was the lowest among all classifiers used – see Table 3.

7.2. FER analysis

Analysis of the FER for the original custom dataset with an SNR of 35 dB, as presented in the Fig. 3, leads
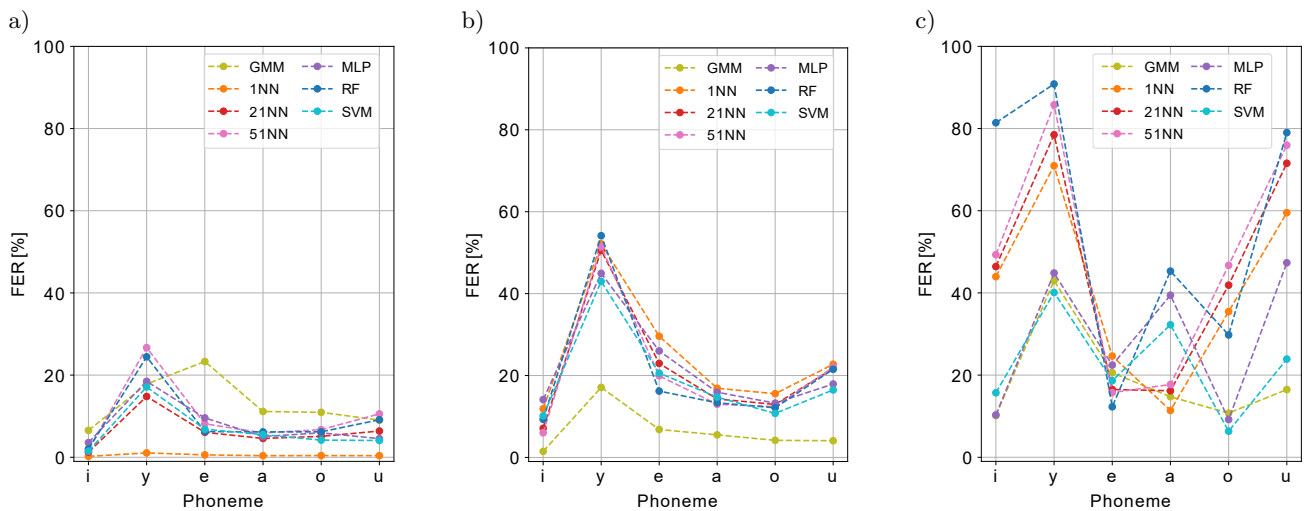

Fig. 3. FER for SNR = 35 dB across three data split approaches used during classifiers training:
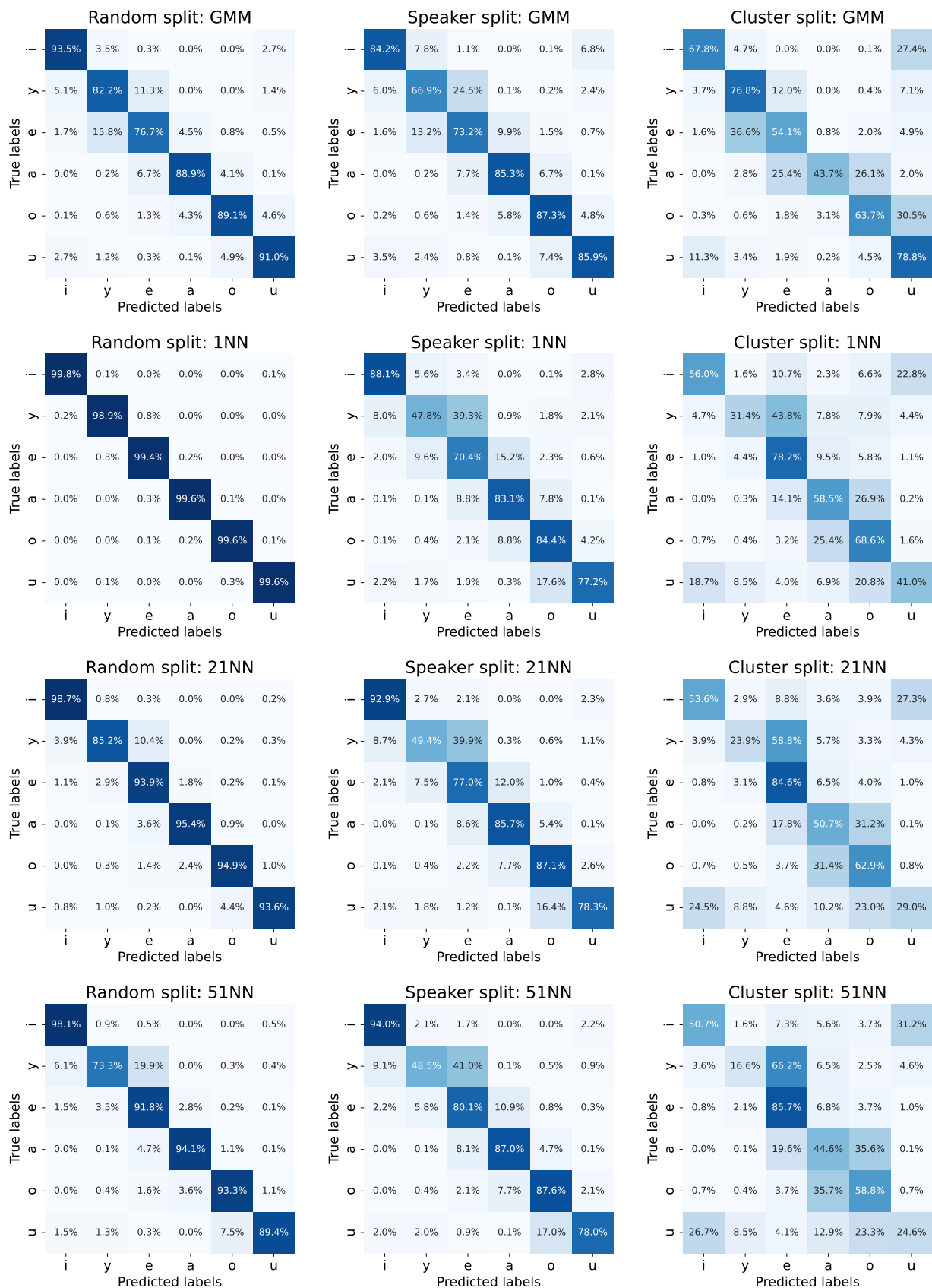a) random split; b) speaker split; c) cluster split.

Fig. 4. Confusion matrices for the classifiers: GMM and KNN with $K = 1$, 21, and 51. Classification was performed using three data split methods between training and testing sets. The results were obtained on the original custom Polish vowels dataset with an SNR of 35 dB.

to the following general conclusions. For the random split of data between training and testing sets (Fig. 3a), the 1NN classifier performed worse than all others, demonstrating a strong overfitting. For the speaker split (Fig. 3b), the only well and consistently performing classifier was GMM. Under the cluster split condition (Fig. 3c), the comparable results were obtained for GMM, MLP, and SVM methods, with a slight advantage of GMM.

The most problematic phoneme in terms of recognition accuracy was the vowel 'y', which exhibited the highest error values for all three splitting methods. The class 'y' was most frequently misclassified as 'i'. However, for the GMM with a random split, the highest FER was observed for the vowel 'e', which was again confused with 'y' sounds.

7.3. Local error analysis

We analyzed the confusion matrices (Figs. 4 and 5) for all tested classifiers to determine local errors, i.e., the confusion between true and predicted vowels. The most problematic vowel pair is 'y' and 'e'. This is due to the close proximity of their formant frequencies

Fig. 5. Confusion matrices for the classifiers: MLP, RF, and SVM. Classification was performed using three methods of data splitting between training and testing sets. The results were obtained on the original custom Polish vowels dataset with an SNR of 35 dB.

(Makowski, 2011). This can be verified by analyzing widely available formant frequency tables for Polish vowels, for example, such as those in (Jassem, 1973). Consequently, these vowels are often confused.

The confusion between vowels 'y' and 'e' is the strongest in the speaker split for both training and testing sets. All classifiers for the speaker split predict frames from the vowel class 'y' as class 'e' quite often, with the frequency (estimated probabilities) ranging from 24.5 % for GMM to as high as 45.7 % for RF – see Figs. 4 and 5. The worst performance was obtained with RF for all three methods of data split.

The confusion between 'o' and 'u' in the speaker split is highest for the KNN and RF algorithms (around 17 %). The best result, with the lowest number of misclassified phonemes 'o' and 'u' in the speaker split, was achieved again by GMM (7.4 %). However, for the cluster split, GMM performed worst with 30.5 % of 'o' phonemes misclassified as 'u'. This is due to the characteristic of the GMM training process, where a separate model is built for each class and some phonemes are not properly represented in the training clusters after K-means clustering of the data.

The confusion between 'i' and 'u' and 'a' and 'o' appears most prominently in the cluster split. For the first phoneme pair ('i' and 'u'), the errors rate rage from 14.0 % for MLP to 77.5 % for RF. Moreover, under the speaker split all classifiers except GMM, achieved error rates of around 3 % or lower, while only GMM exceeded twice that level. Phoneme 'u' naturally occurs quite rarely in Polish speech compared to other phonemes. Confusion between 'a' and 'o' is mainly visible for the cluster split with quite high errors from around 25 % to 35 %, except MLP and SVM classifiers, which performed very well in that case. We guess well trained nonlinearity of decision function is responsible for those correct classifications.

The confusion between 'e' and 'a' is observed for all classifiers under the seakers split and error rate ranges from 9.5 % for RF to 15.2 % for 1NN. For the random split, these two phonemes were generally confused in 4 %–5 % of cases, with two exceptions for the GMM and 1NN algorithms. For GMM, we observe error rates of 4.5 % and 6.7 %, while for the 1NN – extremely low error rates of 0.2 % and 0.3 %. These low error rates for 1NN indicate, a strong overfitting tendency (only one neighbor decides on predicted class). In contrast, GMM demonstrates to well-generalizable performance.

8. Conclusions

Overfitting in classifiers remains a challenging phenomenon to quantify in a rigorous scientific manner, especially in real-world applications. However, it can have a detrimental effect, causing models to make inaccurate predictions, even when the tested data suggest otherwise.

The experiments conducted in this study involved a comparison of the learning performance of the following seven classifiers: GMM, 1NN, 21NN, 51NN, MLP, RF, and SVM. These classifiers were trained on three data setups, each applyinh different training strategies for splitting the data into training and testing sets:

- random split using all frames;
- speaker split, where speakers were grouped to avoid repetition;
- cluster split, based on the most distant feature vectors, selected according to a chosen metric.

Using an ASR system, outside the conditions it was trained on, can lead to a significant drop in performance. Speech signals are highly variable, influenced by factors such as speaker characteristics (e.g., gender, age, vocal tract anatomy), intra-speaker variability, linguistic diversity, as wells as regional, cultural, and contextual factors. Therefore, ASR systems must account for these variations to maintain accuracy across different environments and user populations.

The PS-STFT, a generalization of classical cepstral parameterization methods such as MFCC and HFCC, as well as other spectrogram-based approaches, enhances recognition performance in ASR systems. It achieves this by smoothing the amplitude spectra (and, consequently, spectrograms), and by reducing the variance of cepstral coefficient estimators. Our classification results across various SNRs (original 35 dB, 20 dB, 10 dB, and 5 dB) consistently demonstrate the high robustness of the PS-HFCC parametrization to noise in recordings, regardless of the train-test split method used.

The aim of the study was to compare different popular classifiers with – the default algorithm commonly used in speech recognition – the GMM algorithm. GMM is the most robust against overfitting among tested classifiers and well generalizes the data, even in the case of a random split between training and testing sets in the classifier learning process. This is a scenario in which we expected the highest overfitting effect. Considering that the signals were divided into 30 ms frames with 10 ms shifts, it is highly probable that neighboring frames with 20 ms overlap were chosen as nearest neighbors. For this reason, the KNN algorithm served as a reference point for overfit detection.

While random split is the default method used in most studies on classification tasks, speech recognition is a specialized task that often requires alternative splitting strategies, such as speaker grouping by gender or other individual characteristics. In scenarios like ours, where the classifier is trained on one set of speakers and tested on a different set (i.e., a speaker split), this approach is essential. Training an algorithm for all potential speakers is impossible due to the vast voice diversity, as encountered in real-world applications such as smartphone-based speech recognition.

For this reason, popular big-data systems incorporated in everyday software utilize users voice samples to retrain the systems and improve their performance. However, collecting additional samples from new speakers is not always feasible. Therefore, it is important to use methods that are as resistant to overfitting as possible.

Beyond simply separating speakers for training and testing, advanced speaker grouping techniques (e.g., based on gender, dialect, or vocal characteristics) can further enhance the robustness of ASR systems. While our current study utilizes only male recordings, the observed benefits of speaker-independent train-test splitting and the robust performance of PS-HFCC parametrization are expected to extend to other demographic groups. This underscores at broader implication: speaker-independent evaluation, potentially incorporating detailed speaker grouping, is crucial for developing ASR models that truly generalize and avoid overfitting to specific speaker characteristics present in the training data, thereby ensuring reliable performance with unseen users across diverse populations.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Authors' contribution

Stanisław Gmyrek was responsible for developing the MATLAB software, conceptualizing and implementing the signal processing algorithms described in the paper, and editing the corresponding sections of the manuscript. Urszula Libal was responsible for implementing the Python-based software, data classification, visualization, and results interpretation. Robert Hossa contributed to the conceptualization of the signal processing algorithms and to the overall writing and editing of the article. All authors reviewed and approved the final manuscript

References

1. Bishop C.M. (2006), *Pattern Recognition and Machine Learning*, Springer, New York.

2. Breiman L. (2001), Random forests, *Machine Learning*, **45**(1): 5–32, https://doi.org/10.1023/A:1010933404324.

3. Cherifi E., Guerti M. (2021), Conditional random fields applied to Arabic orthographic-phonetic transcription, *Archives of Acoustics*, **46**(2): 237–247, https://doi.org/10.24425/aoa.2021.136574.

4. Cortes C., Vapnik V. (1995), Support-vector networks, *Machine Learning*, **20**(3): 273–297, https://doi.org/10.1007/BF00994018.

5. Cover T., Hart P. (1967), Nearest neighbor pattern classification, *IEEE Transactions on Information Theory*, **13**(1): 21–27, https://doi.org/10.1109/TIT.1967.1053964.

6. Davis S., Mermelstein P. (1980), Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **28**(4): 357–366, https://doi.org/10.1109/TASSP.1980.1163420.

7. de Cheveigné A., Kawahara H. (2002), YIN, a fundamental frequency estimator for speech and music, *The Journal of the Acoustical Society of America*, **111**(4): 1917–1930, https://doi.org/10.1121/1.1458024.

8. Dempster A.P., Laird N.M., Rubin D.B. (1977), Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society: Series B (Methodological)*, **39**(1): 1–22, https://doi.org/10.1111/j.2517-6161.1977.tb01600.x.

9. Gmyrek S., Hossa R. (2025a), Improving the vowel classification accuracy using varying signal frame length, *Vibrations in Physical Systems*, **36**(1): 2025114, https://doi.org/10.21008/j.0860-6897.2025.1.14.

10. Gmyrek S., Hossa R. (2025b), Robust speech parametrization based on pitch synchronized cepstral solutions, *International Journal of Electronics and Telecommunications*, **71**(3): 1–7, https://doi.org/10.24425/ijet.2025.153614.

11. Gmyrek S., Hossa R., Makowski R. (2023), Reducing the impact of fundamental frequency on the HFCC parameters of the speech signal, [in:] *2023 Signal Processing Symposium (SPSympo)*, pp. 49–52, https://doi.org/10.23919/SPSympo57300.2023.10302705.

12. Gmyrek S., Hossa R., Makowski R. (2024), Amplitude spectrum correction to improve speech signal classification quality, *International Journal of Electronics and Telecommunications*, **70**(3): 569–574, https://doi.org/10.24425/ijet.2024.149580.

13. Goodfellow I., Bengio Y., Courville A. (2016), *Deep Learning*, MIT Press.

14. Hastie T., Tibshirani R., Friedman J. (2001), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, New York.

15. HAZEN T. (2000), A comparison of novel techniques for rapid speaker adaptation, *Speech Communication*, **31**(1): 15–33, https://doi.org/10.1016/S0167-63 93(99)00059-X.

16. HOSSA R., MAKOWSKI R. (2016), An effective speaker clustering method using UBM and ultra-short training utterances, *Archives of Acoustics*, **41**(1): 107–118, https://doi.org/10.1515/aoa-2016-0011.

17. JASSEM W. (1973), *Fundamentals of Acoustic Phonetics* [in Polish: *Podstawy Fonetyki Akustycznej*], PWN.

18. KUHN M., JOHNSON K. (2013), *Applied Predictive Modeling*, Springer, New York.

19. KUNDEGORSKI M., JACKSON P.J.B., ZIÓŁKO B. (2014), Two-microphone dereverberation for automatic speech recognition of Polish, *Archives of Acoustics*, **39**(3): 411–420, https://doi.org/10.2478/aoa-2014-0045.

20. LIBAL U., BIERNACKI P. (2024a), Drone flight detection at an entrance to a beehive based on audio signals, *Archives of Acoustics*, **49**(3): 459–468, https://doi.org/10.24425/aoa.2024.148796.

21. LIBAL U., BIERNACKI P. (2024b), MFCC-based sound classification of honey bees, *International Journal of Electronics and Telecommunications*, **70**(4): 849–853, https://doi.org/10.24425/ijet.2024.152069.

22. LIBAL U., BIERNACKI P. (2024c), MFCC selection by LASSO for honey bee classification, *Applied Sciences*, **14**(2): 913, https://doi.org/10.3390/app14020913.

23. MACIEJKO W. (2015), The effect of voice over IP transmission degradations on MAP-EM-GMM speaker verification performance, *Archives of Acoustics*, **40**(3): 407–417, https://doi.org/10.1515/aoa-2015-0042.

24. MAKOWSKI R. (2011), *Automatic Speech Recognition – Selected Problems* [in Polish: *Automatyczne Rozpoznawanie Mowy – Wybrane Zagadnienia*], Oficyna Wydawnicza Politechniki Wrocławskiej.

25. MAUCH M., DIXON S. (2014), Pyin: A fundamental frequency estimator using probabilistic threshold distributions, [in:] *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 659–663, https://doi.org/10.1109/ICASSP.2014.68 53678.

26. MCLACHLAN G., PEEL D. (2000), *Finite Mixture Models*, Wiley-Interscience.

27. NAITO M., DENG L., SAGISAKA Y. (2002), Speaker clustering for speech recognition using vocal tract parameters, *Speech Communication*, **36**(3–4): 305–315, https://doi.org/10.1016/S0167-6393(00)00089-3.

28. NEDELJKOVIĆ Ž., MILOŠEVIĆ M., DUROVIĆ Ž. (2020), Analysis of features and classifiers in emotion recognition systems: Case study of Slavic languages, *Archives of Acoustics*, **45**(1): 129–140, https://doi.org/10.24425/aoa.2020.132489.

29. NG A.Y. (2004), Feature selection, L_1 vs. L_2 regularization, and rotational invariance [in:] *Proceedings of the Twenty-First International Conference on Machine Learning*, pp. 78–85, https://doi.org/10.1145/1015330.1015435.

30. PIĄTEK Z., KŁACZYŃSKI M. (2021), Acoustic methods in identifying symptoms of emotional states, *Archives of Acoustics*, **46**(2): 259–269, https://doi.org/10.24425/aoa.2021.136580.

31. QUATIERI T.F. (2001), *Discrete-Time Speech Signal Processing: Principles and Practice*, Prentice Hall, Upper Saddle River, NJ.

32. RABINER L., SCHAFER R. (2010), *Theory and Application of Digital Speech Processing*, Pearson.

33. REYNOLDS D.A. (2009), Gaussian mixture models, [in:] *Encyclopedia of Biometrics*, Li S.Z., Jain A. [Eds.], Springer, https://doi.org/10.1007/978-0-387-73003-5_196.

34. RUMELHART D.E., HINTON G.E., WILLIAMS R.J. (1986), Learning representations by back-propagating errors, *Nature*, **323**(6088): 533–536, https://doi.org/10.1038/323533a0.

35. SKOWRONSKI M.D., HARRIS J. (2004), Exploiting independent filter bandwidth of human factor cepstral coefficients in automatic speech recognition, *The Journal of the Acoustical Society of America*, **116**(3): 1774–1780, https://doi.org/10.1121/1.1777872.

36. STEFANOWSKA A., ZIELIŃSKI S.K. (2024), Speech emotion recognition using a multi-time-scale approach to feature aggregation and an ensemble of SVM classifiers, *Archives of Acoustics*, **49**(2): 153–168, https://doi.org/10.24425/aoa.2024.148784.

