

Perceptual Evaluation of Dynamic Sound Zones, Part 2: Audio Quality of Experience in a Simulated Hospital Environment

LINE HERMANNSEN,^{1,2,*}  KASPER FANGEL SKOV,²  AND SØREN BECH,^{1,3}  AES Fellow
(line.hermannsen@gmail.com) (kfsk@ucsyd.dk) (sbe@es.aau.dk)

¹*B&O Research, Bang & Olufsen A/S, Struer, Denmark*

²*Department of Computer Science, Aalborg University, Aalborg, Denmark*

³*Department of Electronic Systems, Aalborg University, Aalborg, Denmark*

In the Interactive Sound Zones for Better Living (ISOBEL) project, a sound zone system was developed to create two personal listening experiences simultaneously in health care settings. The perception and audio quality of experience of sound zones were evaluated through listening tests, where perceptual attributes were rated in different scenarios with variations in reproduction mode, audio bandwidth, and reproduction levels. The tests were conducted in a simulated hospital setting to improve the ecological validity of the findings. All perceptual attributes were found to be significantly influenced by one or more of the selected physical factors. In audio-on-audio scenarios, the use of sound zones significantly decreased the perceived distraction and increased the liking compared with wide-angle reproduction. The influence of reproduction levels and audio bandwidth depended on the use case and audio content. Distraction was found to be the main predictor of the audio quality of experience of sound zones. The results help refine the experimental methodology for future perceptual tests within the project. This study provides a valuable contribution to understanding how to enhance listening experiences simultaneously for multiple users.

Keywords: auditory masking, beamforming, psychoacoustics, spatial sound field control, subjective listening tests

1 INTRODUCTION

1.1 Background

Noise is an environmental stressor that can have negative impacts on human health and well-being [1]. In health care environments that should facilitate healing, rest, and recovery, noise can be ubiquitous and may lead to stress, poor sleep quality, delirium, and/or cardiovascular effects [2–4]. Hospital alarms and medical equipment emit frequent high-level sounds that are among the most disturbing noises in intensive care units (ICUs) [5]. Yet, sounds associated with humans may be the main noise contributors in hospitals, with staff sounds being found to account for 57% of the acoustic energy in ICU patient rooms [6] and sounds from other patients being rated the topmost disturbing noise [5].

In contrast to noise, music and natural sounds can have therapeutic effects that can alleviate stress, decrease anxiety, reduce heart rate, and improve sleep and thereby help the healing process [7].

In shared hospital rooms where multiple patients may each desire their own personal listening experience, headphones are an easy solution to reduce noise exposure and improve the individual acoustic environment. However, the resulting auditory isolation and reduced environmental awareness may be undesirable and can also lead to negative effects. Sound mediates the relation of the listener with the environment [8] because auditory scene analysis is continuously conducted to sense the surroundings and detect potential signals of interest. Acoustic communication can help patients avoid harmful isolation and loneliness [9], and to “escape the situation,” which can create meaning and hope [10]. Patients in an unconscious state can also have active hearing and respond actively or neurally to acoustic signals [11]. Furthermore, the health state of a patient may call for nonintrusive equipment that does not cause fatigue and discomfort, risk transferring infections, or reduce the ability for effective and critical communication with hospital staff. Even in single-bed hospital rooms the presence of

*To whom correspondence should be addressed, e-mail: line.hermannsen@gmail.com

patient, staff, and relatives may call for solutions that can create multiple audio streams while avoiding isolation.

A potential solution for improving shared spaces at hospitals is to create personal listening experiences in dynamic sound zones that are less isolating and intrusive than headphones. Sound zones can be created from an array of speakers by beam-forming of multiple audio streams at the same time toward different positions in a room [12]. This enables audio-on-audio scenarios with improved target-to-interferer ratio (TIR)¹ compared with wide-angle scenarios and potential masking of unwanted sounds. By directing sound to the intended users, transmission to the surrounding environment can also be minimized. The use of sound zones in health care environments may thereby not only benefit patients but also hospital staff that are exposed daily to various noise sources with potential negative effects on their well-being and performance [3, 4].

In the Interactive Sound Zones for Better Living (ISOBEL) project, a dynamic sound zone system was developed for use in health care environments. The overall aim was to create sound zones that can adapt in real-time 1) to variations in the acoustic environment with fluctuations in ambient noise from alarms and other hospital equipment and 2) to changes in patient needs through biological feedback mechanisms where patient physiological responses (e.g., heart rate variability) and sleep quality are continuously monitored. The first step to achieve an effective system with optimization of sound zones to specific contexts and needs is to understand how sound zones are perceived in a health care setting, which is the aim of this study.

This study (part 2) builds on previous work (part 0 and 1) [13, 14] to identify and test the physical factors and perceptual dimensions that influence audio quality of experience (QoE) of sound zones in domestic and health care environments, with focus on the confirmed trade-off between bass enhancement and sound zone separation [14]. An overview of all four studies (part 0–3) within the ISOBEL project focusing on QoE of sound zones can be found in part 1 [14]. Part 2 was designed to evaluate various use cases within a hospital environment to best prepare and optimize the experimental methodology for later field studies and also to obtain perceptual results as close to expected field results as possible.² SEC. 2 of this paper explains the methodology of the listening tests and preliminary work conducted to obtain perceptual data. SEC. 3 presents the results, and the primary findings are discussed in SEC. 4. The last section, SEC. 5, concludes on the findings.

