Directional Early-to-Late Energy Ratios to Quantify Clarity: A Case Study in a Large Auditorium

1st Alexis Campos
Perception Research
Lima, Peru
acampos@perception3d.com

2nd Shuichi Sakamoto
RIEC and GSIS
Tohoku University
Sendai, Japan
saka@ais.riec.tohoku.ac.jp

3rd César D. Salvador
Perception Research
Lima, Peru
salvador@perception3d.com

***Abstract*—Early-to-late energy ratios (ELER) are used to quantify speech intelligibility and music clarity in acoustic spaces from measurements of omnidirectional room impulse responses (RIR). Nowadays, the capture of directional RIRs is possible with spherical microphone arrays and the spherical Fourier transform. These tools are thus motivating the enhancement of omnidirectional metrics and the search for new metrics to quantify directional features of sound. This research explores a directional metric of intelligibility and clarity based on ELERs of directional RIRs. The early-to-late transition times are chosen according to the content: 50 ms for speech and 80 ms for music. The proposed metrics can therefore be interpreted as directional versions of the standard clarity indexes of speech (C50) and music (C80). Directional RIRs were captured at many seats in a large auditorium using a first-order ambisonics microphone. Supporting acoustic simulations of a cuboid room with a second-order ambisonics microphone were also used. Directional ELERs were calculated in the octave bands within the operation range of the microphones. Three directional ELER patterns were identified: an omnidirectional pattern, a dipole pointing forward and backward, and a beam pointing towards the source.**

***Index Terms*—Room impulse response, clarity, early-to-late energy ratio, spherical Fourier transform, ambisonics.**

I. Introduction

Quantifying speech intelligibility and music clarity is critical in architectural acoustics to take decisions on material acoustic qualities and sound source positioning. Room impulse responses (RIRs) [1] are a key tool to objectively quantify intelligibility and clarity. Traditionally, RIRs characterize the room response to sound transmission from an omnidirectional sound source to an omnidirectional receiver. The temporal structure of omnidirectional RIRs distinguishes direct waves from the source followed by early and late reflections from the walls, floor, and ceiling. The early-to-late transition times are commonly set by the source material: 50 ms for speech and 80 ms for music [2]. With these transition times, early-to-late energy ratios (ELER) of omnidirectional RIRs are used as objective metrics of intelligibility and clarity. Subjectively, the early part is regarded as the useful component whereas the late part as the detrimental component; the term useful-to-detrimental ratio is also used when referring to ELERs [3]. Preferred values of ELERs are between -5 and 5 dB [2], [4] whereas their just noticeable differences (JND), are between 1 and 3 dB [5], [6].

This study was partly supported by Universidad Peruana Unión (UPeU), Lima, Perú, and JSPS KAKENHI Grant Number 19H04145, Japan.

Nowadays, there is a growing interest in quantifying the directional features of intelligibility and clarity to discriminate among the different directions in which sounds may reach a listener. The perception of late reflections from specific angles was highlighted by Lachenmayr [7] when evaluating sound envelopment in concert halls. Romblom [8] also found directional changes in reverberant diffuse fields that provide cues for spatial impression. Recently, Alary [9] suggested that accurate sound reproduction should account for directional properties of late reverberation. The directional dependence of early and late reflections is also important to investigate intelligibility of speech in the context of spatial release from masking [10] where maskers from several directions interfere with a target speech. Other authors as Van Wijngaarden [11] proposed the inclusion of binaural hearing effects in a new model for the improvement of the monoaural speech transmission index prediction. Although binaural cues are crucial to discriminate among targets and maskers, reverberant spaces produce reflections from multiple directions that reduce this discrimination capacity [12]. In reverberant conditions, early or useful components can be regarded as targets, whereas late or detrimental components, as maskers.

To investigate the directional dependence of intelligibility and clarity in reverberant spaces, directional RIRs can be measured with microphone arrays [13], [14] or calculated with 3D models and numerical methods [15], [16]. In recent years, spherical microphone arrays (SMA) have made possible the analysis of reverberant sound fields and the estimation of directions of arrival (DOA) in concert halls [17]. Amengual [18] compared the accuracy of DOA using the spatial decomposition method and the sound intensity vector method considering early and late reverberation. Dick [19], on the other hand, investigated the relationship between the perception of the listener envelopment and DOA with different transition times. It was found that SMAs yield better metrics than omnidirectional or dipole microphones, suggesting that a better understanding of actual metrics and the definition of new metrics might emerge from directional analyses with SMAs.

In this paper, we explore a directional metric of intelligibility and clarity based on ELERs of directional RIRs captured

TABLE I
OBJECTIVE METRICS OF SPEECH INTELLIGIBILITY AND MUSIC CLARITY CALCULATED FROM AN OMNIDIRECTIONAL RIR, $h(t)$.