37. UMA MAHESWARI S., SHAHINA A., RISHICKESH R., NAYEEMULLA KHAN A. (2020), A study on the impact of Lombard effect on recognition of Hindi syllabic units using CNN based multimodal ASR systems, *Archives of Acoustics*, **45**(3): 419–431, https://doi.org/10.24425/aoa.2020.134058.

38. UPADHYAYA P., FAROOQ O., ABIDI M.R., VARSHNEY P. (2015), Comparative study of visual feature for bimodal Hindi speech recognition, *Archives of Acoustics*, **40**(4): 609–619, https://doi.org/10.1515/aoa-2015-0061.

Research Paper

Benchmarking the First Realistic Dataset for Speech Separation

Rawad MELHEM*, Oumayma AL DAKKAK, Assef JAFAR

Higher Institute for Applied Sciences and Technology
Damascus, Syria

*Corresponding Author e-mail: rawad.melhem@hiast.edu.sy

This paper presents a thorough benchmarking analysis of a recently introduced realistic dataset for speech separation tasks. The dataset contains audio mixtures that replicate real-life scenarios and is accompanied by ground truths, making it a valuable resource for researchers. Although the dataset construction methodology was recently disclosed, its benchmarking and detailed performance analysis have not yet been conducted. In this study, we evaluate the performance of four speech separation models using two distinct testing sets, ensuring a robust evaluation. Our findings underscore the dataset's efficacy to advance speech separation research within authentic environments. Furthermore, we propose a novel approach for assessing metrics in real-world speech separation systems, where ground truths are unavailable. This method aims to improve accuracy evaluations and refine models for practical applications. We make the dataset publicly available to encourage innovation and collaboration in the field.

Keywords: single-channel; speech separation; deep learning; corpus; datasets.

1. Introduction

Speech separation remains an active research area, with the primary challenge being the separation of speech mixtures in realistic environments. This field has numerous applications, including automatic speech recognition (ASR), speaker verification, automatic captioning for audio and video recordings, human-machine interaction, and hearing aid devices. Traditional methods, such as non-negative matrix factorization (NMF) (ISMAEL, KADHIM, 2024), complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) (MELHEM *et al.*, 2024a), and independent component analysis (ICA) (KARIYAPPA *et al.*, 2023) have attempted to address these challenges, but their effectiveness has often been limited, particularly since they typically require prior knowledge of the speakers' data.

Recently, deep learning has significantly advanced speech separation techniques. Many studies have employed supervised learning with clean datasets containing ground truths; however, performance tends to decline in real-world scenarios. To enhance robustness, some researchers have explored training supervised models using noisy datasets that better reflect actual recordings. Additionally, there has been a shift toward unsupervised learning to improve separation accuracy when dealing with real mixtures.

Creating a realistic dataset with ground truths for speech separation is particularly challenging, as it is impossible to record the same utterance twice, once for a clean ground truth and then in a mixture with another speaker. In (MELHEM *et al.*, 2024b), there was introduced the first realistic dataset for speech separation Realistic_TIMIT_2mix, which includes ground truths. We detailed the methodology for its construction and compared it with a synthetic dataset to demonstrate its effectiveness.

In this study, we extend our earlier efforts by conducting an analysis of the realistic dataset, examining its characteristics and evaluating its potential to enhance the efficacy of speech separation models and make it available online for public[1]. In addition,

---

[1] Avaiable at https://drive.google.com/drive/folders/1ViMQ BN04ct0sKw66hSZytQIo89INahx-?usp=sharing.

we show how to enhance metric assessments in speech separation. The key contributions of our study include:

- analyzing the specifications of the realistic dataset;
- benchmarking the dataset against state-of-the-art deep learning models;
- publicly releasing the realistic dataset for broader use;
- introducing a novel methodology for evaluating speech separation metrics in real-world environments based on the construction of the realistic dataset.

The paper is organized as follows: Sec. 2 reviews related work on realistic datasets for speech separation; Sec. 3 provides a description of the dataset, Sec. 4 presents experiments for benchmarking the dataset; Sec. 5 discusses the obtained results, Sec. 6 explores how leveraging our dataset can enhance metric assessment in speech separation; and finally, Sec. 7 concludes the paper.

2. Related work

Most proposed solutions in speech separation have mainly employed supervised learning methodologies, as demonstrated in studies by SAIJO *et al.* (2024) and WANG (2024). These approaches typically utilize a synthetic corpus for training, such as WSJ0_2mix dataset (HERSHEY *et al.*, 2016), which consists of clean, read speech in near-field conditions. While the results in idealized environments were commendable, the accuracy of these methods tends to degrade when confronted with more realistic scenarios. In response, some researchers have ventured into unsupervised techniques, directly addressing realistic mixtures to enhance separation accuracy in practical settings, as noted in (WANG, WATANABE, 2023; HAN, LONG, 2023). However, tackling realistic mixtures in an unsupervised manner presents significant challenges, necessitating extensive analyses of the auditory scene.

Numerous studies have sought to find effective solutions through supervised methods, leading to the development of specialized training sets. For instance, in (WICHERN *et al.*, 2019; MACIEJEWSKI *et al.*, 2020; COSENTINO *et al.*, 2020), the authors constructed a synthetic noisy datasets by separately recording noise signals and clean speech. These recordings were then mathematically combined to create noisy mixtures, ensuring that clean ground truths were available for the corresponding clean mixtures. Conversely, some researchers opted to capture speech signals in conjunction with noise, resulting in authentic noisy speech signals. This approach involved the addition of pairs of signals to generate noisy mixtures, as seen in datasets such as Mixer6 (BRANDSCHAIN *et al.*, 2010) and VoxCeleb (NAGRANI *et al.*, 2017). In these cases, the ground truths do not necessarily correspond to the mixtures, complicating the training process.

The CHiME datasets represent a comprehensive collection of speech data carefully designed for research in speech processing, particularly focusing on robust speech separation and recognition in challenging acoustic environments. These datasets include speech recordings in various noisy conditions, including background chatter, music, and meeting room sounds, thereby simulating real-world scenarios where speech processing systems may encounter difficulties. The CHiME datasets are available in multiple versions, including CHiME-3, CHiME-5 (BARKER *et al.*, 2015; 2018), CHiME-7, CHiME-8 (CORNELL *et al.*, 2023; 2024), and are frequently employed to evaluate algorithms designed to improve the performance of speech processing systems in noisy environments. Notably, the CHiME datasets feature multichannel recordings and lack ground truths, rendering them unsuitable for single-channel supervised learning approaches.

SUBAKAN *et al.* (2021) introduced a realistic dataset called REAL-M for speech separation. This dataset comprises utterances collected from contributors who simultaneously read predefined sentences from the LibriSpeech dataset (PANAYOTOV *et al.*, 2015) across various acoustic environments and utilizing different recording devices to replicate real-world conditions. Although REAL-M effectively captures authentic scenarios, it does not provide ground truths, which limits its applicability to unsupervised learning methodologies.

3. Dataset description

The dataset in this work is referred to as 'Realistic_TIMIT_2mix', named for its construction based on the TIMIT corpus. This process can be replicated with other corpora to generate similarly realistic datasets. Using a corpus such as TIMIT is essential for creating realistic audio mixtures. Ground truths are required for each mixture, and it is not feasible for a single person to record a sentence in isolation (to establish the ground truth) and then repeat it while another person speaks concurrently to create a mixture. The realistic dataset along with the codes utilized for recording and synchronizing the ground truth signals, is publicly accessible online under the Creative Commons Attribution 4.0 International (CC BY 4.0) license.

3.1. Dataset properties

The Realistic_TIMIT_2mix dataset is organized into four main folders: training, validation, testing, and ground_truths. The training folder contains 55 000 mixtures, totaling over 30 h of speech. The validation folder includes around 6000 mixtures, which is equivalent to approximately 5 h of speech, while the testing

folder holds around 5000 mixtures, corresponding to nearly 3 h of speech. The mixtures are named according to a common format, such as 'speaker1-speaker2.wav'. Overall, the Realistic_TIMIT_2mix dataset comprises 66 000 mono-channel audio files, each sampled at 16 KHz with 16 bits per sample, resulting in a total size exceeding 3.5 GB.

3.2. Recording procedure

The primary tool utilized for building the Realistic_TIMIT_2mix dataset is the AudioPlayerRecorder (APR) MATLAB function. This function interfaces with the sound card to both play and record audio on the left and right channels simultaneously. It specifically necessitates the use of an ASIO driver for the sound device on Windows operating systems.

During the recording process for ground truths, APR is used for a single speaker to play and record their audio. However, for mixtures, both audio files corresponding to the speakers are played simultaneously, one on the left channel and the other on the right channel of the output audio device. These mixed audio streams are then recorded into a single WAV file. The detailed algorithm is elaborated in Algorithm 1 in (Melhem *et al.*, 2024b).

The recording process consumed a total of 45 working hours, making it more time-intensive compared to constructing a synthetic dataset. In synthetic datasets, mixtures are generated programmatically by sequentially combining samples from two speakers, a process that depends on PC configurations, and it can be accomplished in under 2 h. In contrast, for the Realistic_TIMIT_2mix dataset, each audio file must be played through the PC to allow the microphone to capture it, which contributes to the increased time investment.

The recordings were performed on a PC featuring the following hardware configuration: ASUS ROG STRIX Z390-F GAMING motherboard, Intel Core i9 CPU, 64 GB of RAM, NVIDIA GeForce RTX 2080 Ti GPU, and an ROG SupremeFX 8-Channel High-Definition Audio CODEC S1220A for audio processing. The distance between the microphone and audio output device was about 2 m, with the right and left channels placed 50 cm apart. The recording took place in a lab, far away from noise and disturbances.

4. Experiments

4.1. Dataset preprocessing

The TIMIT corpus consists of clean, read speech samples recorded at 16 KHz, featuring 630 speakers from eight distinct American English dialects. Each speaker utters ten unique sentences. During this experiment, the recording process was conducted in a quiet,

noise-minimized environment. Within TIMIT dataset, each file is renamed to reflect the speaker's dialect, ID, and sentence type, as depicted in Fig. 1. This naming convention facilitates the creation of diverse audio mixes by selecting speakers of various genders, dialects, and speech patterns.
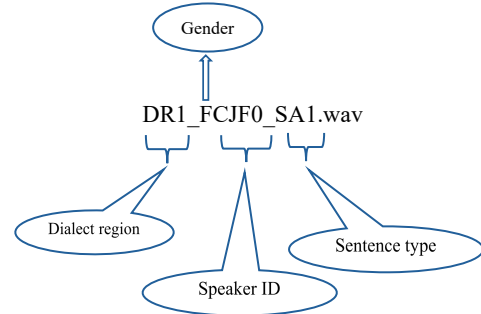

Fig. 1. File naming convention in the dataset.

4.2. Distortion and noise characterization

Each clean file within the TIMIT corpus undergoes playback via an output audio device, subsequently captured by a microphone, serving as the ground truth. As the clean file traverses a channel, comprising the output audio device, air, and microphone, distortions occur, due to attenuation and delay. To quantify these distortions, we compute the signal-to-noise ratio (SNR), where the signal is the original TIMIT file, and the noise is the difference between original file and the one captured by the microphone. Additionally, we compute the perceptual evaluation of speech quality (PESQ) and short-time objective intelligibility (STOI) to measure the quality and intelligibility of the signals yielding an average metrics presented in Table 1. These metrics were computed after applying RMS scaling (Eqs. (1) and (2)) and cross-correlation time alignment.

Table 1. SNR of ground truths in the realistic dataset.

	SNR	PESQ	STOI
Ground truths	13.34 dB	2.06	0.56

To estimate the background noise, we recorded a 30-second silent segment in the lab environment and computed its root-mean-square (RMS) level, yielding a noise floor of -62 dBFS (decibels relative to Full Scale). The lab measures 10 m × 6 m × 4 m (L × W × H). During recordings, the space is unoccupied by people but contains tables, equipment, and curtains. Under test conditions, the lab exhibits no perceptible echo.

The total loudspeaker-to-microphone distortion was quantified by averaging the difference between the original and recorded signals across 4531 files (the number of ground truths) and computing the RMS of the averaged difference (RMS_Distortion), giv-

ing −40.6 dBFS. The procedure for calculating the RMS_Distortion is as follows:

- time alignment for each file pair using cross correlation;
- RMS-scaling for each recorded signal using the following formula:

$$\text{Scaling\_Factor} = \frac{\text{RMS(Original)}}{\text{RMS(Recorded)}}, \qquad (1)$$

$$\text{Recorded} = \text{Recorded} * \text{Scaling}_{\text{Factor}}; \qquad (2)$$

- compute the difference (distortion) signal for each file;
- calculate RMS of each distortion signal;
- average the RMS values across files;
- convert the RMS distortion to dBFS using the following formula:

$$\text{distortion (dBFS)} = 20\log_{10}\left(\frac{\text{RMS\_Distortion}}{\text{Full\_Scale}}\right), \qquad (3)$$

where Full_Scale = 1 in MATLAB.

The time shift between the recorded and original signals of ground truths was determined through cross-correlation analysis, yielding an average shift of 1200 samples (approximately 75 ms at our sampling rate of 16 KHz).

For spectral analysis, we calculated the power spectral density (PSD) using a 1024-sample window (for high-frequency resolution) with 50 % overlap for each file pair (original and recorded signals). Before calculating the PSDs, we first applied RMS-scaling and time-alignment using cross correlation. The resulting spectrum, shown in Fig. 2, illustrates the distortion characteristics. The recorded and original signals match closely, except below 150 Hz, where microphone may increase low-frequency power. The recorded PSD shows higher baseline power across all frequencies because of the noise floor.
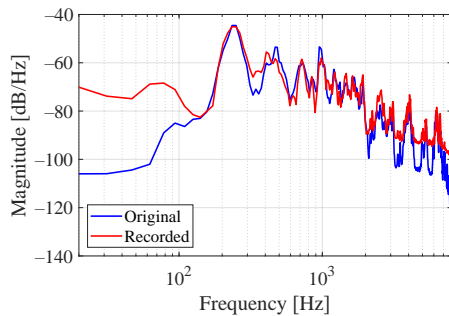

Fig. 2. PSD of the total distortion and the original signal.

The ground truths for Realistic_TIMIT_2mix must be clear and understandable. However, the primary focus is to ensure that each ground truth closely resembles itself in the mixed files. This requirement precludes using clean files as ground truths. In speech separation methods, the process depends on generating masks for each speaker and then applying these masks to the mixtures to isolate individual speakers. Any differences between the estimated masks and the corresponding ground truths can introduce distortions in the reconstructed speech of the separated speaker.

4.3. Deep learning models used

We chose four deep learning models for speech separation to benchmark the realistic dataset. TF-GridNet (Wang et al., 2023), Sepformer (Subakan et al., 2020), deep attractor network (DANet) (Zhuo et al., 2017), and DANet with bidirectional gated recurrent units (DANet-BGRU) (Melhem et al., 2021). The reasons behind choosing them is, first, to check the dataset on both modern and old versions of speech separation models, and second, to test the dataset on various architectures of deep models (e.g., transformer, long short-term memory (LSTM), and gated recurrent units (GRU)). For training Sepformer, we utilized the mixtures and ground truths in the temporal domain. However, for training TF-GridNet, DANet, and DANet-BGRU models, we employed the log magnitude of the spectrogram. The spectrogram was computed using a Hanning window of 32 ms length and an 8 ms hop size.

4.3.1. TF-GridNet

The architecture of this model consists of stacked blocks, each comprising three modules. The first module is an intra-frame spectral module, which is implemented as a single-layer bidirectional long short-term memory (BLSTM) network. The second module is a sub-band temporal module, which is also a single-layer BLSTM. The third module is a full-band self-attention mechanism implemented as a transformer, enabling the model to capture global sequence information (Wang et al., 2023). For the input, TF-GridNet uses the log magnitude of the spectrum of the mixture.

4.3.2. Sepformer

It is a RNN-free model designed for speech separation. It consists of three components: the encoder, the masking network, and the decoder. The encoder consists of a single convolutional neural network (CNN) layer that processes the temporal mixtures, while the decoder utilizes a transposed convolution layer. The masking network is structured as a block that is iterated twice. Within this block, there are also two consecutive transformers: an intra-transformer, which captures relationships among samples within the same frame, and an inter-transformer, which captures relationships across frames (Subakan et al., 2020). The parameters for the SepFormer architecture are provided in Table 2.

Table 2. Parameters for SepFormer architecture.

Parameter	Value
Number of convolutional filters	256
Kernel size	16
Number of attention heads	8
Chunk size	250
Number of DualPathBlocks	2
Number of transformers in each DualPathBlock	4

4.3.3. DANet

This model consists of four BLSTM layers followed by a fully connected layer that estimates masks for speakers (Zhuo *et al.*, 2017). As input, the model takes the logarithm of the amplitude of the spectrogram. The BLSTM, a non-causal network, aims to utilize information from both past and future frames of the speech signal. To reconstruct the separated speech signals for each speaker, the phase of the mixture is combined with inverse short-time Fourier transform (ISTFT).

4.3.4. DANet-BGRU

It is similar to the previous model but it replaces the BLSTM layers with BGRU (Melhem *et al.*, 2021). BGRUs are simpler and faster to train compared to BLTSMs. BGRU, with fewer gates compared to LSTM, streamlines the architecture, resulting in improved training efficiency and computational effectiveness.

All previous models were trained using the parameters presented in Table 3.

Table 3. Parameters for training AI models.

Parameter	Value
Learning rate	(10^{-5})
Batch size	8
Stop criterion	No improvement in validation error for three epochs
Optimizer	Adam
Normalization	z-score

The loss function for all models is the scale-invariant signal-to-distortion ratio (SI-SDR), defined as follows:

$$s_{\text{target}} = \frac{\langle \widehat{s}, s \rangle \, s}{\|s\|},$$

$$e_{\text{noise}} = \widehat{s} - s_{\text{target}}, \tag{4}$$

$$\text{SI-SDR} = 10\log\frac{\|s_{\text{target}}\|^2}{\|e_{\text{noise}}\|^2}.$$

The metrics used for evaluation are: SI-SDR (Le Roux *et al.*, 2019), PESQ (Rix *et al.*, 2001), STOI (Taal *et al.*, 2010). We evaluated the trained models on two distinct datasets. The first dataset, Realistic_TIMIT_test, was constructed similarly to Realistic_TIMIT_2mix but

with files that were not part of the training set. The second testing set is the noisy Libri2Mix (Cosentino *et al.*, 2020), which is a synthetic dataset.

5. Discussion

Figure 3 illustrates the effectiveness of using Realistic_TIMIT_2mix for training deep learning models. The models – TF-GridNet, Sepformer, DANet, and DANet-BGRU – all achieved convergence during training on Realistic_TIMIT_2mix. Specifically, Sepformer and TF-GridNet reached their minimum loss value around epoch 70, while DANet and DANet-BGRU did so around epoch 200. This discrepancy in convergence speed is attributed to the architectural differences: Sepformer and TF-GridNet leverage transformer blocks for parallel context feature extraction, whereas DANet and DANet-BGRU rely on recurrent neural network (RNN) architectures, such as LSTM and BGRU.

Notably, TF-GridNet exhibited less stable learning compared to Sepformer, likely due to fluctuations in the loss function. This instability may be linked to TF-GridNet's architecture, which includes two layers of BLSTM before the transformer, potentially complicating convergence.

Table 4 displays the evaluation results on the Realistic_TIMIT_test dataset, comprising realistic speech mixtures. Notably, Sepformer excelled across all evaluation metrics. With an SI-SDR of 9.57 dB, a challenging feature under realistic settings, alongside high PESQ (2.94) and STOI (0.66) scores, the quality of the separated speech was remarkably high. These results highlight the advantages of training models on data that closely mirrors real-world conditions, enhancing their performance. These findings underscore the effectiveness of Realistic_TIMIT_2mix in proficiently training deep learning models for speech separation.

Table 4. Performance of the learned models on the Realistic_TIMIT_test dataset.