1.2 Main Research Questions

In the present study, a sound zone system was used to either 1) create two personal sound zones or 2) create two

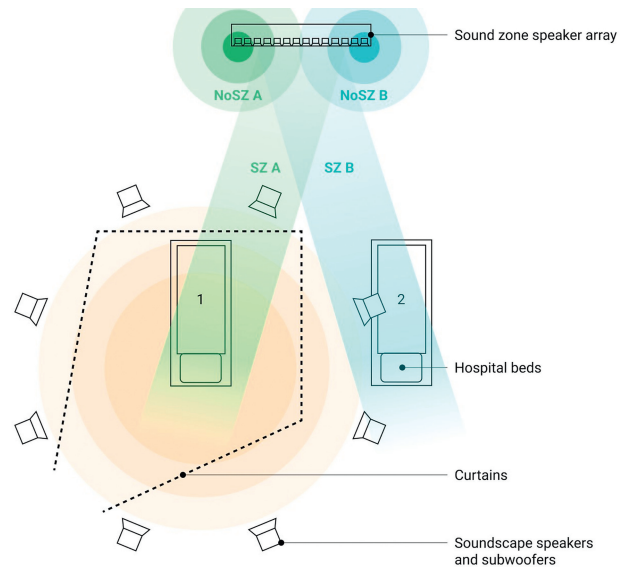


Fig. 1. The experimental setup with the health care system emitting two sound zones (SZ A and SZ B) or two wide-angle audio streams from each end of the systems (NoSZ A and NoSZ B). The assessor was placed in SZ A surrounded by acoustically transparent curtains that covered the laboratory equipment to make the settings more realistic. A hospital soundscape was simulated by an eight-loudspeaker and eight-subwoofer setup placed around the hospital bed.

wide-angle audio streams from each end of a custom-built health care system in a simulated hospital environment. Sound zone A (SZ A) was directed toward the assessor placed in hospital bed 1, and sound zone B (SZ B) was directed toward hospital bed 2 to simulate the presence of another patient (Fig. 1). The primary research questions were as follows: 1) Does audio reproduction in sound zones improve the perceived audio QoE during audio-on-audio scenarios compared with reproduction of wide-angle audio? 2) How do changes in TIR and audio content (related to different use cases) influence perception of and affection for stimuli? 3) Does removal of low-frequency audio components reproduced in sound zones decrease the audio QoE because of poorer bass enhancement, and/or does it improve the listening experience by optimizing the separation of zones?

2 METHODS

2.1 Laboratory Settings

This study was conducted in a sound-proof laboratory³ at Aalborg University, Denmark, from December 2021 to February 2022. The experimental settings simulated a hospital environment with hospital beds centered in a section of the laboratory that was visually shielded

¹ The TIR is defined as the difference between the target and interferer levels (e.g., LAeq [10 s], which is the A-weighted equivalent continuous sound level measured over a 10-second time interval) of the target and interferer signals measured in the same zone.

² In case it was not possible to access a real hospital environment, for example, because of COVID-19.

³ The laboratory has dimensions of 7.0 m (width) × 8.12 m (length) × 3.0 m (height) and a reverberation time (T60) within the International Telecommunication Union Recommendations (ITU-R) BS.1116 targets except that it is 0.2 s in the frequency range 0.5–6.3 kHz, which is below the lower limit (0.25 s) of ITU-R BS.1116.

from technical equipment by acoustically transparent curtains (Fig. 1). Light levels and room temperatures were monitored to ensure that the potential influence on the study results would be similar across all listening tests. Light levels were measured at the headrest of the hospital bed with a handheld Elma CA1110 luxmeter and kept constant (15–21 lux) with dense window curtains and artificial lights. The room temperature was monitored continuously with a basic thermometer placed below the hospital bed during data collection and varied by 2.3°C (18.9–21.2°C).

2.2 The Simulated Hospital Soundscape

To obtain subjective evaluations of sound zones closer to real sound zone experiences in a hospital, a simulated hospital soundscape was added to the study setup. The hospital soundscape simulation was based on recordings from an ICU at North Zealand Hospital, Denmark, and was also used in another experiment within the ISOBEL project [15]. The recordings were modified to include a range of different sound sources, including low and high-intensity alarms (with peak frequencies at 1–3 kHz), equipment beeps, continuous positive airway pressure machinery, and staff conversation within a 4-min clip, and generated in three versions with a randomized order of the sound sources. The three audio versions were then combined to one 12-min-long ambiance soundscape file (excerpt available at Zenodo [16]), which was considered sufficiently long to avoid assessors remembering the order of the sounds and thereby minimizing their degree of sensitization or habituation to the soundscape.

The soundscape file was rendered as a 16-channel Ambisonics file and played back on eight stand-mounted loudspeakers (Dynaudio BM5, 1.1 m above the floor) and eight custom floor-standing subwoofers (10-in units from Tympany) placed in a circle around and directed toward the headrest of hospital bed 1 (Fig. 1). Speakers and subwoofers were placed 2.2–2.4 m from the center of the headrest. The signals sent to the subwoofers were low-pass filtered with a first-order filter at 150 Hz applied in the mixer (RME TotalMix FX) to minimize distortion resulting from reproduction of higher-frequency sounds.

The reproduction level of the soundscape was adjusted to match realistic hospital levels by comparing A-weighted peak-peak levels of distinctive alarm signals between laboratory and actual hospital recordings to reproduce low-intensity alarms at a peak-to-peak level of 63 dB-SPL at the listening position. The low-intensity alarms were played multiple times in the soundscape with 1–9 transient pulses following each other within a 1–17-s window, whereas the high-intensity alarms were only played once during the 12-min-long ambiance file, containing 11 transient pulses within an 11-s window.