Source material	Name	Metric
Speech	Definition [20]	$D_{50} = \frac{\int_{0\text{ms}}^{50\text{ms}} h(t)^2 dt}{\int_{0\text{ms}}^{\infty} h(t)^2 dt}$.
Speech	U_{50} [21]	$U_{50} = 10 \log \frac{\int_{0\text{ms}}^{50\text{ms}} h(t)^2 dt}{\int_{50\text{ms}}^{\infty} h(t)^2 dt + \eta}$
Speech	Speech clarity [22]	$C_{50} = 10 \log \frac{\int_{0\text{ms}}^{50\text{ms}} h(t)^2 dt}{\int_{50\text{ms}}^{\infty} h(t)^2 dt}$.
Speech	Articulation loss of consonants [23]	$\text{AL}_{\text{cons}} = \left(\frac{200 D^2 T^2}{V} + a \right) \%$
Speech	Speech transmission index [24], [25]	$\text{STI} = \sum_{k=1}^{7} \alpha_k * \text{MTI}_k$
Music	Music clarity [26]	$C_{80} = 10 \log \frac{\int_{0}^{80\text{ ms}} h^2(t) dt}{\int_{80\text{ ms}}^{\infty} h^2(t) dt}$.
Music	Center time [27], [28]	$T_{\text{S}} = \frac{\int_{0\text{ms}}^{\infty} t * h(t)^2 dt}{\int_{0\text{ms}}^{\infty} h(t)^2 dt}$.

with SMAs. We first review existing omnidirectional metrics and then we select the ones based on ELERs that are suitable for being extended to their directional versions. The remainder of this paper is structured as follows. Section II reviews metrics extracted from omnidirectional RIRs. Section III presents the main proposal defining directional ELERs. Section IV details two case studies of directional ELERs: A) The simulation of a cuboid room, and B) The assessment of a large auditorium. Finally, section V states the conclusions.

II. METRICS FROM OMNIDIRECTIONAL RIRS

Table I overviews objective metrics of speech intelligibility and music clarity obtained from omnidirectional RIRs. By $h(t)$, we denote an omnidirectional RIR along time t.

Thiele [20] developed a room acoustical criterion for speech intelligibility referred to as definition (D_{50}); it is defined as the early-to-total energy ratio of h. An extension to D_{50} that includes the noise floor η is U_{50} [21], which is defined in a logarithmic scale.

The metric C_{50} also assess the influence of room acoustics in speech intelligibility [22]. In contrast to D_{50} and U_{50}, it is defined as an ELER of h. An alternative to C_{50} is the percentage of articulation loss of consonants (AL_{CONS}) [23]. However, in addition to h to estimate the reverberation time (T), AL_{CONS} requires geometric knowledge such as the critical distance (D) and the volume (V) of the room.

The speech transmission index (STI) is another metric of speech intelligibility [24]. It uses the modulation transfer index (MTI) to measure the change of the modulation in the signal over time, between the source and the listener, considering the reverberation time and the signal-to-noise ratio (SNR). STI considers a weighting factor α_k for each octave band indexed by k. Schroeder [25] proposed an alternative method to calculate STI from an impulse response, the so called indirect method.

Regarding metrics for music clarity, Reichardt [26] conducted subjective experiments to define C_{80} as an ELER of h. An alternative to C_{80} is the center time (T_S) proposed by Kürer [27], [28]. However, since T_S is defined as the centre of gravity of the room impulse response, a parameter strongly linked to the decay time, T_S is more correlated to reverberance rather than to intelligibility [3].

Among the energy ratios in Table I, C_{50} and C_{80} are well defined as ELERs; they are also linearly related to the subjective response in terms of useful-to-detrimental ratios [21]. This paper proposes the evaluation of these metrics on directional RIRs. Because the standard definitions of C_{50} and C_{80} only consider omnidirectional RIRs [2], the term directional ELER is used in this paper to avoid confusions.

III. DIRECTIONAL ELER

The directional features of sound have an important role in the decisions taken by acousticians to decide on the treatment of surfaces and the positioning of sources in acoustic spaces. Omnidirectional metrics provide general information that help with the use of materials but lacks to identify a desire position for those component. On the other hand, directional RIRs wit the use of ELERs can provide not only useful information on positioning but can indicate the directions where early reflections dominates among the late components, providing then practical guidance for clarity or intelligibility.

Figure 1 overviews the proposed assessment of directional ELERs. A SMA with a sparse number of microphones mounted on an acoustically rigid spherical baffle is used as a directional receiver to capture directional RIRs in the space under test. The spherical Fourier transform (SFT) [29] analyses the directional RIRs with the sparse grid where microphones are placed. A scattering removal filter operates in the SFT domain to compensate for the effects of having a rigid baffle,

providing a free-field representation on the surface of the spherical baffle [29]. The inverse spherical Fourier transform (ISFT) synthesizes RIRs for the directions Ω in the dense grid, yielding a dense set of directional RIRs $g(\Omega, t)$.

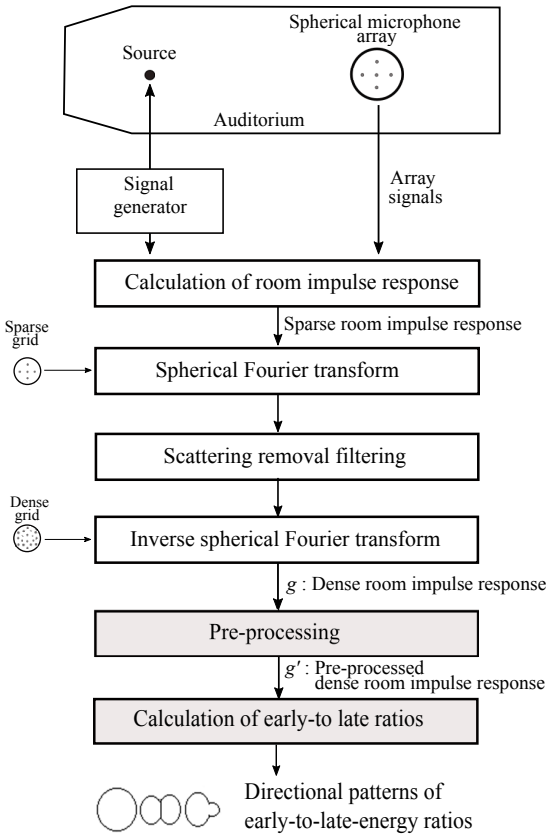

Fig. 1. Assessment of directional ELERs with a spherical microphone array.