Model	SI-SDR [dB]	PESQ	STOI
TF-Gridnet	9.23	2.87	0.64
Sepformer	9.57	2.94	0.66
DANet	8.62	2.02	0.51
DANet-GRU	8.66	2.09	0.52

The models were further assessed on the noisy Libri2Mix dataset, a synthetic collection with added noise, and the results are summarized in Table 5. Impressively, all models demonstrated strong performance, affirming the dataset's value for training purposes. Sepformer once again stood out, achieving superior results with an SI-SDR of 13.07 dB, indicative of highly accurate separation. The quality of the separated speech was notably high, as evidenced by high
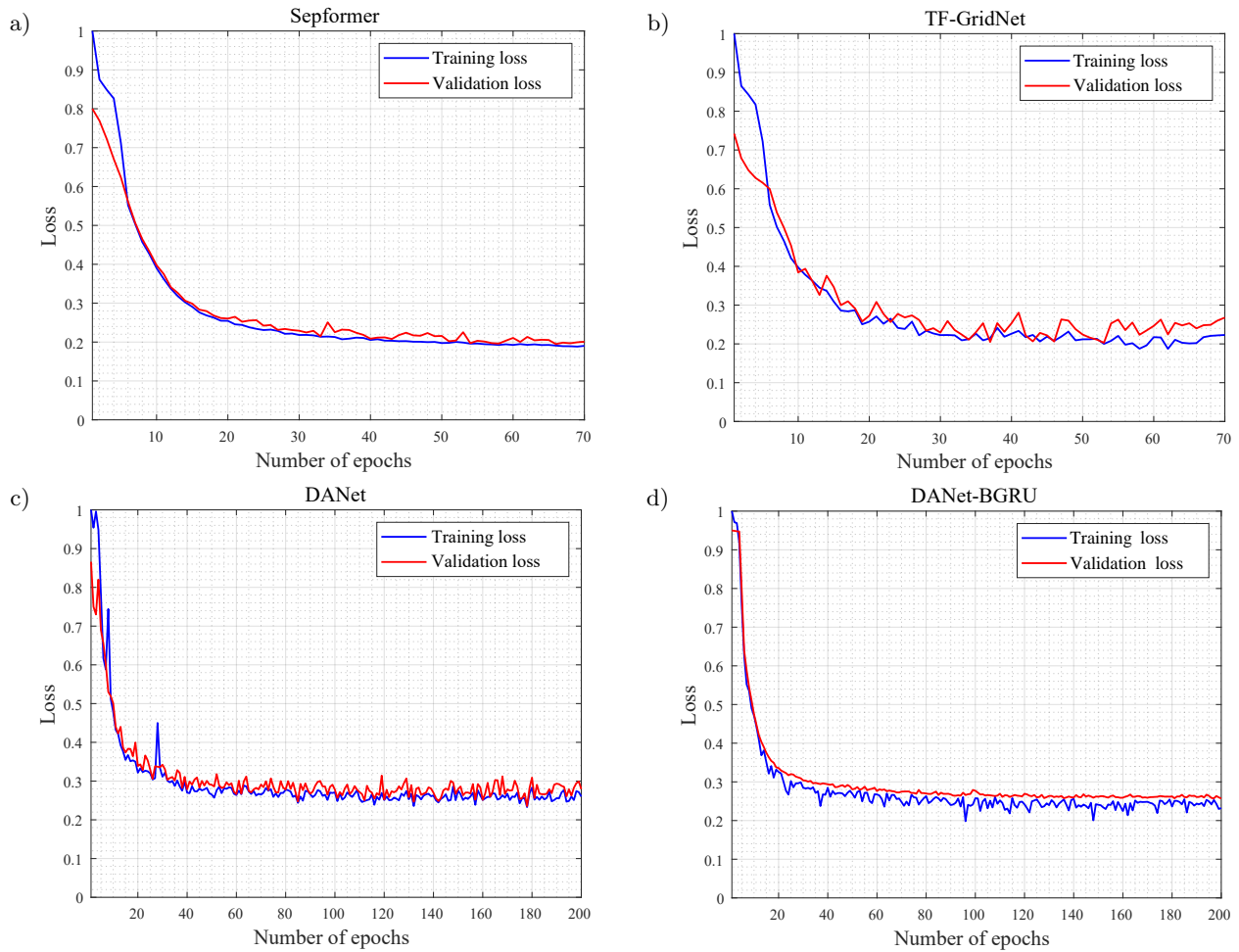
a)

b)

c)

d)

Fig. 3. Training and validation loss of the deep learning models.

Table 5. Performance of the learned models
on the noisy Libri2Mix.

Model	SI-SDR [dB]	PESQ	STOI
TF-Gridnet	12.96	3.00	0.69
Sepformer	13.07	3.21	0.70
DANet	11.01	1.94	0.57
DANet-GRU	11.08	2.01	0.59

PESQ (3.21) and STOI (0.7) scores. These findings reinforce the notion that training models on realistic datasets enhances their ability to extract speech patterns from distorted, noisy mixtures.

6. Enhancing metric assessment with Realistic_TIMIT_2mix

Assessing the performance of trained models for separating speech in real-world mixtures presents a major challenge, as there are no reference signals for comparison. While synthetic datasets offer ground truth signals, they differ from real-world scenarios. Several researchers have approached this challenge by training neural networks to forecast SI-SNR by inputting speech mixtures into them (SUBAKAN *et al.*,

2021), resulting in promising and predictive outcomes. Another alternative popular method for assessing the efficacy of speech separation models is the mean opinion score (MOS). However, MOS employment can be resource-intensive and time-consuming, as necessitates a specific demographic group with excellent hearing ability to hear separated signals and evaluate their quality.

To introduce a novel approach for measuring speech separation metrics in realistic environments, we propose leveraging the methodology employed in constructing the Realistic_TIMIT_2mix dataset. The process, illustrated in Fig. 4, consists of the following steps:

- record a speech signal for Speaker 1, saved as sp1.wav;
- record a speech signal for Speaker 2 saved asin sp2.wav;
- utilize the MATLAB function (APR) to play and independently record sp1.wav and sp2.wav, generating ground truths gt1 and gt2;
- use MATLAB function APR again to simultaneously play and record sp1 and sp2, producing a realistic speech mixture;
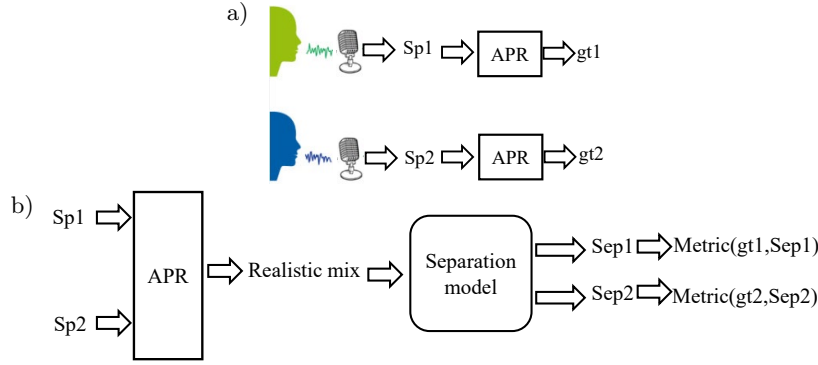
Fig. 4. Metric assessment algorithm with Realistic_TIMIT_2mix:
a) for generating ground truths; b) for realistic speech mixture.

– apply the learned model (the one under evaluation) to separate the realistic mixture into Sep1 and Sep2;

– with the separated speakers and corresponding ground truths available, calculate desired metrics (such as SI-SNR, PESQ, STOI, SDR) to assess the model's performance.

7. Conclusion

This study presented an analysis and benchmarking of Realistic_TIMIT_2mix, the first realistic dataset designed for speech separation. We conducted an in-depth examination of its characteristics and quality, subsequently releasing it online along with accompanying codes. Furthermore, we trained four deep learning models on this dataset to demonstrate its effectiveness for model training.

Additionally, we proposed a novel approach to enhance the assessment of speech separation accuracy by leveraging the methodology employed in constructing Realistic_TIMIT_2mix. Moving forward, we intend to utilize this dataset in conjunction with robust input features to further enhance the performance of speech separation systems.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Authors' contribution

Oumayma Al Dakkak, as the main supervisor for the research, conceptualized the study, reviewed the first draft, and validated the findings. Rawad Melhem conducted the methodology and wrote the first draft. Assef Jafar contributed as a co-supervisor for the study. All authors reviewed and approved the final manuscript.

References

1. Barker J., Marxer R., Vincent E., Watanabe S. (2015), The third 'CHiME' speech separation and recognition challenge: Dataset, task and baselines, [in:] *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, https://doi.org/10.1109/ASRU.2015.7404837.

2. Barker J., Watanabe S., Vincent E., Trmal J. (2018), The fifth 'CHiME' speech separation and recognition challenge: Dataset, task and baselines, [in:] *Proceedings of Interspeech*, pp. 1561–1565, https://doi.org/10.21437/Interspeech.2018-1768.

3. Brandschain L., Graff D., Cieri C., Walker K., Caruso C., Neely A. (2010), Mixer 6, [in:] *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*.

4. Cornell S. *et al.* (2024), The CHiME-8 DASR challenge for generalizable and array agnostic distant automatic speech recognition and diarization, [in:] *8th International Workshop on Speech Processing in Everyday Environments (CHiME 2024)*, https://doi.org/10.21437/CHiME.2024-1.

5. Cornell S. *et al.* (2023), The CHiME-7 DASR challenge: Distant meeting transcription with multiple devices in diverse scenarios, [in:] *7th International Workshop on Speech Processing in Everyday Environments (CHiME 2023)*, https://doi.org/10.21437/CHiME.2023-1.

6. Cosentino J., Pariente M., Cornell S., Deleforge A., Vincent E. (2020), LibriMix: An open-source dataset for generalizable speech separation, arXiv, https://doi.org/10.48550/arXiv.2005.11262.

7. Han J., Long Y. (2023), Heterogeneous separation consistency training for adaptation of unsupervised speech separation, *EURASIP Journal on Audio, Speech, and Music Processing*, **2023**: 6, https://doi.org/10.1186/s13636-023-00273-y.

8. Hershey J.R., Chen Z., Le Roux J., Watanabe S. (2016), Deep clustering: Discriminative embeddings for segmentation and separation, [in:] *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 31–35, https://doi.org/10.1109/ICASSP.2016.7471631.

9. Ismael R.N., Kadhim H.M. (2024), NNMF with speaker clustering in a uniform filter-bank for blind speech separation, *Iraqi Journal for Electrical, Electronic Engineering*, **20**(1): 111–121, https://doi.org/10.37917/ijeee.20.1.12.

10. Kariyappa S. *et al.* (2023), Cocktail party attack: Breaking aggregation-based privacy in federated learning using independent component analysis, [in:] *Proceedings of the 40th International Conference on Machine Learning*, pp. 15884–15899.

11. Le Roux J., Wisdom S., Erdogan H., Hershey J.R. (2019), SDR – Half-baked or well done?, [in:] *ICASSP 2019 – 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, https://doi.org/10.1109/ICASSP.2019.8683855.

12. Maciejewski M., Wichern G., McQuinn E., Le Roux J. (2020), WHAMR!: Noisy and reverberant single-channel speech separation, [in:] *ICASSP 2020 – 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, https://doi.org/10.1109/ICASSP40776.2020.9053327.

13. Melhem R., Hamadeh R., Jafar A. (2024a), Study of the performance of CEEMDAN in underdetermined speech separation, arXiv, https://doi.org/10.48550/arXiv.2411.11312.

14. Melhem R., Jafar A., Dakkak O.A. (2024b), Towards solving cocktail-party: The first method to build a realistic dataset with ground truths for speech separation, *Romanian Journal of Acoustics and Vibration*, **20**(1): 103–113.

15. Melhem R., Jafar A., Hamadeh R. (2021), Improving deep attractor network by BGRU and GMM for speech separation, *Journal of Harbin Institute of Technology (New Series)*, **28**(3): 90–96, https://doi.org/10.11916/j.issn.1005-9113.2019044.

16. Nagrani A., Chung J.S., Zisserman A. (2017), VoxCeleb: A large-scale speaker identification dataset, [in:] *Proceedings of INTERSPEECH 2017*, pp. 2616–2620, https://doi.org/10.21437/Interspeech.2017-950.

17. Panayotov V., Chen G., Povey D., Khudanpur S. (2015), Librispeech: An ASR corpus based on public domain audio books, [in:] *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5206–5210, https://doi.org/10.1109/ICASSP.2015.7178964.

18. Rix A.W., Beerends J.G., Hollier M.P., Hekstra A.P. (2001), Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs, [in:] *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 749–752, https://doi.org/10.1109/ICASSP.2001.941023.

19. Saijo K., Wichern G., Germain F.G., Pan Z., Le Roux J. (2024), TF-Locoformer: Transformer with local modeling by convolution for speech separation and enhancement, [in:] *2024 18th International Workshop on Acoustic Signal Enhancement (IWAENC)*, https://doi.org/10.1109/IWAENC61483.2024.10694313.

20. Subakan C., Ravanelli M., Cornell S., Bronzi M., Zhong J. (2020), Attention is all you need in speech separation, [in:] *ICASSP 2021 – 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, https://doi.org/10.1109/ICASSP39728.2021.9413901.

21. Subakan C., Ravanelli M., Cornell S., Grondin F. (2021), Real-M: Towards speech separation on real mixtures, [in:] *ICASSP 2022 – 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, https://doi.org/10.1109/ICASSP43922.2022.9746662.

22. Taal C.H., Hendriks R.C., Heusdens R., Jensen J. (2010), A short-time objective intelligibility measure for time-frequency weighted noisy speech, [in:] *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 4214–4217, https://doi.org/10.1109/ICASSP.2010.5495701.

23. Wang Z.-Q. (2024), Mixture to mixture: Leveraging close-talk mixtures as weak-supervision for speech separation, [in:] *IEEE Signal Processing Letters*, **31**: 1715–1719, https://doi.org/10.1109/LSP.2024.3417284.

24. Wang Z.-Q., Cornell S., Choi S., Lee Y., Kim B.-Y., Watanabe S. (2023), TF-GRIDNET: Making time-frequency domain models great again for monaural speaker separation, [in:] *ICASSP 2023 – 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, https://doi.org/10.1109/ICASSP49357.2023.10094992.

25. Wang Z.-Q., Watanabe S. (2023), UNSSOR: Unsupervised neural speech separation by leveraging overdetermined training mixtures, arXiv, https://doi.org/10.48550/arXiv.2305.20054.

26. Wichern G. *et al.* (2019), WHAM!: Extending speech separation to noisy environments, [in:] *Proceedings Interspeech 2019*, https://doi.org/10.21437/Interspeech.2019-2821.

27. Zhuo C., Luo Y., Mesgarani N. (2017), Deep attractor network for single-microphone speaker separation, [in:] *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, https://doi.org/10.1109/ICASSP.2017.7952155.

Research Paper

Simulation and Experiment of a Medium-Distance Underwater Ultrasonic Wireless Power Transfer System

Chenyu LIAO[(1)], Huaiqing ZHANG[(2)*], Zhi LI[(1)], Qian ZHANG[(1)]

[(1)] *College of Electrical Engineering, Chongqing University*
Chongqing, China

[(2)] *Department of Electrical Engineering, Chongqing University*
Chongqing, China

[*]Corresponding Author e-mail: zhanghuaiqing@cqu.edu.cn

This paper focuses on ultrasonic wireless power transfer (UWPT), a form of underwater wireless power transfer (WPT) using acoustic (ultrasonic) waves. We propose an ultrasonic transducer designed based on the Langevin transducer model, along with its equivalent circuit model for underwater transmission, tailored for efficient medium-distance underwater WPT. The performance of this transducer is simulated and analyzed using the acoustic-piezoelectric structure module in COMSOL. The results demonstrate that the transducer exhibits excellent underwater transmission characteristics. Additionally, an equivalent circuit model of the underwater UWPT system is developed and analyzed to characterize its transmission properties. For validation, a prototype UWPT is fabricated. At a 35 cm separation between the transmitter and receiver, the transmitter power is 2.5 W. With a 1.2 kΩ load, the received root mean square (RMS) voltage is measured at 22.8 V, corresponding to a received power of 433 mW and a transmission efficiency of 17 %. These results verify the proposed model, demonstrate a significant improvement in the transmission distance of underwater UWPT systems, and confirm the feasibility of medium-distance UWPT.

Keywords: underwater; wireless power transfer (WPT); ultrasonic; ultrasonic transducer.

1. Introduction

With the development of science and technology along with rising living standards, the global demand for electricity continues to grow, highlighting the limitations of traditional wireless power transmission (WPT) methods and increasing interest in WPT (Song *et al.*, 2021; Zhang *et al.*, 2019). WPT, which enables power transmission without physical contact, has demonstrated significant application potential through various methods such as magnetic resonance, inductive coupling, and far-field microwave power transfer (Roes *et al.*, 2013; Kim *et al.*, 2022; Singer, Robinson, 2021; Shin *et al.*, 2014; Sanni *et al.*, 2012; Valenta, Durgin, 2014). At present, electromagnetic wireless power transfer (EWPT) technologies are relatively mature. However, in underwater environment, the presence of electrolytes in sea-

water reduces the efficiency of EWPT, especially for long-distance transmission (Xu *et al.*, 2015). In the case of inductive coupling, the maximum operating distance is typically limited to only a few centimeters. In contrast, ultrasonic waves offer excellent transmission performance underwater, as sound waves undergo low attenuation in aqueous environments. Compared to electromagnetic waves, sound waves can propagate over long distances in water with lower energy loss. Therefore, ultrasonic wireless power transfer (UWPT) holds significant promise as an alternative to electromagnetic induction for underwater power delivery systems (Guida *et al.*, 2022).

Current research in UWPT technology has mostly focused on power transmission through metal walls and for subcutaneous medic. For instance, Toshihiko Ishiyama presented the concept of wireless charging for mobile electronic devices using ultrasonic trans-

ducer in air, transmitting 0.8 W of power over a distance of 0.3 m (ROES *et al.*, 2011). BAO *et al.* (2008) developed a kilowatt-scale high-power device that penetrated a metal wall at a frequency of 24.5 kHz to achieve 1 kW of UWPT through a metal wall. CHEN *et al.* (2018) demonstrated underwater UWPT with a transmission efficiency of 31 % at a distance of 5 cm between the transmitter and receiver for transducer charging. BASAERI *et al.* (2019) employed piezoelectric ceramics to power an implantable medical device, delivering 0.7 mW of power to the receiving transducer over a distance of 20 mm.

In previous underwater UWPT systems, the transmission distance has typically been confined to short ranges (less than 10 cm). In this work, by employing 28 kHz ultrasonic transducers, we extend the transfer distance to 35 cm – placing it within the medium-range regime (10 cm–50 cm). We develop a Langevin-type transducer and employ COMSOL Multiphysics to simulate its underwater propagation characteristics. Additionally, we propose an equivalent circuit model of the underwater UWPT system and derive its theoretical transfer characteristics. The accuracy of the theoretical model and the performance of the underwater WPT system are further validated through experimental testing. The feasibility of mid-range UWPT is effectively demonstrated, which is of great significance for underwater wireless power delivery.

2. Underwater WPT system

The initial step in developing an UWPT system involves constructing the underwater WPT setup. The system comprises a signal source, ultrasonic transducers at both the transmitting and receiving ends, and an impedance matching circuit designed to optimize the receiving impedance. Since piezoelectric transducers operate most efficiently at their resonant frequency, a high-frequency signal generator supplies the transmitting ultrasonic transducer with a high-frequency signal at its resonant frequency of f_0 to generate ultrasonic waves. These waves propagate through the water and are received by the ultrasonic transducer at the receiving end, where they are converted back into electrical signals. Due to the capacitive nature of the ultrasonic transducer, an impedance matching circuit is employed to eliminate reactive power generated by

the transducer and ensure maximum power transfer efficiency. The received signal is then rectified and filtered through a circuit to supply power for downstream electronic components. The structure of the underwater UWPT system is shown in Fig. 1.

The underwater ultrasonic transducer discussed in this paper is based on a Langevin-type design. It consists of a backing layer, piezoelectric ceramic elements, and a front matching layer. The Langevin transducer is constructed by stacking piezoelectric elements, which are compressed and held together in place by bolts positioned between the backing and the metal matching layer. This compression bolt applies a preload to the piezoelectric elements, maintaining them in a compressed state to prevent material fractures during high-power operation (BAO *et al.*, 2008).

From an acoustic perspective, the impedance mismatch between solid and liquid necessitates the inclusion of a matching layer to enhance transmission efficiency. The acoustic impedance of the metal at the front of the transducer is approximately 17.3 MRayl, while that of water is around 1.48 MRayl. When there is a significant mismatch in acoustic impedance between two materials, most of the sound waves will be reflected at the boundary, resulting in minimal transmission through the interface. Conversely, when the difference in acoustic impedance is relatively small, most of the sound waves can pass through the boundary. To ensure efficient underwater operation and achieve optimal acoustic impedance matching with water, a two-layer matching structure is required. One of these layers is composed of epoxy resin adhesive (CALLENS *et al.*, 2004).