2.3 The Sound Zone System

In the ISOBEL project, two different systems were developed to create sound zones; one for domestic environments (termed domestic system) and another for health

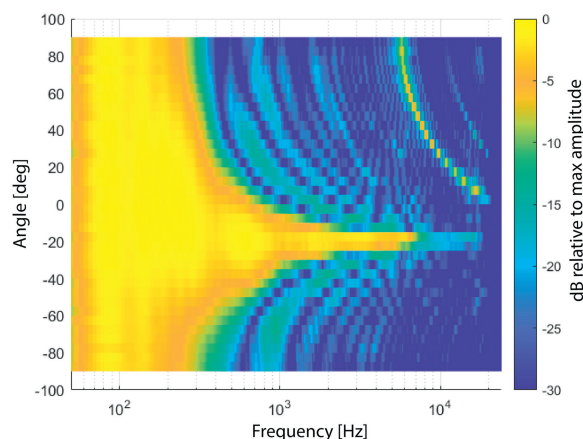


Fig. 2. The simulated free field directivity responses of the simulated health care system when reproducing audio in SZ A. Figure courtesy of Martin Bo Møller, Bang & Olufsen A/S.

care environments (termed health care system). The differences between the two systems were mainly driven by the restrictions in health care environments requiring compact and hygienic equipment, whereas domestic environments generally allow for larger system widths and thereby a better ability to separate lower-frequency components, which can improve the sound zone separation over a larger bandwidth.

Because the sound zone system intended for tests in health care settings (SoundFocus ApS, ISOBEL project partner) was still under development during the experimental period, the domestic sound zone prototype (see Fig. 2 in [14]) was used with characteristics simulating the planned health care prototype. The domestic system simulating the health care system, therefore, had a reduced number of drivers active for this experiment, which included twenty of the 1.5-in drivers and the two most central 5-in drivers, to approximate the sixteen 2-in drivers of the health care prototype. The domestic system had previously been used for other listening tests within the ISOBEL project [13–15].

The sound zone system had two audio modes: 1) wide-angle audio (i.e., no sound zones [NoSZ]) and 2) audio in sound zones (SZ) (Fig. 1) to evaluate how sound zones can influence perception and improve listening experiences. In scenarios with wide-angle audio, the system simulated two distinct sound sources by activating only two drivers (one 1.5-in and one 5-in driver) on each end of the system and reproduced sound at an angle of ± 90 degrees. The decision to use the simulated health care system to simulate a two-way systems instead of using two separate loudspeakers was to ensure that the loudspeaker driver characteristics were the same throughout the tests and thereby did not influence the perception. The sound sources were placed 1.97 m apart center-to-center. In the sound zone scenarios, the system created two sound zones (SZ A and SZ B; Fig. 1) at angles of $\pm 20^\circ$ relative to the center of the system using delay and sum beamforming (Fig. 2).

Table 1. Audio information for audio material used in the listening tests.

Audio name	Audio information <i>Artist and track name, time in track (h:min:s)</i>
Calm music 1	Wavecare— <i>ICU</i> , 00:00:00–00:01:00
Calm music 2	Brian Eno— <i>An Ending (Ascent)</i> , 00:00:45–00:01:48
Calm music 3	Wavecare— <i>Piano</i> , 00:00:00–00:01:01
Pop rock music 1	Kh Marie featuring Bisse— <i>Ryggem Mod Muren</i> , 00:02:04–00:03:25
Pop rock music 2*	The Killers— <i>On Top</i> , 00:01:53–00:02:53
Podcast	DR— <i>Om Hundrede Aar</i> , del 1, 1:5, 00:05:26–00:06:19
Radio Sports*	DR— <i>Liga</i> , June 21 2020, 02:25:38–02:26:38

*Tested in a previous experiment [14].

DR = Danish Broadcasting Corporation.

2.4 Audio Stimuli

2.4.1 Audio Content

Because sound perception is influenced by the intention of listening [17], audio stimuli within three different listening tasks were in focus: rest, entertainment, and information gathering. These use cases were also used and evaluated in two previous experiments [13, 14]. The scenarios were found suitable to cover the two environments planned for field tests: an ICU and a cardiology department in Denmark. The rest scenarios were intended to represent use cases within both environments but focused on the ICU, where patients are in a poor state and a calm and quiet environment is desired. The entertainment and information gathering scenarios were intended to primarily represent use cases at the cardiology department, where patients are awake and may desire recreation and entertainment through listening to music and/or gather new information in the form of radio news, podcasts, or similar speech signals. All audio information can be found in Table 1.

At the listening position (hospital bed 1; Fig. 1), two pieces of calm instrumental music composed by Wavecare (ISOBEL project partner) were used as target stimuli for the rest scenarios. One piece was an ambiance track (*calm music 1*; Table 1), and the other piece (*calm music 3*) was based on the soundscape at an ICU, with the intention of using spectral masking⁴ [18] of unwanted sounds, including hospital alarms. In the user experience study by Skov et al. [15], which used a similar experimental setup as this present study, the track *calm music 3* had also been used for sound zone scenarios in a simulated hospital soundscape and was found to result in higher experienced calmness and reassurance.

⁴ Spectral masking occurs when a sound with a certain spectral content makes the detection of another sound with different spectral content harder.

For entertainment, a music piece with Danish female and male vocals and broadband components that overlap with the simulated hospital soundscape were chosen, whereas a podcast in Danish with a male narrator was chosen for information gathering. Target stimuli were focused on Danish, the mother tongue of the assessors, to create experiences that were not influenced by the ability to speak and understand a second language. Toward hospital bed 2 (Fig. 1), three different stimuli representing interfering audio for the assessor were reproduced: 1) pop/rock vocal music with English lyrics (*pop/rock music 2*); 2) radio sports commentary in Danish (*radio sports*); and 3) instrumental calm music (*calm music 2*). Silence was used as a control to assess the perception of attributes without the interferer. All stimuli were 1-min long and played back at a 44.1-kHz sampling rate.