A pre-processing stage takes $g(\Omega, t)$ as input to yield $g'(\Omega, t)$. This stage is independently applied along each direction Ω. Pre-processing comprises the steps below.

- *Spatial anti-aliasing low-pass filtering*. The low-pass filter cutoff frequency is

$$f_{\max} = \frac{cN}{2\pi r}, \tag{1}$$

  where c is the speed of sound in air, N is the order of the SFT, and r is the radius of the spherical array.
- *Normalization*. Along time, in each direction.
- *Onset-time truncation*. An amplitude threshold of -30 dB defines the onset time before which all samples are discarded.
- *Octave-band filter bank*. Each RIR is analyzed with octave-band filters, the center frequencies of which are the ones recommended by IEC [30].
- *Noise floor removal*. In each IEC frequency band, the noise floor is estimated from the last RIR samples and subsequently the noise floor power is subtracted from the same RIR.

Finally, a directional ELER is defined as

$$\mathrm{ELER}_\tau(\Omega) = \frac{\int_0^\tau g'(\Omega, t)^2 dt}{\int_\tau^\infty g'(\Omega, t)^2 dt}, \tag{2}$$

where τ denotes the early-to-late transition time. A transition time $\tau = 50$ ms is recommended for speech, defining ELER_{50}, whereas $\tau = 80$ ms is recommended for music, defining ELER_{80}. When ELERs are displayed in the logarithmic scale of decibels (dB), 0 dB indicates that the energies of the early and late components are equal. Positive values of ELER indicate that the early component is dominant, whereas negative values indicate that the late component is dominant.

IV. Case studies

We discuss two case studies: A) Simulations of ELERs in a cuboid room using an omnidirectional source and a rigid spherical array of 12 microphones ($N = 2$), and B) Assessment of ELERs in a large auditorium with a dipole source and a first-order ambisonic (FOA) microphone ($N = 1$).

A. Simulation of a cuboid room

A cuboid room was used first to numerically test the ELER values and their directional patterns. Using the Cox aspect ratios [31], [32] for smooth distribution of modes in a room, the following dimensions were chosen: 4.62 m wide (along x), 3.84 m long (along y), and 3 m high (along z). The algorithm in [33] was used to calculate RIRs for an omnidirectional sound source and a spherical array of 12 microphones mounted on a rigid spherical baffle of radius 8.5 cm. Diffuse-field equalization was also applied to the calculated RIRs using 12 additional sources surrounding the SMA with the algorithm in [33]. According to (1), the maximum frequency for the SMA radius 8.5 cm is $f_{\max} = 1285$ Hz. Figure 2 shows the top-view geometry of two cases: a source in the front of the array and a source in a rear corner.

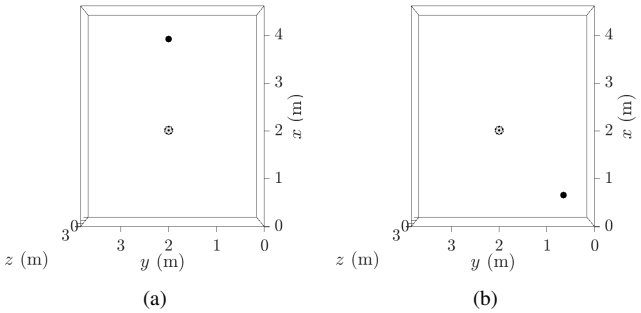

Fig. 2. Top-view of the cuboid rooms containing an omnidirectional source (black point) and a spherical array with 12 microphones near the center of the room. (a) Source at front. (b) Source at rear corner.

The algorithm in [33] is based on the image source method and includes high-order reflections from the walls, floor, and ceiling. All surfaces had a constant reflection coefficient β. For each geometry in Fig. 2, two reverberant conditions were tested: $\beta = 0.2$ and $\beta = 0.5$.

Figure 3 shows directional ELER patterns in a horizontal plane for the reverberant condition $\beta = 0.2$. Panels (a) and

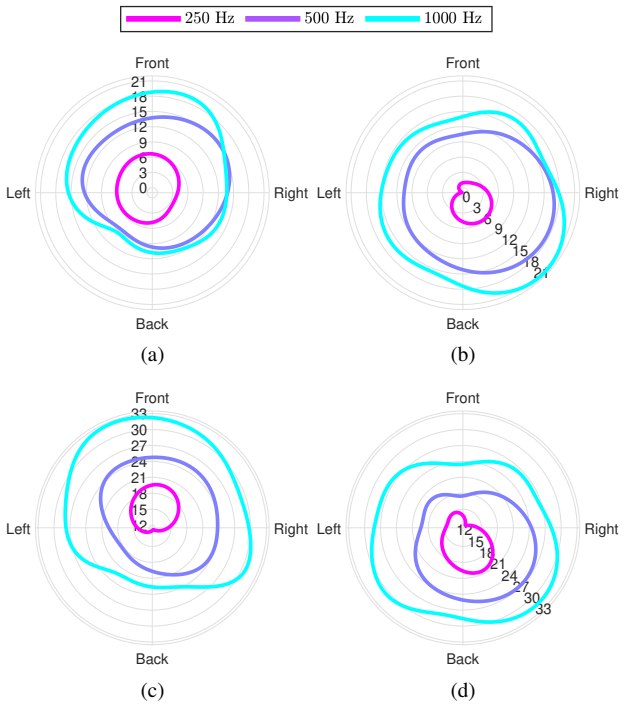

Fig. 3. Directional ELER patterns for the cuboid room. Reflection coefficient $\beta = 0.2$. (a) ELER_{50}, source at front. (b) ELER_{50}, source at rear corner. (c) ELER_{80}, source at front. (d) ELER_{80}, source at rear corner.