The design principle is shown in Fig. 2. This method uses the classical transfer matrix method, where the thickness of each matching layer is T_n, and an acoustic impedance of the matching layer is Z_n. The transfer matrix of the n-th matching layer can be obtained as (CALLENS *et al.*, 2004):

$$T_n = \begin{bmatrix} \cos\theta_n & jZ_n\sin\theta_n \\ \dfrac{j}{Z_n}\sin\theta_n & \cos\theta_n \end{bmatrix}, \qquad (1)$$

where λ_n is the wavelength of the n-th layer, t_n is the thickness of the n-th layer, $n = 1, 2, 3, 4$, and the phase shift $\theta_n = \frac{2\pi t_n}{\lambda_n}$.
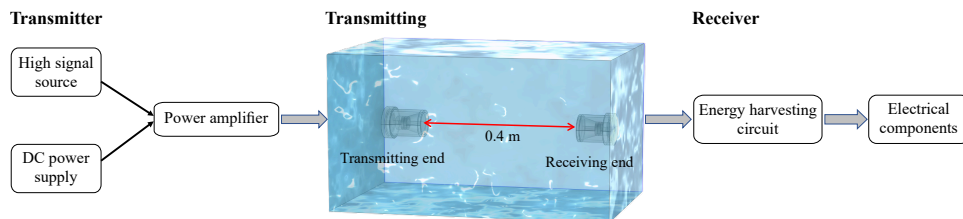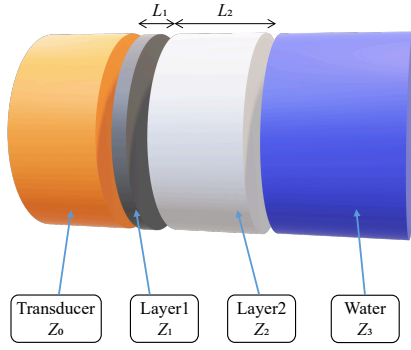

Fig. 1. Underwater UWPT system.

Fig. 2. Schematic diagram of the multi-layer matching transducer.

Impedance matching is achieved by adjusting the equivalent impedance of the ultrasonic transducer $Z_{eq} = Z_3$ at the transducer-dielectric boundary (Z_0 is the impedance of the transmission medium), where Z_{eq} is

$$Z_{eq} = \frac{T_{11}Z_0 + T_{12}}{T_{21}Z_0 + T_{22}}, \tag{2}$$

where T_{ij} is an element of the multilayer matching structure transfer matrix $\mathbf{T}$, defined by

$$T_{eq} = T_1 T_2 T_3 T_4 = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix}. \tag{3}$$

The primary advantage of this method lies in its ability to broaden the range of materials suitable for acoustic impedance matching, thereby enhancing the efficiency of underwater ultrasonic power transmission. The transducer was designed to operate at a frequency of 28 kHz, with the thickness of each layer calculated using the relevant formula: 0.3 cm for the epoxy adhesive layer and 0.98 cm for the plexiglass layer. Using these parameters, the ultrasonic transducer was modeled in COMSOL Multiphysics, as shown in Fig. 3.
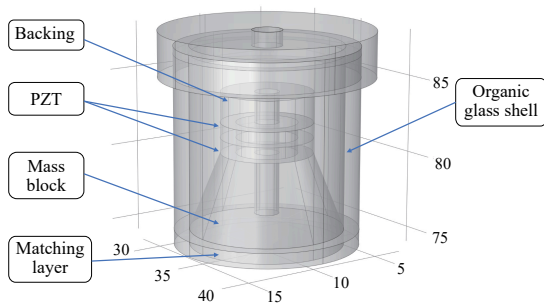

Fig. 3. Transducer simulation model.

The interior of the transducer housing is airtight. Since the acoustic impedance of air is 4×10^{-4} MRayl (SELFRIDGE, 1985), which differs significantly from both the piezoelectric ceramic and the metal acoustic impedance at the front end, the ultrasonic waves encounter a solid-gas boundary inside the housing.
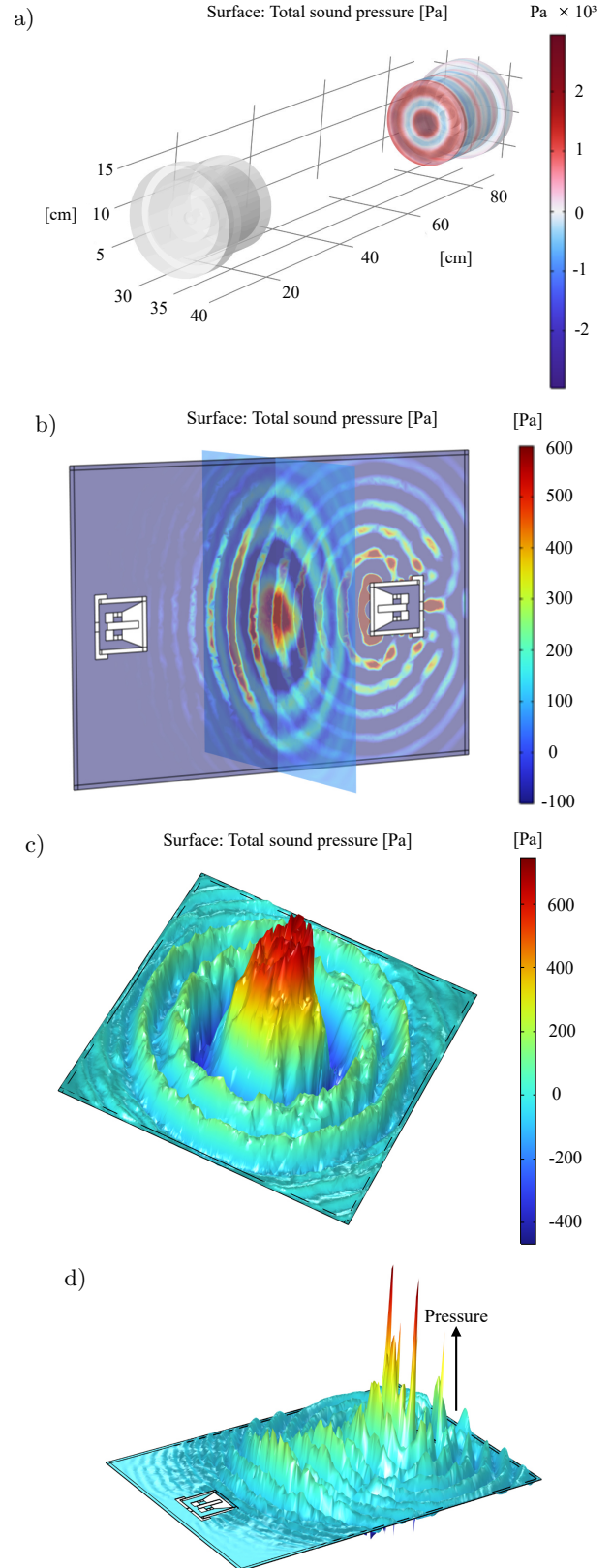

Fig. 4. COMSOL simulation of transducer sound pressure: a) sound pressure distribution at the front face of the transducer; b) multi-slice sound pressure maps; c) sound pressure levels at 25 cm from the transducer; d) *YZ*-plane acoustic pressure distribution map from the COMSOL simulation.

This design effectively minimizes the side lobes of the ultrasonic transducer, thereby ensuring efficient ultrasound emission from the front face.

After that, COMSOL Multiphysics was used to simulate the underwater performance of the ultrasonic transducer. The simulation results reveal that when the transducer operates underwater, the sound field energy generated by the transducer is primarily concentrated at its front face, as shown in Fig. 4a. By constructing an underwater multi-plane acoustic pressure map, as shown in Fig. 4b, it is observed that the side lobe pressure levels of the transducer are minimal, while a pressure level at the center of the transmission is significantly high. This indicates the transducer's excellent transmission characteristics. Furthermore, measurements of the sound pressure level at a distance of 25 cm from the transducer confirm that the central sound pressure is the strongest, as shown in Fig. 4c. The effective underwater transmission characteristics of the transducer can be seen in the YZ-plane, as shown in Fig. 4d.

3. System modeling and experimentation

3.1. Equivalent circuit of ultrasonic transducer

In order to increase the power of the transducer and the received efficiency, proper impedance matching is essential. Therefore, establishing an equivalent circuit model of the ultrasonic transducer has to be established first. Based on the operating principles of the transducer, its equivalent circuit model near the resonant frequency is shown in Fig. 5 (FAI *et al.*, 2015).
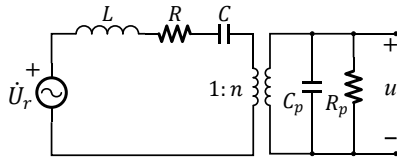

Fig. 5. Equivalent circuit diagram of ultrasonic transducer.

In Fig. 5, $\dot{U}_r$ is the induced voltage, C_p is the intrinsic capacitance of the piezoelectric material, and R_p is the leakage resistance associated with C_p. Additionally, L represents the transducer's mass, C denotes the stiffness, and R corresponds to the mechanical damping. Figure 6 illustrates the equivalent circuit diagram of the underwater UWPT system. When the excitation frequency applied to the transmitting transducer face matches it resonant frequency f_0, the transmitting

transducer generates maximum mechanical vibration, given by

$$f_0 = \frac{1}{2\pi\sqrt{L_0 C_0}}. \tag{4}$$

The transducer at the receiving end converts the incoming ultrasonic signal into an electrical signal. At this point, the input voltage at the receiving end can be modeled as a controlled current voltage source (CCVS), with a magnitude of $k\dot{I}$, where k is the control coefficient. The control coefficient k represents the attenuation coefficient of ultrasonic waves propagating in water. Under fixed experimental conditions and a predetermined propagation distance, the ultrasonic wave traveling from the transmitter to the receiver undergoes attenuation characterized by this constant coefficient k, which quantifies the energy loss incurred within the medium. In the equivalent electrical circuit, this acoustic attenuation is modeled by a CCVS that delivers the attenuated signal at the receiver. Consequently, a larger value of k corresponds to lower acoustic attenuation.

The parameters L_0, R_0, C_0 and L_1, R_1, C_1 are derived from the LRC impedance shown in Fig. 5. C_{p0} and C_{p1} represent the intrinsic capacitance of the piezoelectric transducer, while R_{p1} denotes the equivalent resistance of the energy radiated by the transducer. Furthermore, L_b represents the back-end impedance matching inductance, and R_b denotes the impedance matching resistance.

Next, we calculate the equivalent input resistance Z_1 seen from terminals a and b. When $\dot{U}_r = 0$, the equivalent impedance as viewed from the right end is Z_1, and the equivalent output impedance of the ultrasonic transducer at its receiving end is calculated as

$$Z_1 = Z_{p1}//Z_2 = \frac{Z_1 Z_2}{Z_1 + Z_2}. \tag{5}$$

When the transducer operates at its resonant frequency f_0, L_1 resonates in series with C_1. At this point, the equivalent impedance Z_1 is given by

$$Z_1 = \frac{R_1}{(R_1 \omega C_{p1})^2 + 1} - j\frac{R_1^2 \omega C_{p1}}{(R_1 \omega C_{p1})^2 + 1}. \tag{6}$$

To eliminate the reactive power, the imaginary component of the impedance Z must be set to zero. Consequently, an external inductance L_b is

$$L_b = \frac{R_1^2 C_{p1}}{(R_1 \omega C_{p1})^2 + 1}. \tag{7}$$
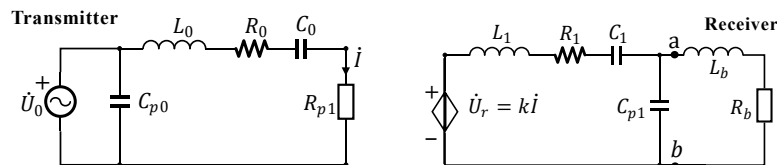

Fig. 6. Equivalent circuit of an underwater UWPT system.

After compensation, the equivalent output impedance is

$$Z_2 = \frac{R_1}{(R_1 \omega C_{p1})^2 + 1}. \tag{8}$$

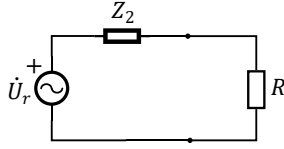The simplified equivalent circuit diagram after impedance matching is shown in Fig. 7.

Fig. 7. Simplified circuit modeling.

The voltage across R can be calculated as U_1, and according to the maximum power transfer theorem, the power at the receiving end is

$$U_1 = \frac{R \dot{U}_r}{R + Z_2}, \tag{9}$$

$$P_1 = \frac{\dot{U}_r^2}{4 Z_2}, \tag{10}$$

where P_1 is the maximum accepted power obtained when $R = Z_2$ according to the maximum power theory.

To obtain the impedance of the ultrasonic transducer, an impedance analyzer (Agilent 4294A) is used to measure the impedance of the ultrasonic transducer. Its internal impedance is shown in Fig. 8, indicating that its impedance reaches a minimum at 26.9 kHz, corresponding to its resonant frequency, with an equivalent resistance of 1200 Ω.
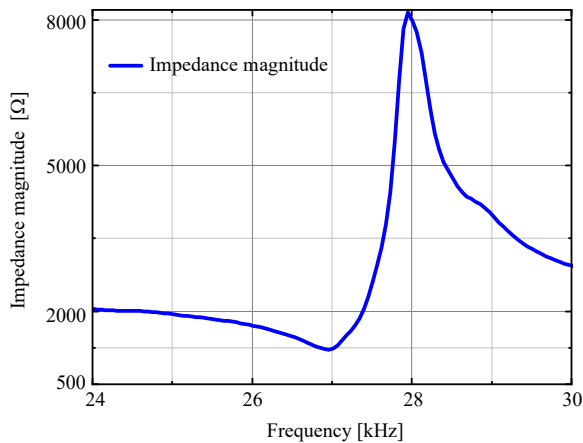

Fig. 8. Ultrasonic transducer impedance amplitude vs. frequency.

3.2. Underwater ultrasonic power transfer experiment

The underwater ultrasonic transducer used in this experiment is based on a Langevin structural design machined by the manufacturer. The receiver is identical in size to the transmitter, with a circular vibrating surface measuring 80 mm in diameter and a resonant frequency of 28 kHz. The actual resonant frequency was measured to be 26.9 kHz using an impedance analyzer. By substituting $C_{p1} = 2$ nF and $R = 1.5$ kΩ into Eqs. (7) and (8), the equivalent impedance is calculated to be approximately 1200 Ω, and the matching inductance is 3.58 mH.

A high-frequency AC signal, matching the resonant frequency of the transducer, is generated by a high-frequency signal source. After amplification by a power amplifier, a voltage of 40 V is applied at the transmitting end. The assembled underwater UWPT system is shown in Fig. 9.
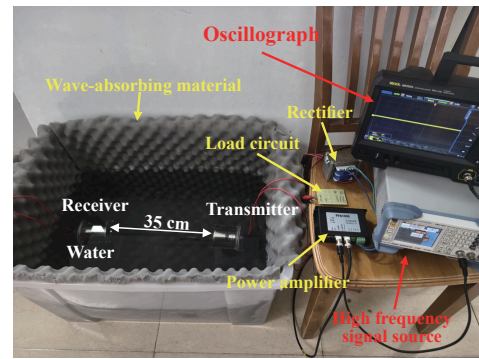

Fig. 9. Underwater UWPT system.

Piezoelectric components in ultrasonic transducers operate exclusively at their inherent resonant frequency, where they achieve maximum amplitude and optimal power conversion efficiency. When the signal frequency supplied to the ultrasonic transducer deviates from this resonant frequency, both the ultrasonic power generated at the transmitting end and the acoustic-to-electric conversion efficiency at the receiving end decrease significantly. This also causes a marked reduction in the received voltage at the receiving end. Moreover, when the receiving ultrasonic transducer is laterally displaced from the transmitting transducer (i.e., off the transmitter's central axis), the received power decreases significantly – falling by 40 % at a 5 cm offset. This indicates that the acoustic power generated by the transmitter is predominantly concentrated in the central region. These observations are consistent with the COMSOL simulation results.

As shown in Fig. 10, the maximum voltage at the receiving end occurs when the signal source output frequency is 26.88 kHz, with a peak-to-peak value reaching 64 V. Notably, this resonant frequency is slightly lower than previously measured values. This reduction in resonance frequency is attributed to the addition of matching layers onto the exterior of the Langevin transducer (Fei *et al.*, 2015).

Resistors of varying sizes are used as loads at the receiving end of the ultrasonic transducer to examine how the load power changes with different resis-
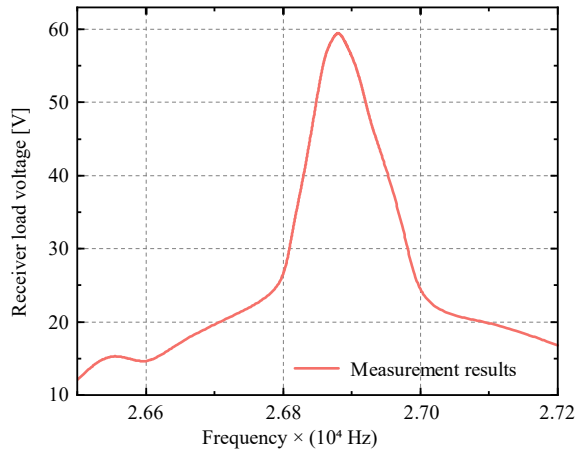
Fig. 10. Load acceptance voltage vs. input signal frequency.

tance values. While the theoretical model closely predicts the output characteristics, slight deviations from the experimental data are observed, likely due to conductance variations induced by changes in load resistance. As shown in Fig. 11, the load power increases with rising resistance, reaching an RMS voltage of
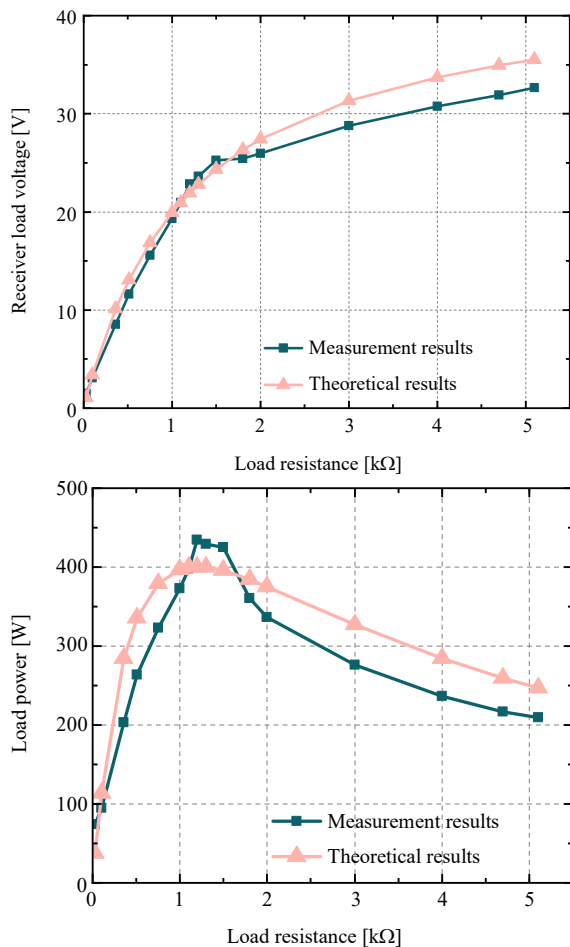

Fig. 11. Output performance characteristics
for different resistive loads.

22.8 V across the load at an impedance of 1.2 kΩ, resulting in a real power output of 433 mW. However, when the resistance deviates from this optimal value, the load power decreases significantly.

When the transmitter delivers a real power of 2.5 W, the system achieves an effective transmission efficiency of 17 % over a 35 cm distance, corresponding to a received power of 433 mW. This study demonstrates the feasibility of medium-distance underwater UWPT, offering a viable solution to the challenges associated with underwater ultrasonic power transmission.

4. Conclusion

In this paper, we proposed an ultrasonic transducer designed based on the Langevin transducer model, along with its equivalent circuit model for underwater UWPT systems. The performance of the transducer was simulated and analyzed using the acoustic-piezoelectric structure module in COMSOL. The results demonstrated that the ultrasonic transducer exhibits excellent underwater transmission characteristics. Subsequently, the equivalent circuit model was employed to theoretically predict the transmission characteristics of the UWPT system. The model provided an effective means of calculating the expected performance of the designed system. An experimental platform for the underwater UWPT system was developed to validate the proposed model. The measured efficiency matched the theoretical predictions within experimental error.

Currently, UWPT distances are typically limited to millimeters or a few centimeters. The underwater ultrasonic transducer developed in this study significantly extends the transmission range, enabling WPT over a distance of 35 cm. The maximum transmission power of the UWPT system was calculated using the equivalent circuit model. Finally, the receiver registered a power output of 433 mW, corresponding to a net transmission efficiency of 17 %. It was demonstrated that the theoretical model aligns with the experimental data and effectively enhances both the transmission distance and efficiency of underwater WPT. This approach provides a non-contact solution for underwater wireless charging, enabling marine monitoring sensors and small autonomous underwater vehicles to recharge without frequent retrieval or cumbersome tethered connections. It may offer valuable benefits for future underwater sensor networks and small-scale autonomous subsea vehicles.

Future work will focus on constructing a larger water tank to enable ultrasonic energy transmission into the far-field region, thereby significantly extending the underwater WPT distance. Additionally, underwater phased-array techniques will be investigated to enable precise acoustic beam focusing and enhanced directivity, further improving transmission power.

References

1. Bao X. *et al.* (2008), High-power piezoelectric acoustic-electric power feedthru for metal walls, [in:] *Proc. SPIE 6930, Industrial and Commercial Applications of Smart Structures Technologies 2008*, https://doi.org/10.1117/12.776473.