2.4.2 Optimal Loudness

A small experiment was conducted before the listening tests to find the optimal loudness of the target stimuli while being exposed to the simulated hospital soundscape. Seven assessors (age 19–33, four women and three men) participated in the optimal loudness test. All stimuli were initially loudness-matched according to the International Telecommunication Union Recommendation (ITU-R) BS.1770-5 Standard [19] to a sound pressure level of 66 dB (unweighted and without the soundscape), which was evaluated as a suitable starting point for the listening levels by the experimenter.

The stimuli were played back through a Max/MSP patch (Cycling '74), which contained a user interface for adjusting gain. The interface was broadcasted through Mira (Cycling '74) to a tablet on which the assessors could adjust the loudness according to personal preferences. The assessors were instructed to set an optimal loudness that they could listen to for a prolonged period, as if they were a patient in a hospital. The interface in the first round of assessors was a simple box with displayed values and a vertical slider. In the second round the slider was shown without displayed values. The difference in interface did not impact selected loudness significantly ($p < 0.1$; Kruskal-Wallis test in R Studio, version 1.4.1103) and all data could therefore be compiled.

In the optimal loudness test, assessors first conducted a familiarization round where no data were collected, followed by evaluation of the four target stimuli in each of three versions (later used in the listening tests): 1. full bandwidth audio in sound zones, 2. band-pass filtered (4th order Butterworth high-pass filter at 150 Hz and low-pass filter at 10 kHz) audio in sound zones, and 3. wide-angle audio reproduction. The twelve stimuli were presented in a randomized order and a hidden repeat was added to test the consistency of assessors.

The results showed that the selected loudness was not significantly different across assessors ($p = 0.05$, Kruskal-Wallis test), and the overall perception of loudness was not significantly different across the 12 stimuli ($p = 0.19$), indicating that loudness-matching of stimuli before the exper-

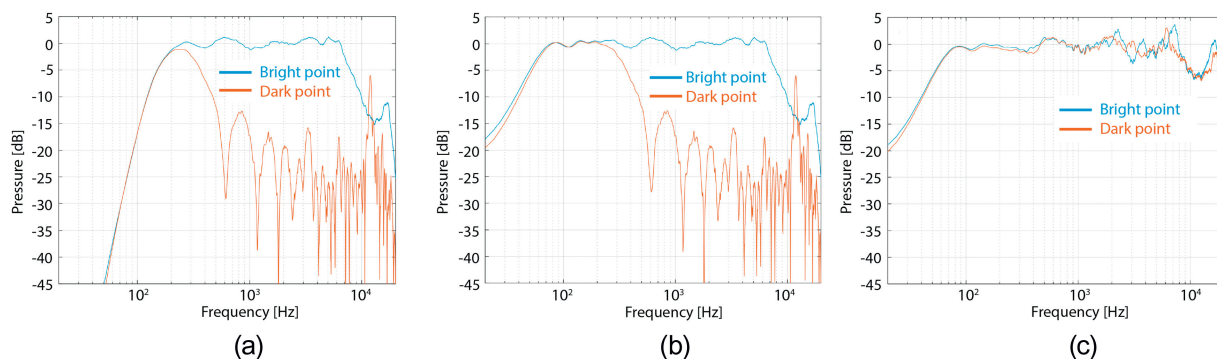


Fig. 3. Reproduction of audio in sound zones bandpass filtered (a) or full bandwidth (b) and in wide-angle scenarios (c). The plots show the amplitude of the target audio (bright point) in contrast to the amplitude of the interference audio (dark point). Figure courtesy of Martin Bo Møller, Bang & Olufsen A/S.

iment was effective. When comparing perceived loudness for two identical stimuli (hidden repeat), assessors were generally consistent in their ratings (mean = 2.3 dB). The mean of selected loudness across all stimuli and assessors was -10.9 dB (95% CI, -12.0 to -9.8) relative to the original reproduction level resulting in a overall selected reproduction level of 55 dB-SPL (unweighted) at the listening position. The reproduction level in the listening test was therefore adjusted to match this level.

2.4.3 Room Correction

Room impulse responses were measured for each reproduction mode (see SEC. 2.5.1) and filter setting (see SEC. 2.5.3) of the sound system to enable room correction of the audio reproduction. This was done using logarithmic sweeps following the method in [20]. All target and interference audio stimuli were then corrected with the measured impulse responses in a custom-written MATLAB script before use in the listening test.

2.5 Changes in Physical Factors

2.5.1 Reproduction Mode

To better understand how sound zones can improve the listening experiences of a listener, sound zone scenarios should be evaluated and compared with wide-angle audio scenarios, which is the normal listening mode. Wide-angle audio was included in a previous study within the ISOBEL project [14] investigating sound zone perception, but in that study it was not possible to evaluate the benefits of sound zones in audio-on-audio scenarios because wide-angle audio was only included in scenarios without interfering audio. Here, audio-on-audio stimuli were presented in two reproduction modes: 1) wide-angle audio (NoSZ) and 2) audio reproduced in sound zones (SZ) (Figs. 1 and 3).

2.5.2 Gain

In both reproduction modes, the influence of gain was assessed by having two reproduction levels of the target audio either with 1) 0-dB gain (“unmodified,” see SEC. 2.4.2) or 2) an increase by 6 dB. This modification was also investigated in the previous experiment [14], which confirmed that a 6-dB increase in gain is perceivable and positively

influences liking within the selected optimal loudness of stimuli. Here, it was included again to evaluate how the TIR influences ratings when noise is a combination of interfering audio and a hospital soundscape.