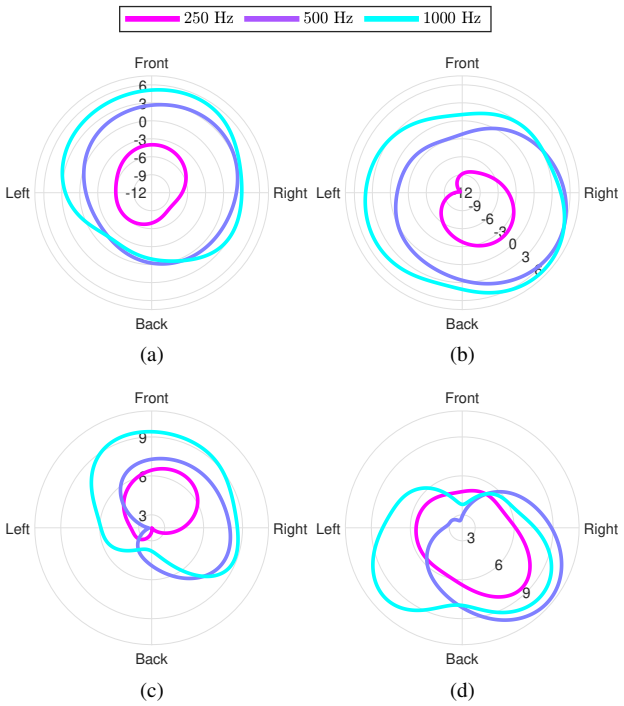

Fig. 4. Directional ELER patterns for the cuboid room. Reflection coefficient $\beta = 0.5$. (a) ELER_{50}, source at front. (b) ELER_{50}, source at rear corner. (c) ELER_{80}, source at front. (d) ELER_{80}, source at rear corner.

(b) show ELER_{50} patterns with values ranging from -1 dB to 22 dB. In all IEC frequency bands, ELER_{50} patterns present positive values with higher values pointing towards the source; this indicates that the early components are more dominant in the side of the source. The 250 Hz band presents the lower values, whereas the 500 Hz and 1000 Hz bands expand the contribution of early energy to all the directions. In panel (b), the 1000 Hz band expands the contribution of early energy to the lateral directions making these lobes more prominent; ELER_{50} patterns in this case tend to have higher values due to oblique reflections near the source position at the corner. Panels (c) and (d) show ELER_{80} patterns with values ranging from 11.5 dB to 33.5 dB. The energy of the early components is higher due to the larger transition time of 80 ms and, therefore, ELER_{80} yield larger values than ELER_{50}. In all IEC frequency bands, ELER_{80} patterns also have a lobe in the direction of the source. The 250 Hz band exposes higher values towards the source than those in the opposite direction. The ELER_{80} patterns for the 250 Hz band forms a cardiod-shaped pattern, whereas the 500 Hz and 1000 Hz bands expand the patterns into the lateral directions.

Table II shows dynamic ranges of ELERs for $\beta = 0.2$ shown in Fig 3. The 250 Hz band yields the lowest dynamic range along directions for ELER_{50}. A tendency exist for the 1000 Hz band showing the highest dynamic range for almost all cases in both ELER_{50} and ELER_{80}.

Figure 4 shows directional ELER patterns in a horizontal plane for the reverberant condition $\beta = 0.5$. This condition yields lower ELER values than the case $\beta = 0.2$ because the diffuse late reflections are more prominent. Panels (a) and (b) show ELER_{50} patterns with values ranging from -12 dB to 7.5 dB. In all IEC frequency bands, ELER_{50} patterns present a lobe with higher values pointing towards the source position that increase with increasing frequency; this indicates that the diffuse components are less dominant in the side of the source. In panel (a), the 500 Hz and 1000 Hz bands expand the

TABLE II
DYNAMIC RANGE OF ELER FOR $\beta = 0.2$

ELER	Frequency band	Source at front	Source at rear corner
ELER_{50}	250 Hz	3.15 dB	6.52 dB
	500 Hz	6.80 dB	7.23 dB
	1000 Hz	9.86 dB	8.02 dB
ELER_{80}	250 Hz	7.79 dB	8.23 dB
	500 Hz	6.74 dB	9.35 dB
	1000 Hz	10.42 dB	8.13 dB

TABLE III
DYNAMIC RANGE OF ELER FOR $\beta = 0.5$

ELER	Frequency band	Source at front	Source at rear corner
ELER_{50}	250 Hz	4.03 dB	9.65 dB
	500 Hz	5.26 dB	8.77 dB
	1000 Hz	7.34 dB	5.78 dB
ELER_{80}	250 Hz	4.75 dB	3.78 dB
	500 Hz	5.95 dB	8.15 dB
	1000 Hz	5.71 dB	6.23 dB

contribution of early energy to all the directions with slightly highers values of ELER_{50} coming from the right wall. In panel (b), the 250 Hz band yields a cardioid-shaped ELER_{50} pattern. The 500 Hz band expand the contribution of early energy from the lateral directions relative to the source at the corner. Panels (c) and (d) show ELER_{80} patterns with values ranging from 2 dB to 11 dB. The energy of the early components is higher due to the larger transition time of 80 ms and, hence, ELER_{80} yields larger values than ELER_{50}. In panel (c) ELER_{80} patterns expand into different directions; the 250 Hz and 500 Hz bands show a clear cardioid-shaped pattern towards the source and the right wall. In contrast, in panel (d), all patterns are more directive towards the source.