2. Basaeri H., Yu Y., Young D., Roundy S. (2019), A MEMS-scale ultrasonic power receiver for biomedical implants, *IEEE Sensors Letters*, **3**(4): 2501104, https://doi.org/10.1109/lsens.2019.2904194.

3. Callens D., Bruneel C., Assaad J. (2004), Matching ultrasonic transducer using two matching layers where one of them is glue, *NDT & E International*, **37**(8): 591–596, https://doi.org/10.1016/j.ndteint.2004.03.005.

4. Chen X., Xu K., Mou X., Li G. (2018), Comparison of inductively coupled and ultrasonically coupled wireless energy transmission technologies in seawater [in Chinese], *Journal of Electrical Machines and Control*, **22**(03): 9–16, https://doi.org/10.15938/j.emc.2018.03.002.

5. Fai L.H., Dai X., Hu A. (2015), Electrical modeling of a wireless ultrasonic power transfer system [in Chinese], *Chinese Journal of Electrical Engineering*, **30**(19): 85–89, https://doi.org/10.19595/j.cnki.1000-6753.tces.2015.19.012.

6. Fei C.L. *et al.* (2015), Design of matching layers for high-frequency ultrasonic transducers, *Applied Physics Letters*, **107**(12): 123505, https://doi.org/10.1063/1.4931703.

7. Guida R., Demirors E., Dave N., Melodia T. (2022), Underwater ultrasonic wireless power transfer: A battery-less platform for the Internet of Underwater Things, *IEEE Transactions on Mobile Computing*, **21**(5): 1861–1873, https://doi.org/10.1109/tmc.2020.3029679.

8. Kim H.S. *et al.* (2022), Ferroelectrically augmented contact electrification enables efficient acoustic energy transfer through liquid and solid media, *Energy & Environmental Science*, **15**(3): 1243–1255, https://doi.org/10.1039/d1ee02623b.

9. Roes M.G.L., Duarte J.L., Hendrix M.A.M., Lomonova E.A. (2013), Acoustic energy transfer: A review, *IEEE Transactions on Industrial Electronics*, **60**(1): 242–248, https://doi.org/10.1109/tie.2012.2202362.

10. Roes M.G.L., Hendrix M.A.M., Duarte J.L. (2011), Contactless energy transfer through air by means of ultrasound, [in:] *IECON 2011 – 37th Annual Conference of the IEEE Industrial Electronics Society*, pp. 1238–1243, https://doi.org/10.1109/IECON.2011.6119486.

11. Sanni A., Vilches A., Toumazou C. (2012), Inductive and ultrasonic multi-tier interface for low-power, deeply implantable medical devices, *IEEE Transactions on Biomedical Circuits and Systems*, **6**(4): 297–308, https://doi.org/10.1109/tbcas.2011.2175390.

12. Selfridge A.R. (1985), Approximate material properties in isotropic materials, *IEEE Transactions on Sonics and Ultrasonics*, **32**(3): 381–394, https://doi.org/10.1109/t-su.1985.31608.

13. Shin J. *et al.* (2014), Design and implementation of shaped magnetic-resonance-based wireless power transfer system for roadway-powered moving electric vehicles, *IEEE Transactions on Industrial Electronics*, **61**(3): 1179–1192, https://doi.org/10.1109/tie.2013.2258294.

14. Singer A., Robinson J.T. (2021), Wireless power delivery techniques for miniature implantable bioelectronics, *Advanced Healthcare Materials*, **10**(17): 2100664, https://doi.org/10.1002/adhm.202100664.

15. Song M. *et al.* (2021), Wireless power transfer based on novel physical concepts, *Nature Electronics*, **4**(10): 707–716, https://doi.org/10.1038/s41928-021-00658-x.

16. Valenta C.R., Durgin G.D. (2014), Harvesting wireless power: Survey of energy-harvester conversion efficiency in far-field, wireless power transfer systems, *IEEE Microwave Magazine*, **15**(4): 108–120, https://doi.org/10.1109/mmm.2014.2309499.

17. Xu K., Chen X., Liu D. (2015), Electrical impedance transformation technique for undersea ultrasonic coupled wireless energy transmission system [in Chinese], *Chinese Journal of Electrical Engineering*, **35**(17): 4461–4467, https://doi.org/10.13334/j.0258-8013.pcsee.2015.17.022.

18. Zhang Z., Pang H., Georgiadis A., Cecati C. (2019), Wireless power transfer – An overview, *IEEE Transactions on Industrial Electronics*, **66**(2): 1044–1058, https://doi.org/10.1109/tie.2018.2835378.

Review Paper

Sound Source Localisation in Digital Hearing Aids: A Review of Critical Factors

Fermín SCALITI[(1)] [iD], Diego A. EVIN[(1), (2)] [iD], Fabián C. TOMMASINI[(1)*] [iD]

[(1)] *Centro de Investigación y Transferencia en Acústica (CINTRA), CONICET*
Universidad Tecnológica Nacional Facultad Regional Córdoba
Córdoba, Argentina

[(2)] *Facultad de Ingeniería, Universidad Nacional de Entre Ríos*
Oro Verde, Argentina

*Corresponding Author e-mail: ftommasini@frc.utn.edu.ar

Sound source localisation is a fundamental ability for a listener's functional interaction with the environment, yet it remains a significant challenge for hearing aid users. In general, they perform worse on spatial hearing tests than individuals with normal hearing. This narrative literature review examines the critical factors affecting sound source localisation when hearing aids are worn, with a focus on direction identification. We analysed peer-reviewed articles published over the past three decades to evaluate the impacts of the type of fitting, form factor, acoustic coupling, processing delay, bandwidth, directional microphones, and dynamic range compression on localisation ability. As a general conclusion, there is a consensus in the literature that binaural and open fittings, with microphones at the ear entrance and with extended bandwidth, significantly improve localisation performance. In addition, this review is intended to provide valuable information that can guide future innovations in hearing aid technology to improve users' hearing experiences and their integration into the environment.

Keywords: hearing aids; sound source localisation; binaural hearing; head-related transfer functions.

1. Introduction

Sound source localisation is an important auditory function that refers to the ability to infer both the direction and distance of a sound source. It enables us not only to localise sound sources but also to separate sounds based on their spatial locations, which is particularly important in situations where visual information is unavailable (BLAUERT, 1997). Spatial awareness comes into play in various scenarios, such as shifting attention between speakers during conversation or understanding speech in highly reverberant rooms (MINNAAR *et al.*, 2010). Despite the higher spatial resolution of vision, the auditory modality allows us to monitor objects located anywhere around us. In addition to its role in survival, sound localisation significantly contributes to aesthetic experiences and spatial orientation (KILLION, 1997).

It is known that the human hearing system is capable of estimating the position and distance of sound sources by exploiting so-called localisation cues. The human auditory system relies on input from both ears to assess the spatial properties of complex and reverberant environments, including the number, distance, direction, and orientation of sound sources, as well as the level of reverberation (BLAUERT, 1997; DERLETH *et al.*, 2021). Normal-hearing listeners naturally master these spatial properties of sound and have a remarkable ability to localise sound sources. However, this is not the case for hearing-impaired individuals. In general, these people tend to perform worse in spatial hearing tests than people with normal hearing (AKEROYD, 2014). Such deficits in directional hearing can significantly diminish quality of life and may compromise rapid responses in potentially hazardous situations. Most of these people resort to hearing aids (HAs)

to compensate for reduced hearing sensitivity, either in one or both ears.

Listeners with hearing loss (HL) tend to localise sounds more accurately without their HAs than with them (Akeroyd, 2014; Kumar *et al.*, 2024; Van den Bogaert *et al.*, 2006). However, some studies suggest that after an adaptation period, most listeners can recover their pre-amplification localisation abilities, particularly in the horizontal plane (Drennan *et al.*, 2005). Furthermore, individuals with hearing impairment often struggle to understand speech in complex acoustic environments, such as group conversations (Gatehouse, Noble, 2004), whereas normal-hearing individuals can do so effortlessly. Several studies suggest that sound source localisation plays a crucial role in the process of spatial separation between speech and competing signals, thereby enhancing intelligibility (Bregman, 1994; Cubick *et al.*, 2018; Freyman *et al.*, 1999; Hawley *et al.*, 1999; 2004; Wang, Brown, 2006).

A pioneering study by Byrne and Noble (1998) reviewed sound localisation with HAs and their characteristics: unilateral vs. bilateral, coupling, type of HA, frequency range, and potential effects of directional microphones and compression. Although this work dates from the early days of digital HAs – which began to be commercially available in 1996 (Bille *et al.*, 1999) – to the best of our knowledge, no updated review on this topic exists. There are other studies, with a more general focus, that address issues such as localisation and spatial hearing. Georganti *et al.* (2020) discussed intelligent hearing instruments, including HAs and cochlear implants, while Akeroyd and Whitmer (2016) investigated whether people with HL are also affected by a loss in sound source localisation ability and the potential benefits of HAs. They reviewed 29 studies published since 1983 that measured acuity in direction perception in the horizontal plane (left-right and front-back acuity). Derleth *et al.* (2021) briefly summarised the advantages of binaural over monaural hearing, followed by a detailed description of related technological advances in modern HAs. Zheng *et al.* (2022) reviewed contemporary studies on localisation ability across different populations in different listening environments, examining how this ability is affected by various hearing devices, including HAs, bone-anchored hearing instruments, and cochlear implants. Shiraishi (2021) discussed sound localisation and lateralisation with bilateral bone conduction devices, middle ear implants, and cartilage conduction hearing aids.

The aim of this work is to present a narrative literature review focusing on the critical factors that influence accurate sound source localisation in digital HAs. Particular emphasis will be given to direction identification rather than distance perception or externalisation. Estimating the distance of a sound source involves more complex variables, such as sound intensity, environmental acoustics, and reverberation. Instead, when a user determines the direction of a sound, they can quickly orient their head or eyes towards the source, allowing them to integrate visual cues and enhance overall localisation accuracy.

Based on the review by Byrne and Noble (1998), the present study attempts to provide an updated state-of-the-art overview of recent advances in the field, help to contextualise the problems, and identify potential solutions. The discussion focuses on factors influencing sound localisation, including fitting type and mode of operation, form factor, acoustic coupling, processing delay, bandwidth, and algorithms for directional microphone processing and wide dynamic range compression. Finally, a section on technological advancements with potential implications on sound localisation is included.

To ensure a thorough review, an initial core set of reference articles was selected and expanded through a comprehensive search of academic databases using various keyword combinations. Additional studies were identified via forward and backwards snowballing techniques. Finally, the authors' exclusion criteria were used to form the final corpus of relevant articles.

2. Background of sound localisation

Binaural hearing enables spatial localisation, noise filtering, and speech comprehension in noisy environments by extracting perceptual cues from sounds reaching both ears (Koenig, 1950; Werner *et al.*, 2012). This section outlines key aspects of auditory localisation, with additional details available in reviews by Yost (2017) and Carlini *et al.* (2024).

The localisation of sounds in the horizontal plane depends on two main cues: the interaural time difference (ITD) and the interaural level difference (ILD). The ITD is the difference in the amount of time it takes for a sound to reach both ears, resulting in interaural phase differences that vary with frequency, and depends on head size and the speed of sound (Blauert, 1997). ITDs are most effective at frequencies below 1500 Hz, corresponding to longer wavelengths, but become ambiguous at higher frequencies, where wavelengths are shorter than the distance between the ears (Byrne, Noble, 1998; Dillon, 2012). On the other hand, the ILD is the difference in intensity between the sounds reaching each ear due to the acoustic shadow on the ear farther from the source. ILDs become effective above 1000 Hz and are particularly robust above 4000 Hz (Shaw, 1974). There is an intermediate region (between 1500 Hz and 4000 Hz) where both cues contribute, but localisation is less precise since neither ITD nor ILD is fully effective (Mills, 1958; Risoud *et al.*, 2018). For broadband sounds, low-frequency components dominate because ITD cues prevail over ILDs (Wightman, Kistler, 1992). Fur-

thermore, ITDs dominate in quiet environments, and ILDs in noisy ones (LORENZI *et al.*, 1999).

The locus of all positions with equal ITD/ILD is known as the cone of confusion. These positions are located in a cone and share an emergent axis from the ear canal (BLAUERT, 1997; SHINN-CUNNINGHAM, 2000; VON HORNBOSTEL, WERTHEIMER, 1920). To localise sound sources within the cone of confusion and in the vertical plane, listeners use spectral cues. These cues, that arise due to the transformations produced by the head and pinnae on incoming sounds, are embedded in the head-related transfer functions (HRTFs) (BLAUERT, 1997). Known as monaural cues, they provide the critical information needed for vertical localisation and for resolving front-back ambiguities.

The encoding of azimuth and elevation information is independent. Vertical localisation is disrupted when spectral cues are altered, whereas azimuthal performance remains unaffected (OLDFIELD, PARKER, 1984). However, HOFMAN *et al.* (1998) demonstrated that humans can gradually adapt to spectral alterations and relearn elevation localisation without compromising their ability to localise sounds horizontally.

Furthermore, sound source localisation is significantly influenced by dynamics. Head movements while listening to a sound increase the accuracy in determining its location. This holds true for resolving front-back ambiguities (MACPHERSON, KERR, 2008; MCANALLY, MARTIN, 2014; WALLACH, 1940) and for determining elevation (KATO *et al.*, 2003; PERRETT, NOBLE, 1997; THURLOW *et al.*, 1967).

3. Hearing aids and sound localisation

Sound localisation relies on the availability of precise timing and intensity cues throughout the frequency spectrum. As stated before, individuals with HL often exhibit diminished sound source localisation capabilities due to compromised hearing thresholds, limiting their access to both binaural and monaural localisation cues (BYRNE, NOBLE, 1998). Certain HAs do not seem to significantly improve this ability in the hearing-impaired individuals. In fact, some studies suggest that they may even worsen it compared to unaided listening (VAN DEN BOGAERT *et al.*, 2006; 2009a). AKEROYD (2014) reviewed five studies from 2005 to 2014 and found that in all of them, normal-hearing listeners performed better than unaided listeners with HL, while aided listeners exhibited equal or greater localisation errors than unaided listeners. This negative impact appears to be due to alterations of localisation cues caused by HAs (DILLON, 2012; MUELLER *et al.*, 2012). There are numerous ways in which HA processing can distort both monaural (DENK *et al.*, 2018b; DURIN *et al.*, 2014; PAUSCH *et al.*, 2018) and binaural cues (BROWN *et al.*, 2016; DILLON, 2012).

In terms of sound localisation in horizontal plane, studies indicate that people with normal hearing have a mean error ranging from 5.3° to 9° in free-field condition, and 6° to 7.5° in listening room. In contrast, bilateral HA users show significantly larger errors: 13°–16° in free-field and 12°–12.9° in listening rooms (BEST *et al.*, 2010; DORMAN *et al.*, 2016; FERNANDEZ *et al.*, 2025). Furthermore, some studies report that individuals with HL may localise sounds more accurately without HAs than with them (e.g., from 13° without HA to 17.5° with adaptive HA for a noise-free signal) (BYRNE, NOBLE, 1998; KEIDSER *et al.*, 2006; VAN DEN BOGAERT *et al.*, 2006). Possible explanations for such poor horizontal localisation with HAs result from alterations of ITD or ILD as a result of independent time delays and/or level compression in both HAs. However, BEST *et al.* (2010) suggested that lateral error angles were not affected by the presence of HAs, likely due to preserved low-frequency thresholds in participants and the direct acoustic input provided through the HA vents. Additionally, there is evidence that people adapt to altered binaural cues for localisation, thus most people become accustomed to their HAs' effects (DILLON, 2012).

Vertical localisation and front-back disambiguation are greatly impaired in individuals with high-frequency HL (BYRNE *et al.*, 1992; BYRNE, NOBLE, 1998; NOBLE *et al.*, 1994) and are not significantly improved by the use of HAs (DILLON, 2012; VAN DEN BOGAERT *et al.*, 2009a). BEST *et al.* (2010) showed that mean polar-angle errors ranged from 32° (without HA) to 37° (with HA) in people with HL, and were considerably lower in people with normal hearing (14°–18°). Furthermore, front-back localisation acuity is about 2.5 times worse than left-right acuity for aided users (AKEROYD, WHITMER, 2016).

Moreover, individuals with conductive or mixed HL exhibit inferior performance in sound source localisation compared to those with sensorineural hearing impairments. NOBLE *et al.* (1994) found that unaided listeners with sensorineural HL showed a localisation accuracy ranging from 18% (vertical) to 77% (horizontal) for frontal sounds, and from 10% (vertical) to 59% (horizontal) for lateral sounds. In contrast, a lower accuracy was observed in unaided individuals with conductive or mixed HL, with scores of 11% (vertical) and 45% (horizontal) for frontal sounds, and 13% (vertical) and 32% (horizontal) for lateral sources. This reduced performance is attributed to a greater proportion of sound being transmitted to the cochleae via bone conduction, as opposed to air conduction, which is more prevalent in individuals with normal hearing or sensorineural HL. Consequently, bone-conducted sound, which reaches both cochleae regardless of the stimulation site, reduces the normal interaural time and intensity difference information (STENFELT, GOODE, 2005). In essence, both cochleae

receive identical rather than different information. The use of HAs leads to improvement in horizontal localisation in individuals with conductive or mixed HL by increasing the proportion of air-conducted sound relative to bone-conducted sound (Byrne, Noble, 1998; Byrne *et al.*, 1995). Studies reported significant improvements in localisation errors of 23° (for induced HL) and 30° (for acquired HL) (Agterberg *et al.*, 2012; Zavdy *et al.*, 2022).

4. Critical factors for sound source localisation in hearing aids

4.1. Types of fitting and operation

There are two main types of HA fittings: unilateral and bilateral. Unilateral fitting involves the use of an HA device in one ear, whereas bilateral fitting involves the use of HA devices in both ears. However, in bilateral fittings, the HAs may operate in different modes: bilateral unlinked, bilateral linked, or binaural. Figure 1 illustrates these configurations.

It is important to clearly distinguish between the terms 'bilateral' and 'binaural' to describe modes of operation in HAs, as these terms are often used interchangeably in the literature. An HA that has a bilateral mode of operation employs one device per ear, where each uses its microphone to produce the output for the corresponding ear, with no exchange of information between the two devices (unlinked) or with exchange of some control parameters such as volume, auditory scene, or directionality mode (linked). In the first case, the original acoustic scenario is not preserved, which affects the user's ability to localise, separate, and track sound sources (Jeub *et al.*, 2010). In contrast, binaural HAs, introduced in the 1990s, allow audio signals to be transmitted between the two de-

vices (Blauert, Braasch, 2020; Kollmeier *et al.*, 1993; Wittkop *et al.*, 1996). In this configuration, both signals and control parameters received by the left and right devices are mutually shared, which better preserves acoustic cues and provides a more natural listening experience (Van den Bogaert *et al.*, 2009b).

Bilateral fitting offers advantages over unilateral fitting in many categories (detection, discrimination, speech intelligibility in quiet, speech intelligibility in noise, and localisation), as confirmed by questionnaires (Boymans *et al.*, 2008; 2009; Noble, Gatehouse, 2006) and by measuring speech perception in noise and localisation (Boymans *et al.*, 2008; Köbler, Rosenhall, 2002). In contrast, unilateral configurations tend to perform worse for horizontal localisation (Akeroyd, Whitmer, 2016; Boymans *et al.*, 2009; Noble, Gatehouse, 2006). In addition, Boymans *et al.* (2008) reported that people with HL preferred bilateral fittings over unilateral fittings because of the potential benefits of using two HAs. Nevertheless, Byrne *et al.* (1992) concluded that localisation performance in individuals with unilateral fittings was similar to that in bilateral fittings when HL was mild (<50 dB). However, these findings might not fully apply today. Given its publication date, the study likely used analogue HAs, which introduced negligible processing delays compared with modern digital HAs. This distinction is important because analogue devices have less impact on ITDs.