2.5.3 Filter

For sound zone scenarios, the effects of filtering the audio was investigated by evaluating the target audio in either 1) full bandwidth or 2) band-pass filtered, where low-frequency (bass) components were digitally removed (fourth-order Butterworth filter at 150 Hz; Fig. 3). This was to assess the effects of low-frequency (bass) components on the trade-off between audio quality and sound zone separation. In a previous experiment [14], a full subwoofer setup was added to the domestic system reproduction and the perceived bass strength was found to be positively correlated with liking. In the present study, a subwoofer setup was not included because this study was focused on health care environments, where a compact sound zone system is desirable. Owing to the limited array size of the health care system, the reproduction of bass may decrease the experienced sound zone separation because the beam-forming performance of the system starts to deteriorate below approximately 1 kHz (Fig. 2). The band-pass filtering of audio stimuli was therefore implemented in MATLAB before the listening tests.

2.5.4 Experimental Design

For each of the target-interferer combinations of audio (see previous section), a full factorial design was used to generate the stimuli based on the three factors reproduction mode (SZ or NoSZ), gain (0 dB or 6 dB), and filter (band-pass or full bandwidth). The final design is shown in the right-hand side of Fig. 4. Note that the two combinations NoSZ-BP-0dB and NoSZ-BP-6dB were not included because they were investigated and the results are shown in Fig. 12 in [14].

2.6 Attributes

Attributes were chosen based on two previous experiments [13, 14] and evaluations of a professional tonmeister (employed at Bang & Olufsen A/S) listening to the audio

Audio content				
Target audio		Interference audio		
		Rest	Entertainment	Information Gathering
Rest	Calm music 1	No interferer		
		Calm music 3		
			Pop rock music 2	
	Calm music 2	No interferer		
		Calm music 3		
			Pop rock music 2	
Entertainment	Pop rock music 1	No interferer		
		Calm music 3		
			Pop rock music 2	
	Podcast	No interferer		
		Calm music 3		
			Pop rock music 2	
Information Gathering	Podcast	No interferer		
		Calm music 3		
			Pop rock music 2	
	Radio sports	No interferer		
		Calm music 3		
			Pop rock music 2	

Stimulus	Mode	Filter	Gain	
			Target audio	Interfering audio
REF	Sound Zones	Band-pass	Unmodified	
1	Sound Zones	Full BW	Unmodified	Unmodified
2	Sound Zones	Full BW	+ 6 dB	Unmodified
3	Sound Zones	Band-pass	Unmodified	Unmodified
4	Sound Zones	Band-pass	+ 6 dB	Unmodified
5	No Sound Zones	Full BW	Unmodified	Unmodified
6	No Sound Zones	Full BW	+ 6 dB	Unmodified

(b)

Fig. 4. Overview of the audio stimuli, with audio content (a) within three use cases. For each combination of target and interfering audio, six different stimuli were generated (b) representing changes in reproduction mode, filter, and gain. The reference (REF) stimulus was available on each page in the user interface to make it possible to hear the target audio without an interferer. For a given target-interferer combination, the stimuli 1–6 and the reference were presented in a randomized order as the seven stimuli A–G in the user interface.

stimuli together with the hospital soundscape in the experimental setup. The three attributes annoyance, confusion, and distraction were included for all listening tasks to capture interference caused by both the hospital soundscape and the interfering audio and therefore reformulated from the previous experiments to better encompass this [e.g., “*Are the surrounding sounds (original: neighboring sound zone) distracting your ability to hear the audio content in your own soundzone?*” (translated from Danish)]. The inclusion of these three attributes in all listening tasks also allowed for evaluations of their correlations and relevance across different use cases.

Based on results from [14], attributes midrange strength, nasal, boomy, and width were excluded in entertainment scenarios. The attribute distance was included for speech stimuli. Perceived liking was considered a proxy for audio QoE by describing the overall liking of the entire acoustic environment (translated from Danish): “*To which degree do you like the sound environment?*” In total, seven attributes and liking were presented to assessors for the listening tasks entertainment and information gathering, whereas five attributes and liking were presented for the rest scenarios (Table 2).

2.7 Assessors

Eleven assessors (all students at Aalborg University, age 19–25, five women and six men) participated in the listening test. One assessor had participated in a previous experiment with ISOBEL sound zones and the simulated hospital soundscape [15], and another had participated in a sound zone listening test to determine audio QoE in a domestic setting [14]. All assessors had self-reported normal hearing.

Table 2. Perceptual attributes evaluated in the three listening tasks: rest, entertainment, and information gathering.

Rest	Entertainment	Information Gathering
Loudness*	Loudness*	Loudness*
Natural*	Clean*	Natural*
	Canny*	...
	...	Distance*
	Bass strength*	...
	...	Understandable†
Confusion‡	Confusion‡	Confusion‡
Annoyance‡	Annoyance‡	Annoyance‡
Distraction‡	Distraction‡	Distraction‡
Liking	Liking	Liking

*From the sound wheel [22]; †defined as “*It is easy to understand, what is being said in the audio clip?*”; ‡from [17]; ||defined as “*To which degree do you like the sound environment?*”

2.8 Experimental Protocol

One assessor at a time participated in the experiment, seated in hospital bed 1 (Fig. 1) with a headrest and pillow in a comfortable back position (approximately a 47° angle). The experimental protocol of the listening test contained four sections:

- (1) Questions about previous experiences in hospitals.
- (2) Familiarization to introduce the listening test interface and methodology.
- (3) Listening tests.
- (4) Post evaluation.