Table III shows dynamic ranges of ELERs for $\beta = 0.5$ shown in Fig 4. The 250 Hz band yields the lowest dynamic range along directions only for ELER_{50} when the source is at front. When the source is at the rear corner, the ELER_{50} dynamic range decreases when the frequency increases. In contrast to table II, the 1000 Hz band does not always posses the highest dynamic range.

When comparing panels (a) and (b) in Fig. 3 and Fig. 4, ELER_{50} shows similar patterns oriented towards the source. However, there are different ranges of ELER_{50} values in the two reverberant conditions; ELER_{50} values are higher in $\beta = 0.2$ than in $\beta = 0.5$ due to detrimental contributions of late components. The ELER_{80} patterns, are different in the two reverberant conditions.

B. Measurements in a large auditorium

Directional RIRs were captured at many seats in a large auditorium (volume 90000 m^3, reverberation time 5 s) using an FOA microphone and the SFT algorithms [29]. The auditorium did not have lateral walls. Considering a small sphere of radius 2.5 cm fully containing the four concentric microphones of the FOA array, the maximum frequency according to (1) is $f_{\max} = 2183$ Hz. However, ELERs have only been calculated in the octave bands of 250 Hz, 500 Hz, and 1000 Hz.

The source position was the stage scenario pointing towards the front (audience) and the back (concave hard wall). Two full range JBL model SRX815P loudspeakers were used as the sound sources, each with 2000 Watts peak of power rating. For the low frequency range a JBL sub woofer model SRX818SP with 1000 Watts peak of power rating was used.

Figure 5 overviews the assesment of directional ELERs using the FOA microphone. The signal generator emitted a logarithmic sweep tone of 11.5 s duration, reproduced over a dipole loudspeaker system in the stage. The FOA microphone (Sennheiser AMBEO) and the audio interface (Zoom F8) were used to capture the RIRs at the three receiver positions shown in Fig. 6, Fig. 7, and Fig. 9. The Zoom F8 built-in FOA encoder (SFT for cardioid microphones) was used to produce the FOA encoded RIRs. Subsequently, the SFT was used to extract RIRs for a dense number of directions in the horizontal and median planes of the FOA microphone. Pre-processing and calculation of ELERs from dense RIRs was performed according to Sec. III. Due to the highly reverberant conditions,

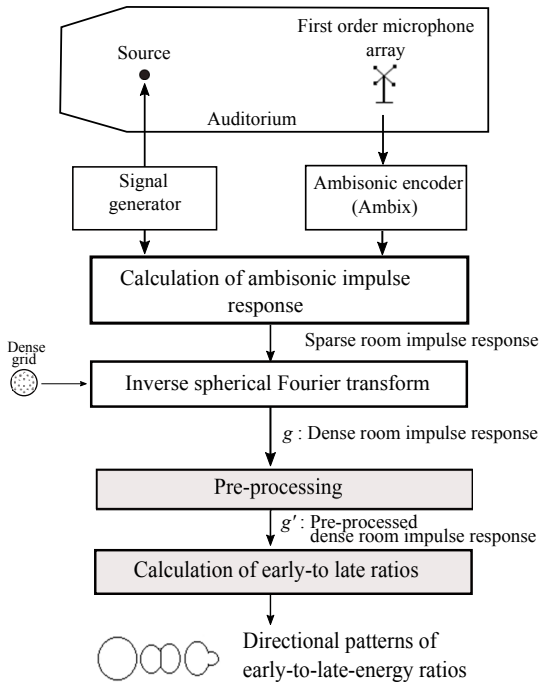

Fig. 5. Assessment of ELERs with a first-order ambisonic microphone.

most of ELER_{50} and ELER_{80} yielded negative values because the diffused field produced late reflections that were more dominant than the late reflections.

Figure 8 shows ELER patterns in the horizontal plane of receivers. Panels (a), (b), and (c) show ELER_{50} patterns with values ranging from -44 dB to 0 dB. Panels (d), (e), and (f) show ELER_{80} patterns with values ranging from -20 dB to 1.5 dB. In panel (a), ELER_{50} at 250 Hz shows that the late components are less dominant in the direction of the source because the values are less negative; a small lobe in the opposite side of the source appears due to the presence of a rear wall, whose early reflections reduces the effect of the overall late reflections. In panels (b) and (c), ELER_{50} at 500 Hz becomes more directive in the direction of the source in the stage. In panels (a), (b), and (c), ELER_{50} patterns at 1000 Hz, in contrast, tend to be omnidirectional. In panel (d), ELER_{80} values in all bands tend to be a dipole pattern. In panels (e) and (f) ELER_{80} values tend to be more directive towards the source.