Although bilateral fitting and operation may support the activation of binaural hearing mechanisms, they do not ensure it (Derleth *et al.*, 2021). Georganti *et al.* (2020) indicated that bilateral HAs with parameter exchange (linked operation) can preserve binaural cues, therefore improving sound localisation performance compared with bilateral unlinked HAs.
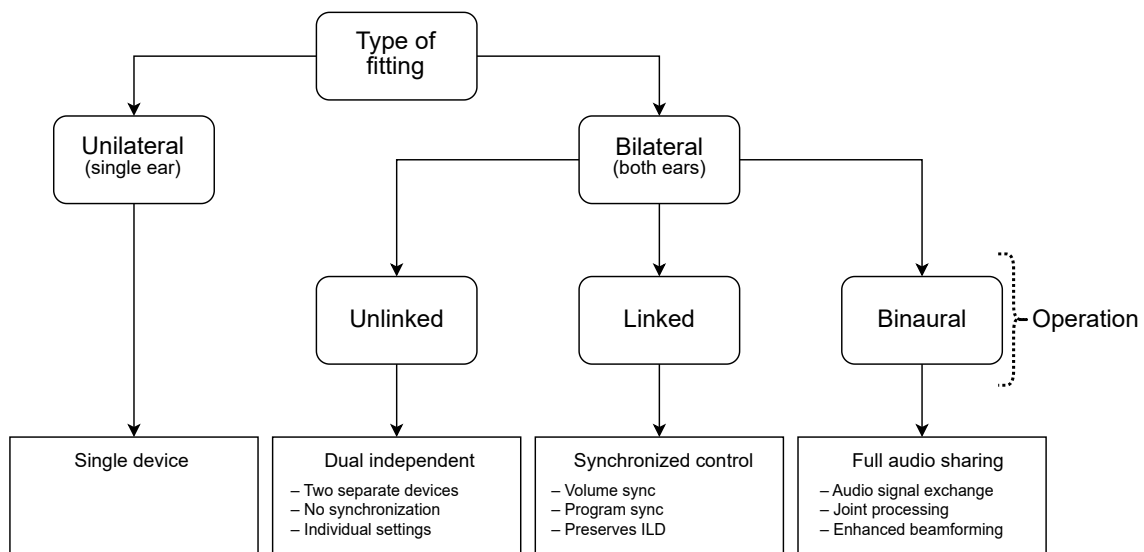

Fig. 1. HAs fitting configuration tree.

However, they do not enhance speech intelligibility (Ibrahim *et al.*, 2012; Smith *et al.*, 2008; Sockalingam *et al.*, 2009). In contrast, binaural HAs can significantly improve the performance of certain signal processing algorithms such as beamforming and noise cancellation, and thus provide better speech intelligibility. Chinnaraj *et al.* (2021) demonstrated that binaural operation can also improve horizontal localisation accuracy by about 5° compared with bilateral unlinked HAs.

4.2. Form factor

Figure 2 shows modern digital HA form factors (also called styles), which range from ultra-compact units that fit deep within the ear canal – invisible-in-the-canal (IIC) to devices that fill the outer ear to varying degrees, including completely-in-the-canal (CIC), in-the-canal (ITC), and in-the-ear (ITE), and finally to devices that are placed behind the pinna – behind-the-ear (BTE), or the receiver-in-the-canal (RIC) subtype, also called receiver-in-the-ear (RITE) (Durin *et al.*, 2014; Mondelli *et al.*, 2015). Note that these form factors also vary with respect to the placement of the microphone(s).
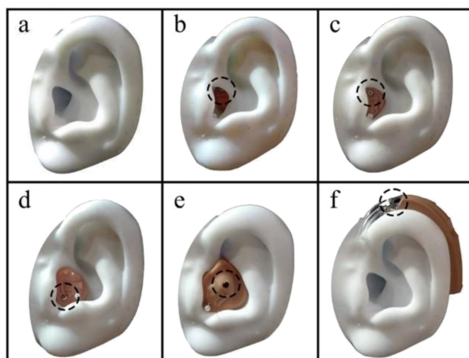

Fig. 2. HAs form factors: a) bare pinna; b) invisible-in-the-canal (IIC); c) completely-in-the-canal (CIC); d) in-the-canal (ITC); e) in-the-ear (ITE); f) behind-the-ear (BTE) HAs. Dashed circles indicate the microphone position. Adapted from Durin *et al.* (2014).[1]

HL itself does not affect a listener's HRTFs, as these are determined by individual anatomical features. When a person with HL wears HAs, HARTFs come into play (Pausch *et al.*, 2018). These HARTFs are strongly influenced by the location of the microphones (Derleth *et al.*, 2021; Diedesch, 2016; Durin *et al.*, 2014). For example, in the case of ITE HAs, the microphone is located at the entrance of the ear canal. In

---

[1] Adapted from [Durin V., Carlile S., Guillon P., Best V., Kalluri S. (2014), Acoustic analysis of the directional information captured by five different hearing aid styles, *The Journal of the Acoustical Society of America*, **136**(2): 818–828, https://doi.org/10.1121/1.4883372] with permission of Acoustical Society of America. Copyright [2014], Acoustical Society of America.

contrast, in BTE HAs, two microphones are typically located above the pinna and capture virtually no directional spectral cues (Best *et al.*, 2010; Denk *et al.*, 2018b; Diedesch, 2016). Given the similar microphone positions, there should be no major performance difference between BTE and RIC in terms of localisation (Kara *et al.*, 2024; Mondelli *et al.*, 2015). Denk *et al.* (2018a) demonstrated that direction-dependent errors can be observed at frequencies above 4000 Hz for microphones located on the concha and above 2000 Hz for microphones located behind the ear, with errors centred around the direction of frontal incidence. In the case of ITE, positioning the microphone closer to the canal does not produce better results in terms of directional cues compared with positioning it at the rear of the concha.

Regarding localisation in the frontal horizontal plane, no significant differences are observed between different HA form factors (Akeroyd, Whitmer, 2016; Van den Bogaert *et al.*, 2009a). For vertical localisation, Denk *et al.* (2019) conducted an experiment simulating different styles of HAs in individuals with normal hearing and found that BTE devices showed poorer accuracy and, in some conditions, approached chance levels. In this sense, devices positioned behind the pinna, such as BTE or its variant RIC, can also lead to increased front-back confusion (Byrne, Noble, 1998; Denk *et al.*, 2019). Other HA form factors, such as ITE, ITC or CIC, preserve most pinna-related directional cues naturally (Diedesch, 2016; Van den Bogaert *et al.*, 2009a) resulting in fewer reversals (e.g., 35 % for BTE and 26 % for CIC) (Best *et al.*, 2010). However, Best *et al.* (2010) found no differences in polar-angle error between CIC and BTE (~37°), probably due to the limited high-frequency audibility of their participants.

Despite microphone position having a strong impact on monaural cues, all HA form factors – except BTE – preserve directional information that can be partially relearned. Previous pinna-occlusion experiments have shown that by continuously wearing an earmould over a period of several days (10–60 days), allows subjects to adapt, to a greater or lesser extent, to distorted HRTFs for spatial localisation both within (Hofman *et al.*, 1998) and outside the visual field (Carlile *et al.*, 2007; 2014; Carlile, Blackman, 2014).

Durin *et al.* (2014) qualitatively described the spectral distortions of each HA form factor compared with the bare ear. Figure 3, adapted from that study, illustrates the directional transfer functions (DTFs) of HARTFs for the left ear on a logarithmic scale, highlighting how different HA form factors impact sound localisation. DTFs eliminate any directionally independent components of HARTFs to focus only on location-dependent variations (Middlebrooks, Green, 1990). Four cones of confusion are observed

for: a) $\alpha = 20°$, i.e., the contralateral hemisphere, b) $\alpha = 0°$, the median plane, c) $\alpha = -30°$, and d) $\alpha = -45°$, the ipsilateral hemisphere. The red line represents the bandwidth limit of conventional HAs (~8 kHz) (see bandwidth Subsec. 4.5).

In the DTFs of the reference condition (without an HA), the first notch (N1) is clearly evident across all cones of confusion, with its centre frequency varying monotonically with polar angle, providing a strong elevation-dependent cue. It is important to note that the N1 in the IIC style most closely resembles the reference condition. The CIC and ITC styles also show good similarity to the reference in terms of spectral shape and N1 characteristics. However, the ITE and BTE styles significantly deviate, with the N1 being unclear in the ITE and completely absent in the BTE. Other differences among HA styles are observed in peaks around 13 kHz for the frontal region and 8 kHz for the superior region. While the IIC style closely mirrors the reference, CIC and ITC styles show slight devi-

ations from it. The BTE style shows a different pattern of peaks across the whole frequency range.

Despite their poorer performance with regards to sound source localisation, STROM (2020) reported that people in the United States prefer BTE-type (BTE and RIC) HAs, which account for ~88 % of all devices sold. The reasons for this could be that such devices are comfortable to wear and virtually invisible on most ears and often provide better noise management algorithms and connectivity features compared with other HA styles.

4.3. Acoustic coupling

The occlusion effect is an increase in ear canal sound pressure level that results when bone-conducted sounds, such as the user's own voice, are prevented from escaping. In the case of HAs, the ear-mould or shell creates this effect by blocking the ear canal (GROTH, BIRKMOSE, 2004). This effect is most pronounced at
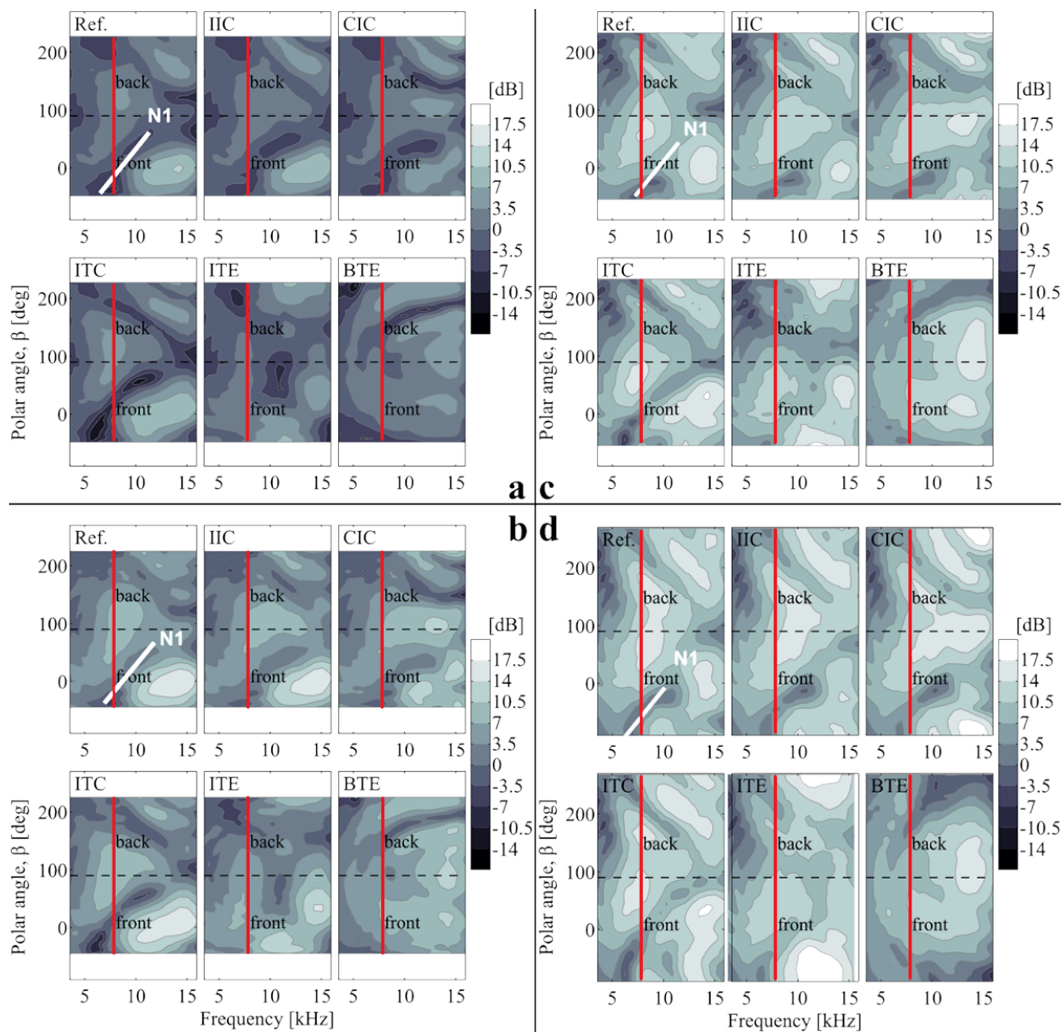

Fig. 3. Left-ear DTF log-magnitude as a function of frequency and polar angle in the planes: a) $\alpha = 20°$; b) $\alpha = 0°$; c) $\alpha = -30°$; d) $\alpha = -45°$, for five HA form factors and a reference. N1 is the first prominent notch in the DTF magnitude spectrum. The red vertical line represents ~8 kHz. Adapted from DURIN *et al.* (2014).[1]

frequencies below 500 Hz and can cause the user to perceive the sound of their own voice as loud or 'boomy' (Dillon, 2012). Self-perceived occlusion decreases as the opening size of the earmould increases (Conrad, Rout, 2013). An open-fit HA reduces or even eliminates the occlusion effect, which results in better sound quality of the user's own voice and potential improvement of the user's localisation ability (Taylor, 2006; Winkler *et al.*, 2016). This improvement is due to the fact that it allows direct and undistorted access to low-frequency content and ITD cues, which are diminished or practically eliminated when closed or semi-open earmoulds (with openings up to 2 mm) are used (Byrne *et al.*, 1996; Mueller *et al.*, 2012; Noble *et al.*, 1998).

Diedesch (2016) showed that people with normal hearing and mild HL achieved the best localisation performance in a room without HAs (~7° error), followed by open earmould fittings (~8°), with the poorest performance observed in the occluded condition with closed earmoulds (~13°). For individuals with conductive and mixed HL, Byrne *et al.* (1996) found equivalent mean localisation performance across different earmoulds. It should be noted, however, that this study likely employed analogue HAs, given its publication date.

However, open-fit earmoulds also present certain disadvantages, such as reduced audibility at low frequencies due to lower maximum gain, diminished control over gain, reduced benefits of directional microphones, compression and noise reduction, as well as impaired sound quality due to the interaction between delayed amplified sound and the non-delayed opening, and, finally, less gain available before feedback occurs (Dillon *et al.*, 2003; Stone, Moore, 2003; Winkler *et al.*, 2016). Because of this, closed fittings are typically used when there is a risk of acoustic feedback or when the user has low-frequency HL.

4.4. Processing delay

Propagation delay refers to the time it takes for sound to travel from the HA microphone to the user's ear canal (Groth, Birkmose, 2004). This delay has been extensively studied for both open and closed fittings. This study focuses on the implications of delay for sound localisation in open fittings. For further details on the impact of delay in closed fittings (see Agnew, Thornton, 2000; Stone, Moore, 1999; 2002; 2003; 2005).

With the advent of digital feedback cancellation, open fittings have become widely used. The open fitting configurations result in a mixture of two versions of the same sound: on the one hand, the real-time sound with low attenuation that enters directly into the ear canal, and, on the other hand, the delayed and amplified sound provided by the HA (Stone *et al.*, 2008). Problems can arise from the interaction between

the delayed signal and the non-delayed reference signal (Roth *et al.*, 2024). Since this delay is mostly due to digital processing, the challenge is to find the balance between the desired functionality and the time required to process it.

One of the outcomes is the comb-filtering effect, which occurs when certain frequency components interfere constructively or destructively, resulting in an alternating pattern of peaks and valleys in the overall frequency response. This leads to alteration of the sound's timbre (colouration), often described as 'barrel', 'sea shell', or 'listening through a pipe'. This effect is most noticeable with short delays (<15 ms) and when the direct and delayed sounds are of equal amplitude (Stone *et al.*, 2008).

The introduction of a high-pass filter in the processed signal path can diminish the comb-filtering effect by reducing the bandwidth of the transition zone (zone affected by both low-frequency direct sound and high-frequency delayed sound) (Denk *et al.*, 2019). Bramsløw (2010) investigated this approach using a Bradley–Terry–Luce (BTL) model (Bradley, 1984) and concluded that a signal path delay of up to 10 ms can be tolerated without compromising sound quality. The study also found no significant impact of different combinations of delays and high-pass filters on speech intelligibility.

Another undesirable consequence of processing delay is the echo effect. With longer delays, the delayed sound may be perceived as an echo. This effect is strongest when the direct and delayed sounds are equal in amplitude. The tolerable delay (i.e., the point at which listeners report being bothered or annoyed by changes in sound quality) has been reported as 24 ms–30 ms for other's speech when the delay is consistent across frequencies, and around 15 ms when the delay varies, as is often the case with HA filtering (Stone, Moore, 1999; 2002). However, for the user's own voice, tolerable delays have been reported to be 9 ms–10 ms (Agnew, Thornton, 2000; Bramsløw, 2010; Groth, Birkmose, 2004; Stone *et al.*, 2008).

It is important to note that Alexander (2016) indicated that commercial HAs have processing delays between 2 ms and 8 ms, with minimal impact even when additional functionalities are activated. These values, which are below the above-mentioned tolerable delays, have remained consistent despite major technological advancements throughout the recent years.

The effect of delay on sound source localisation was studied by Denk *et al.* (2019). They revealed that the influence of a vent and processing delays generally showed no large effect in vertical-plane localisation performance. The spectral ripples produced did not significantly bias relevant spectral directional cues, especially above 4 kHz, where the ripple decreased. In lateral localisation, errors increased under HA conditions. The disruptive effect of delay is approximately additive

to the effect of microphone location. Moreover, Denk et al. (2019) showed that a simulated leakage component that directly enters the ear canal and carries unbiased directional cues in the low-frequency regime did not improve lateral sound localisation. This finding contrasts with the claim that improving sound localisation is one of the many motivations behind the use of vented or open-fitting devices (Akeroyd, Whitmer, 2016; Noble et al., 1998). These alterations originate from fluctuations in ITD and contribute to increased source widths. The effects on lateral localisation were slightly reduced by minimising comb-filtering effects through proper filter design (Denk et al., 2018c).

Derleth et al. (2021) warned that in the case of audio frame swapping (as in binaural HAs), the overall system delay is usually increased. This may degrade sound quality and can be detrimental to localisation in the lateral dimension.

4.5. Bandwidth

Most HAs have an effective frequency limit ranging from 5000 Hz to 6000 Hz (Levy et al., 2015; Moore et al., 2001), with more recent studies showing an extension up to 7000 Hz–8000 Hz (Van Eeckhoutte et al., 2020). Bandwidth limitation modifies spectral cues, increasing front-back confusions and elevation errors, which frequently manifest as biased responses (Hofmann et al., 1998).

Significant deterioration in localisation occurs when the upper cutoff frequency is reduced to 6000 Hz, and vertical localisation is severely impaired when it drops to 4000 Hz (Byrne, Noble, 1998). Therefore, accurate sound localisation performance requires this range to be extended. Although an 8000 Hz cutoff frequency is sufficient for proper speech intelligibility, speech contains information between 8000 Hz and 16 000 Hz that is essential for accurate vertical localisation. Up-down cues are located mainly in the 6 kHz–12 kHz band, while front-back cues are in the 8 kHz–16 kHz range (Langendijk, Bronkhorst, 2002). Best et al. (2005) showed that people exhibited a polar angle error of 30.9° when localising broadband speech, which increased to 46.4° when the speech was low-pass filtered at 8000 Hz. Furthermore, it is important to note that, when the bandwidth is limited to this frequency, there are small differences between HARTFs of different HA form factors. Only the BTE devices exhibit noticeable distortions (see Fig. 3) due to a different pattern of peaks across the whole frequency range (Best et al., 2010; Durin et al., 2014).

4.6. Processing algorithms

Signal-processing features, such as highly directional microphone technology or wide dynamic range compression (WDRC), can affect ILDs. In both cases, the signal appears to originate directly in front of the listener, rather than from their actual location (Diedesch, 2016; Picou et al., 2014; Wiggins, Seeber, 2011).

4.6.1. Directional microphones

The deterioration that HAs cause in horizontal localisation becomes greater when additional algorithms, such as adaptive directional microphones, are activated, resulting in an increase in errors between 1.5° and 10° compared to an omnidirectional microphone condition (Keidser et al., 2006; Van den Bogaert et al., 2006). Directional microphones, also known as beamformers, enhance sounds coming from a particular direction, usually the front, while suppressing noise coming from other directions (Dillon, 2012). Directional microphones are not designed to improve localisation but to improve the signal-to-noise ratio (SNR) and thus speech intelligibility in noisy conditions (Byrne, Noble, 1998). Higher directionality offers better speech recognition, although makes it difficult to detect sources not located in front of the listener (Picou et al., 2014). Directional microphones significantly reduce horizontal localisation accuracy compared to omnidirectional ones, with accuracy decreasing to around 50 % at high directionality levels (Akeroyd, Whitmer, 2016; Picou et al., 2014). ITD cues are likely to be affected because different internal time delays are used to implement each specific polar pattern (Keidser et al., 2006; Van den Bogaert et al., 2006). ILDs are also affected because the polar pattern response shapes differ between the right and left devices, altering the level differences between ears. Spectral cues also change with the direction of the sound depending on the polar pattern used, since it is both frequency- and direction-dependent. However, front-back confusions can be reduced via a weak beamformer for frequencies above 1 kHz, which approximates the directionality of the pinna (Keidser et al., 2006; 2009).

If the directionality is adaptive – meaning the devices can switch between omnidirectional and directional modes within each ear depending on changes in the noise scenario or acoustic environment – a mismatch may occur between the directionality modes of the HAs, causing a distortion of binaural cues and significantly increasing lateral-plane localisation errors (Van den Bogaert et al., 2006).