2.8.1 Previous Experiences With Hospitals

The assessor was welcomed and presented to the laboratory settings with a hospital bed surrounded by cur-

tains that visually concealed the sound sources. No sounds were active while the assessor was asked by the experimenter about previous experiences with hospitals and memory/knowledge of any hospital sounds to assess how previous exposures could affect perception.

2.8.2 Familiarization

The assessor was introduced to the listening test methodology by a familiarization protocol with a music-speech scenario to present the interface and rating methodology of the listening test. The audio content contained two audio stimuli that were not part of the actual listening test. The target stimulus was calm instrumental music composed by Wavecare, and the interfering audio was from the Archimedes Project [21]. The assessor was asked to rate the target-interferer scenario for two attributes that were also not included in the listening tests: “hissing” from the sound wheel [22] relating to the audio characteristics of the target stimulus and “chaotic,” covering perceived interference caused by the mixture of target and interfering audio [17].

2.8.3 Listening Tests

The experiment was divided into three listening tests, one for each of the three listening tasks: 1) rest, 2) entertainment, and 3) information gathering. The listening tests were presented in a randomized order as a modified multiple stimuli with hidden reference and anchor procedure, where the intensity of a given attribute was rated for a number of stimuli. The experimenter handed out written introductions presenting each listening task and setting the scene of a hospital environment before the test began. All introductions began with “*Imagine that you are at a hospital. You are in the hospital bed.*” (translated from Danish) but varied for the different tasks: “*You are tired and need to rest.*” (rest); “*You are awake and would like to be entertained.*” (entertainment); “*You are awake and would like to gather new information (e.g., from news).*” (information gathering).

The listening test was generated in a custom-built Max/MSP patch and broadcasted through Mira to a tablet (Fig. 5) fixed in a side mount holder on the bed frame. Each attribute was presented for one combination of target and interferer per page in the user interface. The order of attributes and audio content was automatically randomized within a listening test by the software. On each page in the user interface, the given audio content was presented as six stimuli given by the two reproduction modes and the two bandwidths within sound zone reproduction: 1) no sound zones, 2) sound zones full bandwidth, and 3) sound zones band-pass filtered (Fig. 3); the two target gains were 1) 0 dB and 2) 6 dB. A randomized hidden repeat was also added in a random position on each page to test the consistency in scores of the assessors, resulting in seven stimuli per page, denoted stimuli A–G.

The assessor was asked to evaluate their perception of audio by placing sliders on vertical bars (bar height on tablet interface = 5.2 cm) that represented the intensity of the given attribute. All attributes were evaluated on the same

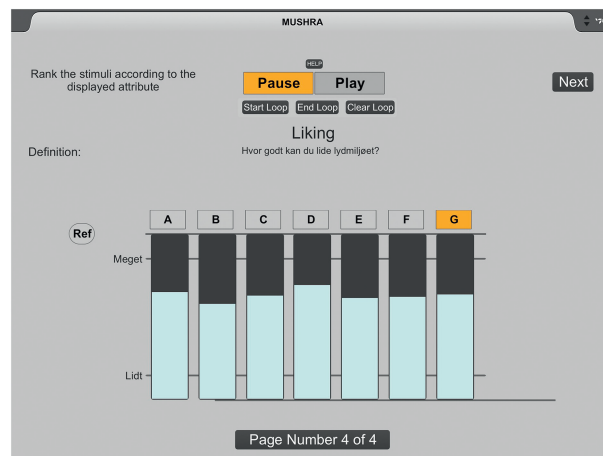


Fig. 5. The listening test user interface displayed on a tablet. Assessors were instructed to rate the given attribute, in this case liking, for each stimuli A through G. By pressing the “Ref” button on the left side, the assessor could listen to the target audio without an interfering audio but still in the presence of the hospital soundscape. Note that the instruction says “Rank the stimuli...,” which is a remnant from a previous setup with a different methodology than used here. Assessors were pretrained with a familiarization protocol to ensure that they understood the test before the data collection began.

scale. The scale covered scores of 0–100, with anchors at 15 (“A little”) and 85 (“A lot”). For the attribute distance, included for speech stimuli, a rating placed in the upper part of the scale meant that the sound was perceived as far away. This was described in the attribute definition on the user interface. For attributes loudness and bass strength, the anchors could also cause confusion in the directions of the scale, and assessors were therefore informed that the scale should be interpreted as “low” (15) to “high” (85) instead (also added to the attribute description). The slider was placed at a score of 50 by default and had to be moved to activate the given scale. The score was shown continuously when moving the slider and disappeared when a position on the scale was selected.

A reference (Ref) button was added to the previous user interface used in [14] to allow for playback of the target audio alone. The assessors were informed to listen to the reference stimuli first, before rating stimuli A–G. By this, the assessor could familiarize themselves with the target audio and obtain a better ability to distinguish the target from the interferer, which could be difficult in some scenarios. The assessor was also given the opportunity of creating audio loops to listen to a specific part of the audio and compare stimuli within that audio snippet. The data on using audio loops were not recorded, and it is therefore not known how many and how often this opportunity was used.

Assessors decided the pace of the test themselves, choosing when and how long they wanted to listen to a stimulus before rating it and whether they wanted to adjust ratings multiple times before going to the next page. In this way, the assessors were not rushed in their judgments. Before proceeding to the next page, the assessor was asked if they had evaluated all stimuli (yes/no). The maximum duration

of the experiment was set to 3 h, including introduction and 5-min breaks every 30–40 minutes, which was found efficient to avoid fatigue and maintain the rating consistency of assessors in a previous study with a similar experimental protocol [14]. Assessors had the option of pressing a “help” button to call in the test leader if they had questions about the test or needed an extra break. Throughout the test, the test leader overviewed the data collection on a monitor outside the laboratory that mirrored the tablet and showed the active audio channels. Furthermore, a visible video camera in the laboratory was used to continuously monitor the assessor to ensure assessor safety. Assessors had given consent to the camera monitoring and no data were logged.