Figure 10 shows ELER patterns in the median plane of receivers with the same scales used in Fig. 10. Panels (a), (b), and (c) show ELER_{50}. Panels (d), (e), and (f) show ELER_{80} patterns. In panel (a), ELER_{50} at 250 Hz shows that the late components are less dominant in the direction of the source; a small lobe in the opposite side of the source appears due to dominant early reflections. In panels (b) and (c), ELER_{50} at 500 Hz becomes more directive in the direction of the source in the stage. In panels (a), (b), and (c), ELER_{50} patterns at 1000 Hz, in contrast, tend to be omnidirectional. In panel (d), ELER_{80} values in all bands tend to be a dipole pattern. In panels (e) and (f) ELER_{80} values tend to be more directive

towards the source.

In all cases, when moving away from the source, ELER_{50} and ELER_{80} patterns at 500 Hz present a more dominant diffuse component. When contrasting the assessment in the auditorium with the cuboid room, it can be observed that the FOA microphone was not as selective as the second-order SMA. Despite this, in Fig. 10, panels (a) and (b), the 250 Hz band shows that the FOA microphone could manage to capture a reflection from the floor. As in the top view, the 500 Hz band shows an improvement in the way the useful energy is related to the source direction.

(a)

(b)

(c)

Fig. 6. Panoramic view from receivers. (a) Receiver 1. (b) Receiver 2. (c) Receiver 3.

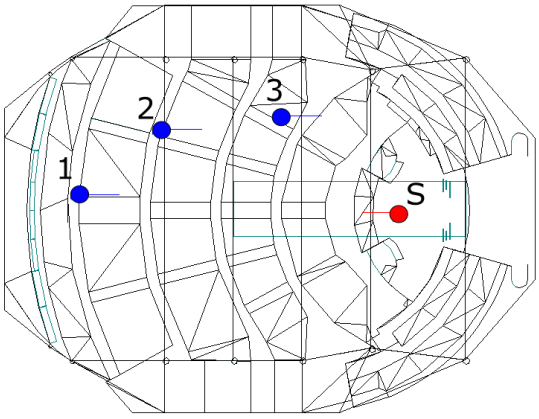

Fig. 7. Schematic top view of the auditorium. The blue dots indicate the receiver positions. The red dot indicate the source position.

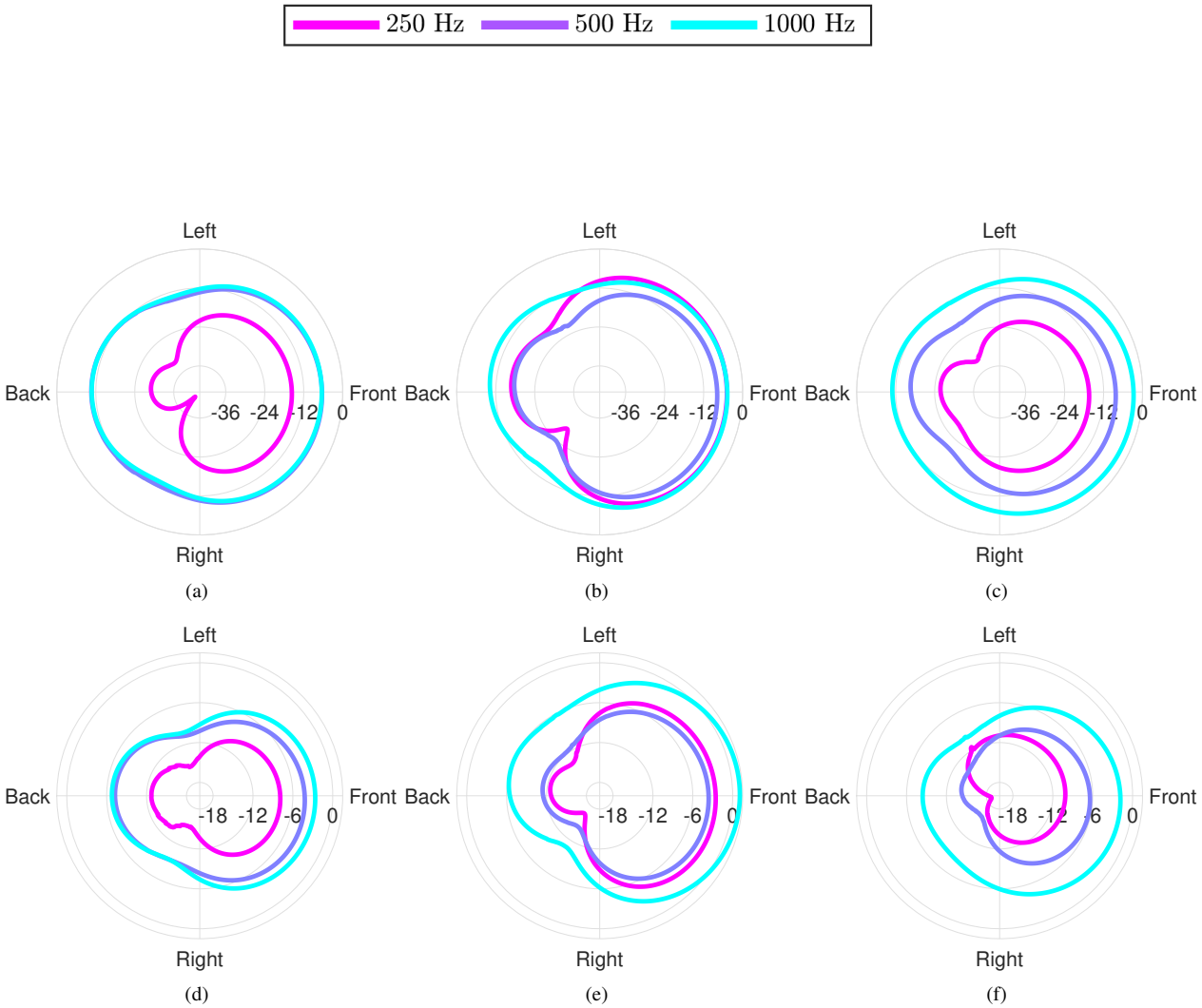

Fig. 8. Directional ELER patterns in a horizontal plane. (a) $ELER_{50}$ at receiver 1. (b) $ELER_{50}$ at receiver 2. (c) $ELER_{50}$ at receiver 3. (d) $ELER_{80}$ at receiver 1. (e) $ELER_{80}$ at receiver 2. (f) $ELER_{80}$ at receiver 3.