Several methods aim to maximise the SNR while partially preserving binaural cues. A basic approach involves using an open-fit earmould or applying the beamformer only at mid and high frequencies, which helps to maintain ITDs intact at low frequencies (Derleth et al., 2021). This is essential for maintaining localisation abilities, as ITDs dominate over ILDs in determining sound localisation. On the other hand, studies like those by Piechowiak et al. (2015)

and JESPERSEN *et al.* (2021) suggest adapting the microphone mode – switching between omni-directional and directional – depending on the acoustic scene. Finally, additional approaches include the binaural multichannel Wiener filter with HRTF preservation (MARQUARDT *et al.*, 2015) and the Jackrabbit method (GOMEZ, 2019), both of which incorporate algorithms designed to preserve spatial cues while enhancing SNR.

4.6.2. Dynamic range compression

WDRC reduces the range of ILD and introduces unwanted fluctuations when applied independently. This produces a conflict between unaltered ITDs and distorted ILDs, which can impair sound localisation for some stimuli, increase diffusivity, cause possible image separations, and reduce the externalisation of sound sources (HASSAGER *et al.*, 2017; WIGGINS, SEEBER, 2011). However, listeners exhibit a considerable capacity to adapt to abnormal ILDs. The impact on localisation ability varies depending on the compression ratio. According to BAKKE (1999), a 2:1 compression ratio (reducing the ILD by half) in a single band does not affect localisation, whereas complete removal of ILDs degrades localisation accuracy. It is worth noting that the exception to this criterion is observed in low-frequency stimuli, where localisation remains unaffected even with 100 % attenuation of ILDs.

In order to preserve the original ILDs, some manufacturers implement a binaural exchange of amplification parameters between bilateral devices (DERLETH *et al.*, 2021) to perform what is known as 'linked compression' or 'binaural compression'. This ensures that the original loudness difference is preserved at the outputs of both HA devices. However, as reported by KORHONEN *et al.* (2015), there is a slight improvement in horizontal localisation yet it is not statistically significant. In contrast, with unlinked fast-acting compression, greater spatial separation between the target source and the masking source is needed to maintain spatial selective auditory attention (SCHWARTZ, SHINN-CUNNINGHAM, 2013). This may require increased auditory effort, especially in noisy environments.

5. Technological advancements with potential impact on sound localisation

Many recent technological advances in HAs are associated with the application of wireless connectivity and deep learning (DL) methods (HOHMANN, 2023).

Wireless connectivity in HAs was developed primarily to enhance speech understanding. However, it also contributes to improved spatial perception and sound localisation. A major advancement in this area is ear-to-ear (E2E) transmission between bilaterally fitted devices, which enables two key modalities: a) control data transmission, which synchronises programs and volume adjustments to preserve ILDs in bilateral linked operation; b) audio data transmission, which facilitates joint processing of acoustic signals for enhanced beamforming and noise-cancellation algorithms in binaural operation (GEORGANTI *et al.*, 2020; MECKLENBURGER, GROTH, 2016). However, E2E transmission involve a trade-off between optimising SNR and preserving spatial binaural cues, requiring careful balancing of these competing objectives (NEHER *et al.*, 2017). Modern HAs primarily utilise two wireless technologies: Bluetooth – whose adoption is increasing (PICOU, 2020; 2022) – and near-field magnetic induction (NFMI). While Bluetooth is mainly used for audio streaming, NFMI prevails in E2E transmission due to its technical advantages. However, there are cases where 2.4 GHz wireless technology (e.g., Bluetooth) is also employed for E2E (DERLETH *et al.*, 2021).

In addition, recent studies applied DL methods to improve sound source localisation in indoor and reverberant environments (CHEN *et al.*, 2025; GRUMIAUX *et al.*, 2022; SONG *et al.*, 2022; VECCHIOTTI *et al.*, 2019; ZHANG *et al.*, 2016), with some approaches incorporating head movements (GARCÍA-BARRIOS *et al.*, 2022; MA *et al.*, 2017). However, their implementation in current HAs remains challenging primarily due to the limited processing power, memory, latency constraints, and energy consumption. Deep neural networks, for example, typically employ multiple layers of interconnected nodes and time-frequency representations of raw audio to preserve optimal spatial representations, demanding substantial computational and memory resources (GOLI, VAN DE PAR, 2023). In terms of latency, some algorithms require relatively long time windows to achieve optimal results (GARCÍA-BARRIOS *et al.*, 2022). High computational demands also lead to increased energy consumption, which can compromise the battery life of compact devices (KHAN *et al.*, 2025). Alternative approaches, such as cloud processing or smartphone co-processors, have been proposed but may introduce additional delays and additional points of failure (FABRY, BHOWMIK, 2021). Addressing these limitations requires developing low-power, high-speed processing architectures, including edge AI and neuromorphic computing (KHAN *et al.*, 2025).

6. Summary

Sound source localisation is an important function of human hearing that is impaired in people with HL, even when HAs are used. However, the extent to which HAs affect localisation performance varies according to different critical factors summarized in a list and in Table 1 for a quick and concise overview.

– Bilateral and binaural fitting and operation: there is evidence supporting the superiority of bilateral

Table 1. List of the main features of HAs affecting sound source localisation;
✓ best performance; – low performance; × worst performance; ⋆ does not significantly affect.

Features		Localisation in the horizontal plane		Localisation in the vertical plane
		ITD	ILD	
Fitting	Unilateral	× Unilateral configurations tend to be worse for horizontal localisation (Akeroyd, Whitmer, 2016)		× Vertical localisation is affected on the ipsilateral side, depending on the HARTFs
	Bilateral	– Bilateral fittings offer better localisation compared to unilateral fittings (Boymans *et al.*, 2008; 2009; Köbler, Rosenhall, 2002; Noble, Gatehouse, 2006)		× Vertical localisation is affected on both sides
	Binaural	✓ Binaural or bilateral-linked HAs provide a better preservation of acoustic cues (Van den Bogaert *et al.*, 2009b)		× Vertical localisation is affected on both sides
Form factor		⋆ No significant differences between different form factors (Akeroyd, Whitmer, 2016; Van den Bogaert *et al.*, 2009a)		× BTE HAs impair vertical cues, increasing front-back confusions, compared to aids with microphones at the ear canal (Best *et al.*, 2010; Denk *et al.*, 2018a; Diedesch, 2016; Durin *et al.*, 2014)
Acoustic coupling	Open	✓ Best localisation performance due to undistorted access to low frequencies and ITDs through the vent (Byrne *et al.*, 1996; Diedesch, 2016; Mueller *et al.*, 2012; Noble *et al.*, 1998)	⋆	⋆
	Closed	× Worst horizontal localisation performance (Diedesch, 2016)	⋆	⋆
Delay		× Delayed output impairs horizontal sound localisation due to fluctuations in ITD (Denk *et al.*, 2019)	⋆	⋆ Spectral ripple created by the delay does not significantly bias the relevant spectral directional cues (Denk *et al.*, 2019)
Bandwidth		⋆	⋆	× Narrow HA bandwidth negatively affects vertical localisation performance, as high-frequency spectral cues are eliminated (Best *et al.*, 2005; 2010; Dillon, 2012; Durin *et al.*, 2014; Langendijk, Bronkhorst, 2002; Levy *et al.*, 2015)
Directional microphones		× ITD cues are likely to be affected because different internal time delays are used to implement each specific polar pattern (Keidser *et al.*, 2006; Van den Bogaert *et al.*, 2006)	× ILD is affected because the polar pattern response shapes differ between the right and left devices, altering the level differences between ears (Keidser *et al.*, 2006; Van den Bogaert *et al.*, 2006)	× The spectral cues also change with the direction of the sound, depending on the polar pattern used (Keidser *et al.*, 2006) ✓ Front/back confusions can be reduced using a weak beamformer for frequencies above 1 kHz (Keidser *et al.*, 2009)
Dynamic range compression		⋆	× Reduces ILDs (Hassager *et al.*, 2017; Wiggins, Seeber, 2011), with impact varying depending on the compression ratio (Bakke, 1999)	⋆

fittings over unilateral ones in improving localisation. Binaural operation or bilateral linked operation, which allows for coordinated processing between the two devices, offers even greater benefits.

– Form factor and microphone location: the location of HA microphones significantly affects HARTFs. Devices with microphones at the ear canal entrance better preserve natural pinna cues and directional information, which users can partially relearn. In contrast, BTE models capture fewer directional spectral cues, leading to impaired vertical localisation and increased front-back confusions.

– Acoustic coupling: open fittings generally preserve spatial hearing cues better than closed fittings, particularly for low-frequency sounds and ITDs. However, this comes at the expense of reduced effectiveness of directional microphones, noise reduction, and overall gain.

– Signal processing delay: while spectral ripples caused by processing delays does not significantly affect monaural cues, they can impact lateral localisation due to fluctuations in ITDs.

– Bandwidth: the typically narrow bandwidth of HAs negatively impacts vertical sound localisation by eliminating high-frequency spectral cues that are crucial for this function.

– Directional microphones and WDRC: these technologies can alter ILDs, potentially affecting localisation performance. However, some modern devices incorporate design constraints to partially preserve binaural cues. Moreover, users often adapt to compression effects and can relearn horizontal-plane localisation over time.

7. Conclusions

This review has examined the impact of various HA features on sound source localisation, revealing several key insights. A general consensus in the literature points to the superiority of binaural and open fittings, particularly those with microphones positioned at the ear canal entrance and with extended bandwidth, in enhancing localisation performance. The findings also highlight the intricate relationship between HA characteristics and their influence on spatial hearing abilities.

This underscores the need for a balanced approach in HA development, weighing the advantages of different technologies and design choices against their potential impact on localisation capabilities. Future developments in HA technology should aim to enhance natural localisation cues without compromising usability or advanced signal-processing features.

Finally, an effort has been made to compile quantitative findings across studies; however, direct comparisons or concise summaries are often challenging due to differences in experimental designs and methodologies. Nevertheless, this review attempted to provide valuable information to guide future innovations in HA technology that seek to improve user's hearing experience and their integration into the environment.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Authors contribution

All authors conceptualized the study and wrote the original draft. Fermín Scaliti curated the data and conducted the investigation process. Diego A. Evin and Fabián C. Tommasini supervised the work and Fabián C. Tommasini acquired funding. All authors reviewed and approved the final manuscript.

References

1. Agnew J., Thornton J.M. (2000), Just noticeable and objectionable group delays in digital hearing aids, *Journal of the American Academy of Audiology*, **11**(6): 330–336, https://doi.org/10.1055/s-0042-1748062.

2. Agterberg M.J.H., Snik A.F.M., Hol M.K.S., Van Wanrooij M.M., Van Opstal A.J. (2012), Contribution of monaural and binaural cues to sound localization in listeners with acquired unilateral conductive hearing loss: Improved directional hearing with a bone-conduction device, *Hearing Research*, **286**(1–2): 9–18, https://doi.org/10.1016/j.heares.2012.02.012.

3. Akeroyd M.A. (2014), An overview of the major phenomena of the localization of sound sources by normal-hearing, hearing-impaired, and aided listeners, *Trends in Hearing*, **18**, https://doi.org/10.1177/2331216514560442.

4. Akeroyd M.A., Whitmer W.M. (2016), Spatial hearing and hearing aids, [in:] *Hearing Aids*, Popelka G.R., Moore B.C.J., Fay R.R., Popper A.N. [Eds.], Springer International Publishing, pp. 181–215, https://doi.org/10.1007/978-3-319-33036-5_7.

5. Alexander J. (2016), Hearing aid delay and current drain in modern digital devices, *Canadian Audiologist*, **3**(4): 4.

6. Bakke M.H. (1999), *The contribution of interaural intensity differences to the horizontal auditory localization of narrow bands of noise*, Ph.D. Thesis, The City University of New York.

7. Best V., Carlile S., Jin C., van Schaik A. (2005), The role of high frequencies in speech localization,

The Journal of the Acoustical Society of America, **118**(1): 353–363, https://doi.org/10.1121/1.1926107.

8. Best V., Kalluri S., McLachlan S., Valentine S., Edwards B., Carlile S. (2010), A comparison of CIC and BTE hearing aids for three-dimensional localization of speech, *International Journal of Audiology,* **49**(10): 723–732, https://doi.org/10.3109/14992027.2010.484827.

9. Bille M., Jensen A.-M., Kjærbøl E., Vesterager V., Sibelle P., Nielsen H. (1999), Clinical study of a digital vs an analogue hearing aid, *Scandinavian Audiology,* **28**(2): 127–135, https://doi.org/10.1080/010503999424851.

10. Blauert J. (1997), *Spatial Hearing: The Psychophysics of Human Sound Localization,* The MIT Press.

11. Blauert J., Braasch J. [Eds.] (2020), *The Technology of Binaural Understanding,* Springer Nature.

12. Boymans M., Goverts S.T., Kramer S.E., Festen J.M., Dreschler W.A. (2009), Candidacy for bilateral hearing aids: A retrospective multicenter study, *Journal of Speech, Language, and Hearing Research,* **52**(1): 130–140, https://doi.org/10.1044/1092-4388(2008/07-0120).

13. Boymans M., Theo Goverts S., Kramer S.E., Festen J.M., Dreschler W.A. (2008), A prospective multi-centre study of the benefits of bilateral hearing aids, *Ear and Hearing,* **29**(6): 930–941, https://doi.org/10.1097/AUD.0b013e31818713a8.

14. Bradley R.A. (1984), Paired comparisons: Some basic procedures and examples, [in:] *Handbook of Statistics,* **4**: 299–326, https://doi.org/10.1016/S0169-7161(84)04016-5.

15. Bramsløw L. (2010), Preferred signal path delay and high-pass cut-off in open fittings, *International Journal of Audiology,* **49**(9): 634–644, https://doi.org/10.3109/14992021003753482.

16. Bregman A.S. (1994), *Auditory Scene Analysis: The Perceptual Organization of Sound,* The MIT Press.

17. Brown A.D., Rodriguez F.A., Portnuff C.D.F., Goupell M.J., Tollin D.J. (2016), Time-varying distortions of binaural information by bilateral hearing aids: Effects of nonlinear frequency compression, *Trends in Hearing,* **20**, https://doi.org/10.1177/2331216516668303.

18. Byrne, D., Noble W. (1998), Optimizing sound localization with hearing aids, *Trends in Amplification,* **3**(2): 51–73, https://doi.org/10.1177/108471389800300202.

19. Byrne D., Noble W., Glauerdt B. (1996), Effects of earmold type on ability to locate sounds when wearing hearing aids, *Ear and Hearing,* **17**(3): 218–228, https://doi.org/10.1097/00003446-199606000-00005.

20. Byrne D., Noble W., LePage B. (1992), Effects of long-term bilateral and unilateral fitting of different hearing aid types on the ability to locate sounds, *Journal of the American Academy of Audiology,* **3**(6): 369–382.

21. Byrne D., Noble W., Ter-Horst K. (1995), Effects of hearing aids on localization of sounds by people with sensorineural and conductive/mixed hearing losses, *The Australian Journal of Audiology,* **17**: 79–86.

22. Carlile S., Balachandar K., Kelly H. (2014), Accommodating to new ears: The effects of sensory and sensory-motor feedback, *The Journal of the Acoustical Society of America,* **135**(4): 2002–2011, https://doi.org/10.1121/1.4868369.

23. Carlile S., Blackman T. (2014), Relearning auditory spectral cues for locations inside and outside the visual field, *JARO: Journal of the Association for Research in Otolaryngology,* **15**: 249–263, https://doi.org/10.1007/s10162-013-0429-5.

24. Carlile S., Blackman T., Cooper J. (2007), The plastic ear: Coping with a life time of change, [in:] *19th International Congress on Acoustics (ICA).*

25. Carlini A., Bordeau C., Ambard M. (2024), Auditory localization: A comprehensive practical review, *Frontiers in Psychology,* **15**, https://doi.org/10.3389/fpsyg.2024.1408073.

26. Chen X., Zhao L., Cui J., Li H., Wang X. (2025), Hybrid convolutional neural network-transformer model for end-to-end binaural sound source localization in reverberant environments, *IEEE Access,* **13**: 36701–36713, https://doi.org/10.1109/ACCESS.2025.3545065.

27. Chinnaraj G., Tanniru K., Rajan Raveendran R. (2021), Speech perception in noise and localization performance of digital noise reduction algorithm in hearing aids with ear-to-ear synchronization, *Journal of All India Institute of Speech and Hearing,* **40**(1): 23–30, https://doi.org/10.4103/jose.JOSE_4_21.

28. Conrad S., Rout A. (2013), Perceived occlusion and comfort in receiver-in-the-ear hearing aids, *American Journal of Audiology,* **22**(2): 283–290, https://doi.org/10.1044/1059-0889(2013/11-0025).

29. Cubick J., Buchholz J.M., Best V., Lavandier M., Dau T. (2018), Listening through hearing aids affects spatial perception and speech intelligibility in normal-hearing listeners, *The Journal of the Acoustical Society of America,* **144**(5): 2896–2905, https://doi.org/10.1121/1.5078582.

30. Denk F., Ewert S.D., Kollmeier B. (2018a), Spectral directional cues captured by hearing device microphones in individual human ears, *The Journal of the Acoustical Society of America,* **144**(4): 2072–2087, https://doi.org/10.1121/1.5056173.

31. Denk F., Ewert S.D., Kollmeier B. (2019), On the limitations of sound localization with hearing devices, *The Journal of the Acoustical Society of America,* **146**(3): 1732–1744, https://doi.org/10.1121/1.5126521.

32. DENK F., HIIPAKKA M., KOLLMEIER B., ERNST S.M.A. (2018b), An individualised acoustically transparent earpiece for hearing devices, *International Journal of Audiology*, **57**(sup3): S62–S70, https://doi.org/10.1080/14992027.2017.1294768.

33. DENK F., SCHEPKER H., DOCLO S., KOLLMEIER B. (2018c), Equalization filter design for achieving acoustic transparency in a semi-open fit hearing device, [in:] *Speech Communication; 13th ITG-Symposium*, pp. 1–5.

34. DERLETH P. *et al.* (2021), Binaural signal processing in hearing aids, *Seminars in Hearing*, **42**(03): 206–223, https://doi.org/10.1055/s-0041-1735176.

35. DIEDESCH A.C. (2016), *Binaural-cue weighting in sound localization with open-fit hearing aids and in simulated reverberation*, Ph.D. Thesis, Vanderbilt University.

36. DILLON H. (2012), *Hearing Aids*, 2nd ed., Thieme, New York.

37. DILLON H., KEIDSER G., O'BRIEN A., SILBERSTEIN H. (2003), Sound quality comparisons of advanced hearing aids, *The Hearing Journal*, **56**(4): 30–40, https://doi.org/10.1097/01.HJ.0000293908.50552.34.

38. DORMAN M.F., LOISELLE L.H., COOK S.J., YOST W.A., GIFFORD R.H. (2016), Sound source localization by normal-hearing listeners, hearing-impaired listeners and cochlear implant listeners, *Audiology and Neurotology*, **21**(3): 127–131, https://doi.org/10.1159/000444740.

39. DRENNAN W.R., GATEHOUSE S., HOWELL P., TASELL D.V., LUND S. (2005), Localization and speech-identification ability of hearing-impaired listeners using phase-preserving amplification, *Ear and Hearing*, **26**(5): 461–472, https://doi.org/10.1097/01.aud.0000179690.30137.21.

40. DURIN V., CARLILE S., GUILLON P., BEST V., KALLURI S. (2014), Acoustic analysis of the directional information captured by five different hearing aid styles, *The Journal of the Acoustical Society of America*, **136**(2): 818–828, https://doi.org/10.1121/1.4883372.

41. FABRY D.A., BHOWMIK A.K. (2021), Improving speech understanding and monitoring health with hearing aids using artificial intelligence and embedded sensors, *Seminars in Hearing*, **42**(03): 295–308, https://doi.org/10.1055/s-0041-1735136.

42. FERNANDEZ J., HYVÄRINEN P., KRESSNER A.A. (2025), Localization accuracy of phantom sound sources on the horizontal plane by bilateral hearing aid users in aided free-field and non–free-field conditions, *The Journal of the Acoustical Society of America*, **157**(2): 1151–1161, https://doi.org/10.1121/10.0035828.

43. FREYMAN R.L., HELFER K.S., MCCALL D.D., CLIFTON R.K. (1999), The role of perceived spatial separation in the unmasking of speech, *The Journal of the Acoustical Society of America*, **106**(6): 3578–3588, https://doi.org/10.1121/1.428211.

44. GARCÍA-BARRIOS G., KRAUSE D.A., POLITIS A., MESAROS, A., GUTIÉRREZ-ARRIOLA J.M., FRAILE R. (2022), Binaural source localization using deep learning and head rotation information, [in:] *2022 30th European Signal Processing Conference (EUSIPCO)*, pp. 36–40, https://doi.org/10.23919/EUSIPCO55093.2022.9909764.