Given the number of audio contents (Fig. 4) and the selected attributes for each listening task (Table 1), the number of pages in the user interface was 48 pages for rest scenarios, 32 pages for entertainment scenarios, and 32 pages for information gathering scenarios. With seven stimuli per page, each assessor was thereby asked to rate a total of 784 stimuli containing 672 original stimuli and 112 hidden repeats. Assessors were paid an hourly rate for their participation.

2.9 Collected Data

The listening tests conducted by two of the assessors had to be ended prematurely because of an error in the filtering of stimuli that randomly affected the reproduction of audio without sound zones. Although scenarios with sound zones were unaffected by the error, the effect on scale use and overall perception was evaluated as too influential to keep any of the data, which therefore led to exclusion of these two datasets. As a result, data from nine assessors (three women, six men) remained in the analysis. Six of these assessors completed all evaluations, and three had a reduced dataset resulting from not attending all sessions.

2.10 Processing and Analytical Methods

Data were exported from Max/MSP to the R environment (R Studio, version 1.4.1103) for statistical analysis. Analyses were conducted separately for each task. The data were first tested with histograms and summary statistics. The scale reliability was tested with Cronbach’s alpha test⁵, whereas the rating consistency of assessors was tested by calculating the absolute differences between original stimuli and hidden repeats on each page in the user interface. The statistical analysis used a general 0.05 criterion in relevant tables shown as an asterisk (*; $p < 0.05$). In cases with higher significance levels, this is indicated by a double asterisk (**; $p < 0.01$) and a triple asterisk (***; $p < 0.001$). To check for multicollinearity issues, correlations of attributes were calculated with Spearman’s correlation coefficients⁶ that can handle non-normality in data, and the results were used to reduce the attribute lists within each use case.

⁵ R package ltm; <https://CRAN.R-project.org/package=ltm>

⁶ R library PerformanceAnalytics, function chart.correlation; <https://CRAN.R-project.org/package=correlation>

An analysis of variance (ANOVA)⁷ was conducted for each attribute individually to test whether the experimental variables had a significant influence on the attribute ratings. The ANOVA models included the following physical variables as fixed factors: mode (reproduction mode with or without sound zones: NoSZ or SZ), gain (0 or 6-dB gain to the initial sound pressure level) and content (the combination of audio in the two audio streams, A and B) including all two-way interactions. For sound zone scenarios, filter (full bandwidth or band-pass filtered) was also added as a factor nested within mode (mode:filter). Assessor was added as a random factor to consider the perception on a population level rather than an individual level [23, 24].

For the listening task rest, the variable content originally contained eight levels, given the two target signals and four interference signals. However, because the ratings for the two target signals were found to not differ significantly (one-way ANOVA, $p = 0.37$ – 0.98), the target signals were grouped together and resulted in content being reduced to four levels, similar to the levels in the two other listening tasks. To explore whether the customized Wavecare audio (*calm music 1*; Table 1) was a better masker of the hospital soundscape than the other audio material, the differences in perceived distraction were evaluated with an ANOVA for compiled data across the three use cases. Here, the predictor task was added to the model that otherwise included the same physical parameters as the ANOVA model above.

Backward elimination⁸ was used to calculate the reduced models containing only the significant ($p < 0.05$) factors for each attribute. Attributes that were found to include one or more significant factors were retained for further analysis. Conditional R^2 values⁹ were calculated to evaluate the explained variance of each model, whereas estimated marginal means¹⁰ were calculated to obtain the significance of paired contrasts, which applies Tukey’s honest significant difference test for p -value adjustment in multiple comparisons.

Next, the attributes in the reduced attribute lists were included in a regression analysis¹¹ to evaluate their relationship with liking as a proxy for audio QoE. Backward elimination was again used to retrieve a reduced model only containing the significant ($p < 0.05$) factors.

3 RESULTS

3.1 Listening Test

3.1.1 Assessor Scale Use and Consistency

In the listening tests, assessors used the majority of the scale (ranges of 83–100 out of 100). All assessors used

⁷ R library lmerTest, function lmer; <https://CRAN.R-project.org/package=lmerTest>

⁸ R function ‘step’; <https://www.rdocumentation.org/packages/lmerTest/versions/2.0-36/topics/step>

⁹ R library MuMIn, function r.squaredGLMM; <https://CRAN.R-project.org/package=r2glmm>

¹⁰ R library lsmeans, function emmeans; <https://CRAN.R-project.org/package=emmeans>

¹¹ R library lme4; <https://CRAN.R-project.org/package=lme4>

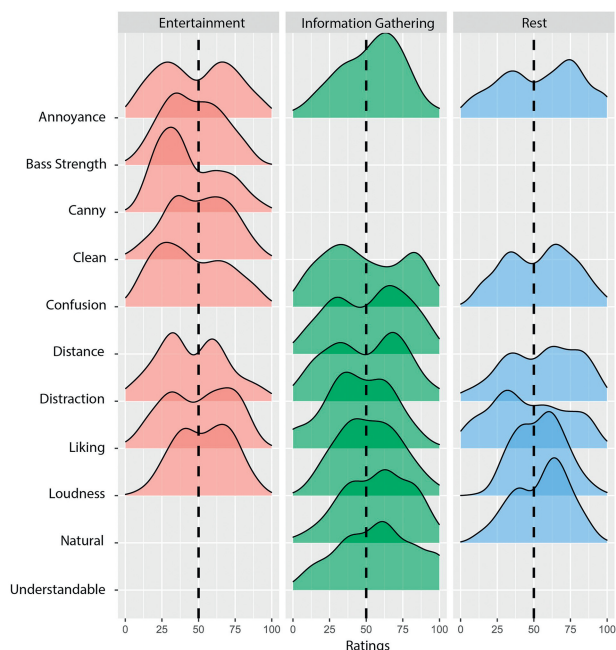


Fig. 6. Distributions of attribute ratings over the rating scale (0–100) for all attributes evaluated in the three use cases.