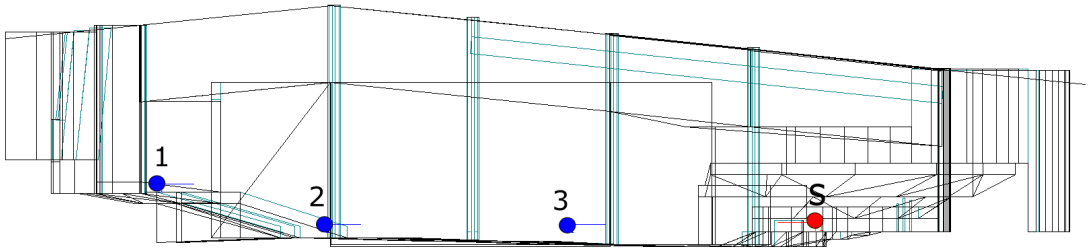

Fig. 9. Schematic lateral view of the auditorium. The blue dots indicate the receiver positions. The red dot indicate the source position.

Fig. 10. Directional ELER patterns in a median plane. (a) $ELER_{50}$ at receiver 1. (b) $ELER_{50}$ at receiver 2. (c) $ELER_{50}$ at receiver 3. (d) $ELER_{80}$ at receiver 1. (e) $ELER_{80}$ at receiver 2. (f) $ELER_{80}$ at receiver 3.

V. Conclusion

The importance of directional ELERs has been addressed in this paper. The current omnidirectional metrics of intelligibility C_{50} and C_{80} were extended to their directional versions. When comparing with these omnidirectional metrics, we found that $ELER_{50}$ and $ELER_{80}$ are also highly dependent on the receiver position since each position in the room is affected by specific reflections.

Directional ELERs were assessed in two situations: a cuboid room and a large auditorium. Directional patterns were identified in both cases. The results highlighted two directional ELER patterns that might correlate with subjective speech intelligibility and music clarity: an omnidirectional pattern and a beam pointing towards the source in the stage. A third pattern with a beam towards the source and a tail with a small pattern in the back was also found in areas where strong reflections from the back wall are present. Further discussion for the correct association of each pattern to the enhancement or detriment of intelligibility and clarity is needed.

Extensions to this work might consider binaural room impulse responses [34] and the paradigm of spatial release from masking to include a target direction Ω_{target} in the numerator of (2) and a masker direction Ω_{masker} in the denominator. This approach would allow to examine speech intelligibility in terms of energy ratios of a useful target from a specific direction and detrimental diffuse maskers from other directions. Furthermore, the inclusion of directional sources simulating the pattern of musical instruments can be also consider as a future line of investigation. Perceptual evaluations by means of detectability of differences, and localization tests along angles, could also provide more insight into the validity of the suggested approach.

Acknowledgment

The authors wish to thank Ayrton Urviola for fruitful discussions.

References

[1] N. Xiang, *Architectural Acoustics Handbook*, ser. Acoustic Series. J. Ross Publishing, 2017.

[2] ISO, *ISO 3382-1:2009 Acoustics — Measurement of room acoustic parameters — Part 1: Performance spaces*, 2009.

[3] J. S. Bradley, "Review of objective room acoustics measures and future needs," *Appl. Acoust.*, vol. 72, no. 10, pp. 713–720, 2011.

[4] T. J. Cox, W. J. Davies, and Y. W. Lam, "The sensitivity of listeners to early sound field changes in auditoria," *Acta Acust. United Ac.*, vol. 79, no. 1, pp. 27–41, 1993.

[5] M. C. Vigeant, R. D. Celmer, C. M. Jasinski, M. J. Ahearn, M. J. Schaeffler, C. B. Giacomoni, A. P. Wells, and C. I. Ormsbee, "The effects of different test methods on the just noticeable difference of clarity index for music," *J. Acoust. Soc. Am.*, vol. 138, no. 1, pp. 476–491, 2015.

[6] J. S. Bradley, R. Reich, and S. G. Norcross, "A just noticeable difference in C50 for speech," *Applied Acoustics*, vol. 58, no. 2, pp. 99–108, 1999.

[7] W. Lachenmayr, A. Haapaniemi, and T. Lokki, "Direction of late reverberation and envelopment in two reproduced Berlin concert halls," in *Proc. 140th Audio Eng. Soc. Convention*, Paris, France, Jun. 2016.

[8] D. Romblom, C. Guastavino, and P. Depalle, "Perceptual thresholds for non-ideal diffuse field reverberation," *J. Acoust. Soc. Am.*, vol. 140, no. 5, pp. 3908–3916, 2016.

[9] B. Alary, P. Massé, S. J. Schlecht, M. Noisternig, and V. Välimäki, "Perceptual analysis of directional late reverberation," *J. Acoust. Soc. Am.*, vol. 149, no. 5, pp. 3189–3199, 2021.