45. GATEHOUSE S., NOBLE W. (2004), The speech, spatial and qualities of hearing scale (SSQ), *International Journal of Audiology*, **43**(2): 85–99, https://doi.org/10.1080/14992020400050014.

46. GEORGANTI E., COURTOIS G., DERLETH P., LAUNER S. (2020), Intelligent hearing instruments – Trends and challenges, [in:] *The Technology of Binaural Understanding*, Blauert J., Braasch J. [Eds.], pp. 733–761, Springer International Publishing, https://doi.org/10.1007/978-3-030-00386-9_24.

47. GOLI P., VAN DE PAR S. (2023), Deep learning-based speech specific source localization by using binaural and monaural microphone arrays in hearing aids, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, **31**: 1652–1666. https://doi.org/10.1109/TASLP.2023.3268734.

48. GOMEZ G. (2019), *Consolidating natural spatial perception and improved SNR in hearing aids: Jackrabbit, a new method*, Ph.D, Thesis, Technische Universität München, https://mediatum.ub.tum.de/1463635.

49. GROTH J., BIRKMOSE M. (2004), Disturbance caused by varying propagation delay in non-occluding hearing aid fittings, *International Journal of Audiology*, **43**(10): 594–599, https://doi.org/10.1080/14992020400050076.

50. GRUMIAUX P.-A., KITIĆ S., GIRIN L., GUÉRIN A. (2022), A survey of sound source localization with deep learning methods, *The Journal of the Acoustical Society of America*, **152**(1): 107–151, https://doi.org/10.1121/10.0011809.

51. HASSAGER H.G., WIINBERG A., DAU T. (2017), Effects of hearing-aid dynamic range compression on spatial perception in a reverberant environment, *The Journal of the Acoustical Society of America*, **141**(4): 2556–2568, https://doi.org/10.1121/1.4979783.

52. HAWLEY M.L., LITOVSKY R.Y., COLBURN H.S. (1999), Speech intelligibility and localization in a multi-source environment, *The Journal of the Acoustical Society of America*, **105**(6): 3436–3448, https://doi.org/10.1121/1.424670.

53. HAWLEY M.L., LITOVSKY R.Y., CULLING J.F. (2004), The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer, *The Journal of the Acoustical Society of America*, **115**(2): 833–843, https://doi.org/10.1121/1.1639908.

54. HOFMAN P.M., VAN RISWICK J.G.A., VAN OPSTAL A.J. (1998), Relearning sound localization with new ears, *Nature Neuroscience*, **1**: 417–421, https://doi.org/10.1038/1633.

55. HOHMANN V. (2023), The future of hearing aid technology: Can technology turn us into superheroes?, *Zeitschrift für Gerontologie und Geriatrie*, **56**: 283–289, https://doi.org/10.1007/s00391-023-02179-y.

56. IBRAHIM I., PARSA V., MACPHERSON E., CHEESMAN M. (2012), Evaluation of speech intelligibility and sound localization abilities with hearing aids using binaural wireless technology, *Audiology Research*, **3**(1): e1, https://doi.org/10.4081/audiores.2013.e1.

57. JESPERSEN C.T., KIRKWOOD B.C., GROTH J. (2021), Increasing the effectiveness of hearing aid directional microphones, *Seminars in Hearing*, **42**(03): 224–236, https://doi.org/10.1055/s-0041-1735131.

58. JEUB M., SCHAFER M., ESCH T., VARY P. (2010), Model-based dereverberation preserving binaural cues, *IEEE Transactions on Audio, Speech, and Language Processing*, **18**(7): 1732–1745, https://doi.org/10.1109/TASL.2010.2052156.

59. KARA E. *et al.* (2024), Improving speech intelligibility in noise and spatial perception: The critical role of hearing aid microphone position, *Frontiers in Neuroscience*, **18**, https://doi.org/10.3389/fnins.2024.1475122.

60. KATO M., UEMATSU H., KASHINO M., HIRAHARA T. (2003), The effect of head motion on the accuracy of sound localization, *Acoustical Science and Technology*, **24**(5): 315–317, https://doi.org/10.1250/ast.24.315.

61. KEIDSER G. *et al.* (2006), The effect of multi-channel wide dynamic range compression, noise reduction, and the directional microphone on horizontal localization performance in hearing aid wearers, *International Journal of Audiology*, **45**(10): 563–579, https://doi.org/10.1080/14992020600920804.

62. KEIDSER G., O'BRIEN A., HAIN J.-U., MCLELLAND M., YEEND I. (2009), The effect of frequency-dependent microphone directionality on horizontal localization performance in hearing-aid users, *International Journal of Audiology*, **48**(11): 789–803, https://doi.org/10.3109/14992020903036357.

63. KHAN A., WAQAR A., KIM B., PARK D. (2025), A review on recent advances in sound source localization techniques, challenges, and applications, *Sensors and Actuators Reports*, **9**: 100313, https://doi.org/10.1016/j.snr.2025.100313.

64. KILLION M.C. (1997), A critique of four popular statements about compression, *The Hearing Review*, **4**(2): 36–38.

65. KÖBLER S., ROSENHALL U. (2002), Horizontal localization and speech intelligibility with bilateral and unilateral hearing aid amplification: Localización horizontal y discriminación del lenguaje con adaptación unilateral y bilateral de auxiliares auditivos, *International Journal of Audiology*, **41**(7): 395–400, https://doi.org/10.3109/14992020209090416.

66. KOENIG W. (1950), Subjective effects in binaural hearing, *The Journal of the Acoustical Society of America*, **22**(1): 61–62, https://doi.org/10.1121/1.1906578.

67. KOLLMEIER B., PEISSIG J., HOHMANN V. (1993), Binaural noise-reduction hearing aid scheme with real-time processing in the frequency domain, *Scandinavian Audiology Supplementum*, **38**: 28–38.

68. KORHONEN P., LAU C., KUK F., KEENAN D., SCHUMACHER J. (2015), Effects of coordinated compression and pinna compensation features on horizontal localization performance in hearing aid users, *Journal of the American Academy of Audiology*, **26**(01): 80–92, https://doi.org/10.3766/jaaa.26.1.9.

69. KUMAR S., NAYAK S., KANAGOKAR V., PITCHAI MUTHU A.N. (2024), Does bilateral hearing aid fitting improve spatial hearing ability: A systematic review and meta-analysis, *Disability and Rehabilitation: Assistive Technology*, **19**(8): 2729–2741, https://doi.org/10.1080/17483107.2024.2316293.

70. LANGENDIJK E.H.A., BRONKHORST A.W. (2002), Contribution of spectral cues to human sound localization, *The Journal of the Acoustical Society of America*, **112**(4): 1583–1596, https://doi.org/10.1121/1.1501901.

71. LEVY S.C., FREED D.J., NILSSON M., MOORE B.C.J., PURIA S. (2015), Extended high-frequency bandwidth improves speech reception in the presence of spatially separated masking speech, *Ear and Hearing*, **36**(5): e214–e224, https://doi.org/10.1097/AUD.0000000000000161.

72. LORENZI C., GATEHOUSE S., LEVER C. (1999), Sound localization in noise in normal-hearing listeners, *The Journal of the Acoustical Society of America*, **105**(3): 1810–1820, https://doi.org/10.1121/1.426719.

73. MA N., MAY T., BROWN G.J. (2017), Exploiting deep neural networks and head movements for robust binaural localization of multiple sources in reverberant environments, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, **25**(12): 2444–2453, https://doi.org/10.1109/TASLP.2017.2750760.

74. MACPHERSON E.A., KERR D. (2008), Minimum head movements required to localize narrowband sounds, [in:] *American Audiology Society 2008 Annual Meeting*.

75. MARQUARDT D., HADAD E., GANNOT S., DOCLO S. (2015), Theoretical analysis of linearly constrained multi-channel wiener filtering algorithms for combined noise reduction and binaural cue preservation in binaural hearing aids, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*,

23(12): 2384–2397, https://doi.org/10.1109/TASLP.2015.2479940.

76. McAnally K.I., Martin R.L. (2014), Sound localization with head movement: Implications for 3-d audio displays, *Frontiers in Neuroscience*, **8**, https://doi.org/10.3389/fnins.2014.00210.

77. Mecklenburger J., Groth T. (2016), Wireless technologies and hearing aid connectivity, [in:] *Hearing Aids*, Popelka G.R., Moore B.C.J., Fay R.R., Popper A.N. [Eds.], pp. 131–149, Springer International Publishing, https://doi.org/10.1007/978-3-319-33036-5_5.

78. Middlebrooks J.C., Green D.M. (1990), Directional dependence of interaural envelope delays, *The Journal of the Acoustical Society of America*, **87**(5): 2149–2162, https://doi.org/10.1121/1.399183.

79. Mills A.W. (1958), On the minimum audible angle, *The Journal of the Acoustical Society of America*, **30**(4): 237–246, https://doi.org/10.1121/1.1909553.

80. Minnaar P., Favrot S., Buchholz J.M. (2010), Improving hearing aids through listening tests in a virtual sound environment, *The Hearing Journal*, **63**(10): 40–44, https://doi.org/10.1097/01.HJ.0000389926.64797.3e.

81. Mondelli M.F.C.G., Garcia T.M., Hashimoto F.M.T., Rocha A.V. (2015), Open fitting: Performance verification of receiver in the ear and receiver in the aid, *Brazilian Journal of Otorhinolaryngology*, **81**(3): 270–275, https://doi.org/10.1016/j.bjorl.2014.08.013.

82. Moore B.C.J., Stone M.A., Alcántara J.I. (2001), Comparison of the electroacoustic characteristics of five hearing aids, *British Journal of Audiology*, **35**(5): 307–325, https://doi.org/10.1080/00305364.2001.11745249.

83. Mueller M.F., Kegel A., Schimmel S.M., Dillier N., Hofbauer M. (2012), Localization of virtual sound sources with bilateral hearing aids in realistic acoustical scenes, *The Journal of the Acoustical Society of America*, **131**(6): 4732–4742, https://doi.org/10.1121/1.4705292.

84. Neher T., Wagener K.C., Latzel M. (2017), Speech reception with different bilateral directional processing schemes: Influence of binaural hearing, audiometric asymmetry, and acoustic scenario, *Hearing Research*, **353**: 36–48, https://doi.org/10.1016/j.heares.2017.07.014.

85. Noble W., Byrne D., Lepage B. (1994), Effects on sound localization of configuration and type of hearing impairment, *The Journal of the Acoustical Society of America*, **95**(2): 992–1005, https://doi.org/10.1121/1.408404.

86. Noble W., Gatehouse S. (2006), Effects of bilateral versus unilateral hearing aid fitting on abilities measured by the speech, spatial, and qualities of hearing scale (SSQ): Efectos de la adaptación uni o bilateral de auxiliares auditivos en las habilidades medidas la escala de cualidades auditiva, espacial y del lenguaje (SSQ), *International Journal of Audiology*, **45**(3): 172–181, https://doi.org/10.1080/14992020500376933.

87. Noble W., Sinclair S., Byrne D. (1998), Improvement in aided sound localization with open earmolds: Observations in people with high-frequency hearing loss, *Journal of the American Academy of Audiology*, **9**(1): 25–34.

88. Oldfield S.R., Parker S.P.A. (1984), Acuity of sound localisation: A topography of auditory space. II. Pinna cues absent, *Perception*, **13**(5): 601–617, https://doi.org/10.1068/p130601.

89. Pausch F., Aspöck L., Vorländer M., Fels J. (2018), An extended binaural real-time auralization system with an interface to research hearing aids for experiments on subjects with hearing loss, *Trends in Hearing*, **22**, https://doi.org/10.1177/2331216518800871.

90. Perrett S., Noble W. (1997), The contribution of head motion cues to localization of low-pass noise, *Perception & Psychophysics*, **59**: 1018–1026, https://doi.org/10.3758/BF03205517.

91. Picou E.M., Aspell E., Ricketts T.A. (2014), Potential benefits and limitations of three types of directional processing in hearing aids, *Ear and Hearing*, **35**(3): 339–352, https://doi.org/10.1097/AUD.0000000000000004.

92. Picou E.M. (2020), MarkeTrak 10 (MT10) survey results demonstrate high satisfaction with and benefits from hearing aids, *Seminars in Hearing*, **41**(01): 021–036, https://doi.org/10.1055/s-0040-1701243.

93. Picou E.M. (2022), Hearing aid benefit and satisfaction results from the MarkeTrak 2022 survey: Importance of features and hearing care professionals, *Seminars in Hearing*, **43**(04): 301–316, https://doi.org/10.1055/s-0042-1758375.

94. Piechowiak T., Udesen J., Moeller K., Gran F., Dittberner A. (2015), Promoting off-axis listening and preserving spatial cues with binaural directionality II, [in:] *Proceedings of the International Symposium on Auditory and Audiological Research*, **5**: 285–292.

95. Risoud M. *et al.* (2018), Sound source localization, *European Annals of Otorhinolaryngology, Head and Neck Diseases*, **135**(4): 259–264, https://doi.org/10.1016/j.anorl.2018.04.009.

96. Roth S., Müller F.-U., Angermeier J., Hemmert W., Zirn S. (2024), Effect of a processing delay between direct and delayed sound in simulated open fit hearing aids on speech intelligibility in noise, *Frontiers in Neuroscience*, **17**, https://doi.org/10.3389/fnins.2023.1257720.

97. Schwartz A.H., Shinn-Cunningham B.G. (2013), Effects of dynamic range compression on spatial selective auditory attention in normal-hearing listeners, *The Journal of the Acoustical Society of America*, **133**(4): 2329–2339, https://doi.org/10.1121/1.4794386.

98. Shaw E.A.G. (1974), Transformation of sound pressure level from the free field to the eardrum in the horizontal plane, *The Journal of the Acoustical Society of America*, **56**(6): 1848–1861, https://doi.org/10.1121/1.1903522.

99. Shinn-Cunningham B. (2000), Learning reverberation: Consideration for spatial auditory displays, [in:] *International Conference on Auditory Display (ICAD 2000)*, Atlanta, Georgia, USA.

100. Shiraishi K. (2021), Sound localization and lateralization by bilateral bone conduction devices, middle ear implants, and cartilage conduction hearing aids, *Audiology Research*, **11**(4): 508–523, https://doi.org/10.3390/audiolres11040046.

101. Smith P., Davis A., Day J., Unwin S., Day G., Chalupper J. (2008), Real-world preferences for linked bilateral processing, *The Hearing Journal*, **61**(7): 33–38, https://doi.org/10.1097/01.HJ.0000325657.80281.9c.

102. Sockalingam R., Holmberg M., Eneroth K., Shulte M. (2009), Binaural hearing aid communication shown to improve sound quality and localization, *The Hearing Journal*, **62**(10): 46–47, https://doi.org/10.1097/01.HJ.0000361850.27208.35.

103. Song T., Zhang W., Chen J. (2022), An end-to-end binaural sound localization model based on the equalization and cancellation theory, [in:] *AES Europe Spring 2022 – 152nd Audio Engineering Society Convention 2022*, pp. 275–283.

104. Stenfelt S., Goode R.L. (2005), Bone-conducted sound: Physiological and clinical aspects, *Otology & Neurotology*, **26**(6): 1245–1261, https://doi.org/10.1097/01.mao.0000187236.10842.d5.

105. Stone M.A., Moore B.C.J. (1999), Tolerable hearing aid delays. I. Estimation of limits imposed by the auditory path alone using simulated hearing losses, *Ear and Hearing*, **20**(3): 182–192, https://doi.org/10.1097/00003446-199906000-00002.

106. Stone M.A., Moore B.C.J. (2002), Tolerable hearing aid delays. II. Estimation of limits imposed during speech production, *Ear and Hearing*, **23**(4): 325–338, https://doi.org/10.1097/00003446-200208000-00008.

107. Stone M.A., Moore B.C.J. (2003), Tolerable hearing aid delays. III. Effects on speech production and perception of across-frequency variation in delay, *Ear and Hearing*, **24**(2): 175–183, https://doi.org/10.1097/01.AUD.0000058106.68049.9C.

108. Stone M.A., Moore B.C.J. (2005), Tolerable hearing-aid delays: IV. Effects on subjective disturbance during speech production by hearing-impaired subjects, *Ear and Hearing*, **26**(2): 225–235, https://doi.org/10.1097/00003446-200504000-00009.

109. Stone M.A., Moore B.C.J., Meisenbacher K., Derleth R.P. (2008), Tolerable hearing aid delays. V. Estimation of limits for open canal fittings, *Ear and Hearing*, **29**(4): 601–617, https://doi.org/10.1097/AUD.0b013e3181734ef2.

110. Strom K. (2020), Hearing aid unit sales increase by 6.5% in 2019, *The Hearing Review*, **27**(2): 2.

111. Taylor B. (2006), Real-world satisfaction and benefit with open-canal fittings, *The Hearing Journal*, **59**(11): 74–82, https://doi.org/10.1097/01.HJ.0000286222.91963.39.

112. Thurlow W.R., Mangels J.W., Runge P.S. (1967), Head movements during sound localization, *The Journal of the Acoustical Society of America*, **42**(2): 489–493, https://doi.org/10.1121/1.1910605.

113. Van den Bogaert T., Carette E., Wouters J. (2009a), Sound localization with and without hearing aids, [in:] *NAG-DAGA International Conference on Acoustics*, pp. 1314–1317.

114. Van den Bogaert T., Doclo S., Wouters J., Moonen M. (2009b), Speech enhancement with multichannel Wiener filter techniques in multimicrophone binaural hearing aids, *The Journal of the Acoustical Society of America*, **125**(1): 360–371, https://doi.org/10.1121/1.3023069.

115. Van den Bogaert T., Klasen T.J., Moonen M., Van Deun L., Wouters J. (2006), Horizontal localization with bilateral hearing aids: Without is better than with, *The Journal of the Acoustical Society of America*, **119**(1): 515–526, https://doi.org/10.1121/1.2139653.

116. Van Eeckhoutte M., Folkeard P., Glista D., Scollie S. (2020), Speech recognition, loudness, and preference with extended bandwidth hearing aids for adult hearing aid users, *International Journal of Audiology*, **59**(10): 780–791, https://doi.org/10.1080/14992027.2020.1750718.

117. Vecchiotti P., Ma N., Squartini S., Brown G.J. (2019), End-to-end binaural sound localisation from the raw waveform, [in:] *ICASSP 2019 – 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 451–455, https://doi.org/10.1109/ICASSP.2019.8683732.

118. Von Hornbostel E.M., Wertheimer M. (1920), On the perception of sound direction [in German], [in:] *Proceedings of the Prussian Academy of Sciences*, pp. 388-396.

119. Wallach H. (1940), The role of head movements and vestibular and visual cues in sound localization, *Journal of Experimental Psychology*, **27**(4): 339–368, https://doi.org/10.1037/h0054629.

120. WANG D., BROWN G.J. [Eds.] (2006), *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*, Wiley-IEEE Press.

121. WERNER L., FAY R.R., POPPER A.N. [Eds.] (2012), *Human Auditory Development*, Springer, https://doi.org/10.1007/978-1-4614-1421-6.

122. WIGGINS I.M., SEEBER B.U. (2011), Dynamic-range compression affects the lateral position of sounds, *The Journal of the Acoustical Society of America*, **130**(6): 3939–3953, https://doi.org/10.1121/1.3652887.

123. WIGHTMAN F.L., KISTLER D.J. (1992). The dominant role of low-frequency interaural time differences in sound localization, *The Journal of the Acoustical Society of America*, **91**(3): 1648–1661, https://doi.org/10.1121/1.402445.

124. WINKLER A., LATZEL M., HOLUBE I. (2016), Open versus closed hearing-aid fittings: A literature review of both fitting approaches, *Trends in Hearing*, **20**, https://doi.org/10.1177/2331216516631741.

125. WITTKOP T., HOHMANN V., KOLLMEIER B. (1996), Noise reduction strategies in digital binaural hearing aids, [in:] *Psychoacoustics, Speech and Hearing Aids*, Singapore: World Scientific, pp. 245–251.

126. YOST W.A. (2017), History of sound source localization: 1850–1950, [in:] *Proceedings of Meetings on Acoustics*, **30**(1): 050002, https://doi.org/10.1121/2.0000529.

127. ZAVDY O. *et al.* (2022), The effect of hearing aids on sound localization in mild unilateral conductive hearing loss, *Journal of the American Academy of Audiology*, **33**(6): 357–363, https://doi.org/10.1055/a-1889-6578.

128. ZHANG T., MUSTIERE F., MICHEYL C. (2016), Intelligent hearing aids: The next revolution, [in:] *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 72–76, https://doi.org/10.1109/EMBC.2016.7590643.

129. ZHENG Y., SWANSON J., KOEHNKE J., GUAN J. (2022), Sound localization of listeners with normal hearing, impaired hearing, hearing aids, bone-anchored hearing instruments, and cochlear implants: A review, *American Journal of Audiology*, **31**(3): 819–834, https://doi.org/10.1044/2022_AJA-22-00006.