[10] R. Y. Litovsky, "Spatial release from masking," *Acoustics today*, vol. 8, no. 2, pp. 18–25, 2012.

[11] S. J. van Wijngaarden and R. Drullman, "Binaural intelligibility prediction based on the speech transmission index," *The Journal of the Acoustical Society of America*, vol. 123, no. 6, pp. 4514–4523, 2008.

[12] T. Leclère, M. Lavandier, and J. F. Culling, "Speech intelligibility prediction in reverberation: Towards an integrated model of speech transmission, spatial unmasking, and binaural de-reverberation," *J. Acoust. Soc. Am.*, vol. 137, no. 6, pp. 3335–3345, 2015.

[13] D. Khaykin and B. Rafaely, "Acoustic analysis by spherical microphone array processing of room impulse responses," *J. Acoust. Soc. Am.*, vol. 132, no. 1, pp. 261–270, 2012.

[14] J. Pätynen, S. Tervo, P. W. Robinson, and T. Lokki, "Concert halls with strong lateral reflections enhance musical dynamics," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 111, no. 12, pp. 4409–4414, 2014.

[15] N. Raghuvanshi, R. Narain, and M. C. Lin, "Efficient and accurate sound propagation using adaptive rectangular decomposition," *IEEE Trans. Vis. Comput. Graphics*, vol. 15, no. 5, pp. 789–801, Sep. 2009.

[16] J. Shi, C. D. Salvador, J. Treviño, S. Sakamoto, and Y. Suzuki, "Spherical harmonic representation of rectangular domain sound fields," *Acoust. Sci. Technol.*, vol. 41, no. 1, pp. 451–453, 2020.

[17] S. Tervo and A. Politis, "Direction of arrival estimation of reflections from room impulse responses using a spherical microphone array," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 10, pp. 1539–1551, Oct. 2015.

[18] S. V. Amengual Garí, W. Lachenmayr, and E. Mommertz, "Spatial analysis and auralization of room acoustics using a tetrahedral microphone," *J. Acoust. Soc. Am.*, vol. 141, no. 4, pp. EL369–EL374, 2017.

[19] D. A. Dick and M. C. Vigeant, "An investigation of listener envelopment utilizing a spherical microphone array and third-order ambisonics reproduction," *J. Acoust. Soc. Am.*, vol. 145, no. 4, pp. 2795–2809, 2019.

[20] R. Thiele, "Richtungsverteilung und Zeitfolge der Schallrückwürfe in Räumen," *Acta Acust. United Ac.*, vol. 3, no. 4, pp. 291–302, 1953.

[21] J. S. Bradley, R. D. Reich, and S. G. Norcross, "On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility," *J. Acoust. Soc. Am.*, vol. 106, no. 4, pp. 1820–1828, 1999.

[22] J. S. Bradley, "Predictors of speech intelligibility in rooms," *J. Acoust. Soc. Am.*, vol. 80, no. 3, pp. 837–845, 1986.

[23] V. M. A. Peutz, "Articulation loss of consonants as a criterion for speech transmission in a room," *J. Audio Eng. Soc*, vol. 19, no. 11, pp. 915–919, 1971.

[24] IEC, *IEC 60268-16:2020 Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index*, 5th ed. IEC, 2020.

[25] M. R. Schroeder, "Modulation transfer functions: Definition and measurement," *Acta Acust. United Ac.*, vol. 49, no. 3, pp. 179–182, 1981.

[26] W. Reichardt, O. A. Alim, and W. Schmidt, "Definition and basis of making an objective evaluation to distinguish between useful and useless clarity defining musical performances," *Acta Acust. United Ac.*, vol. 32, no. 3, pp. 126–137, 1975.

[27] R. Kürer, "Einfaches messverfahren zur bestimmung der schwerpunktzeit raumakustischer impulsantworten (simple measuring procedure for determining the center time of room acoustical impulse responses)," in *AES 116 Convention*, Budapest, Hungary, May 1971.

[28] J. Segura, A. Gimenez, J. Romero, and S. Cerda, "A comparison of different techniques for simulating and measuring acoustic parameters in a place of worship: Sant jaume basilica in valencia, spain," *Acta Acustica united with Acustica*, vol. 97, no. 1, pp. 155–170, Jan. 2011.

[29] C. D. Salvador, S. Sakamoto, J. Treviño, and Y. Suzuki, "Boundary matching filters for spherical microphone and loudspeaker arrays," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 3, pp. 461–474, Mar. 2018.

[30] IEC, *IEC 61260-1:2014 Electroacoustics - Octave-band and fractional-octave-band filters - Part 1: Specifications*, 1st ed. IEC, 2014.

[31] T. J. Cox, P. D'Antonio, and M. R. Avis, "Room sizing and optimization at low frequencies," *J. Audio Eng. Soc*, vol. 52, no. 6, pp. 640–651, 2004.

[32] J. H. Rindel, "Preferred dimension ratios of small rectangular rooms," *JASA-EL*, vol. 1, no. 2, p. 021601, 2021.

[33] D. P. Jarret, E. A. P. Habets, M. R. P. Thomas, and P. A. Naylor, "Rigid sphere room impulse response simulation: algorithm and applications," *J. Acoust. Soc. Am.*, vol. 132, no. 3, pp. 1462–1472, Sep. 2012.

[34] W. Zhang, P. N. Samarasinghe, and T. D. Abhayapala, "Parameterization of the binaural room transfer function using modal decomposition," *J. Acoust. Soc. Am.*, vol. 146, no. 1, pp. EL8–EL14, 2019.