the scale below and above the anchors at 15 and 85. The rating consistency was evaluated as score differences between original stimuli and hidden repeats and found to have medians of 2–12 (interquartile range 1–26) across all assessors and attributes. The reliability of the rating scales was generally high across most attributes, including liking in rest and entertainment scenarios.¹² A lower scale reliability was found for distance and liking rated in the information gathering task.¹³ The distributions of ratings in the three use cases varied (Fig. 6) but with the majority of attributes showing bimodality in line with the previous experiment [14] with a dip in ratings around the central score of 50.

3.1.2 Attribute Correlations

Attribute correlations were evaluated for scenarios with an interfering audio present (i.e., silence in SZ B scenarios were not included). Across all listening tasks, liking was found to be negatively correlated with distraction (Spearman's correlation coefficient $\rho = -0.34$ to -0.42 ; $p < 0.001$), annoyance ($\rho = -0.29$ to -0.56 ; $p < 0.01$), and confusion ($\rho = -0.33$ to -0.53 ; $p < 0.001$). Liking was also found to be significantly negatively correlated with canny ($\rho = -0.29$; $p < 0.001$) and distance ($\rho = -0.22$; $p < 0.05$). A high rating of distance means that the sound is perceived as far away, and the negative correlation thereby indicates that assessors in general liked stimuli more when the sound was perceived close by.

Liking was positively correlated with loudness in entertainment ($\rho = 0.18$; $p < 0.05$) and information gathering

scenarios ($\rho = 0.28$; $p < 0.01$) but not in rest scenarios ($\rho = 0.062$). Liking was also positively correlated with bass strength ($\rho = 0.39$; $p < 0.001$), clean ($\rho = 0.31$; $p < 0.001$), natural ($\rho = 0.31$; $p < 0.001$) and understandable ($\rho = 0.23$; $p < 0.001$). For speech stimuli, loudness was positively correlated with understandable ($\rho = 0.44$; $p < 0.001$) and negatively correlated with distance ($\rho = -0.23$; $p < 0.05$).

Annoyance, distraction, and confusion were significantly positively correlated with each other ($\rho = 0.33$ – 0.49 ; $p < 0.001$). This high correlation indicates that one attribute could represent the perceived interference. As previous studies have showed that perceived distraction is a good predictor of perceived audio-on-audio interference [17] and the availability of a real-time model for distraction based on five measurable acoustic parameters [25], this attribute was chosen as representative for perceived interference onward in the analysis, and attributes annoyance and confusion were therefore left out of further analysis.

3.1.3 Perceptual Changes Resulting From Physical Variables

In the variance analyses (Table 3), all attributes presented in the three listening tasks were found to be significantly influenced by at least one of the physical factors, mode and gain. Note that attributes confusion and annoyance were excluded because of their high correlation with distraction (see SEC. 3.1.2). The audio content also significantly influenced the ratings of many attributes and often interacted with mode. The effects of filtering were nested within the reproduction mode because the band-pass filter was only applied in sound zone scenarios.

The perceived loudness was significantly influenced by gain across all use cases, which verifies that the selected 6-dB change in gain was perceivable by the assessors. In line with this, the addition of 6 dB to the reproduction level also influenced the perception of distance of a speech signal. Furthermore, gain affected how understandable a speech stimulus was perceived, likely a result of improvements in the TIR, which is supported by the positive correlation of understandable with loudness (see SEC. 3.1.2). In Fig. 7, it is seen that gain influenced ratings of distraction across all use cases. The addition of a 6-dB gain also clearly influenced liking for entertainment and information gathering scenarios, whereas the effects are less obvious in rest scenarios. This is confirmed by the variance analysis in Table 3, where gain was not found to significantly predict liking in the rest use case.

Despite all stimuli being loudness-matched and calibrated for the listening position, the perceived loudness also depended on the audio content for the rest and information gathering scenarios. For pop/rock music (entertainment), which contains a broad range of frequencies including bass components, filter also affected loudness in an interaction with mode. This shows that the presence/absence of low-frequency components is perceivable for scenarios with pop/rock music as the target. However, filtering of stimuli does not significantly affect liking or perceived distraction.

¹² Alpha values of 0.69–0.79.

¹³ Distance had an alpha value of 0.58, whereas liking had an alpha value of 0.56.

Table 3. Variance analysis for attributes and liking presented for the three listening tasks (columns).

Rest	Entertainment	Information gathering
Loudness ~ gain*** + content** Mode:Content* + gain:content* SW- $p = 0.122$, $R^2c = 0.38$	Loudness ~ mode*** + gain*** + mode:filter*** SW- $p = 0.774$, $R^2c = 0.74$	Loudness ~ gain*** + content* SW- $p = 0.187$, $R^2c = 0.54$
Natural ~ mode** + gain* + content*** + mode:content*** SW- $p = 0.065$, $R^2c = 0.24$	Clean ~ mode + gain*** + content* + mode:content* SW- $p = 0.323$, $R^2c = 0.48$ Canny ~ gain*** SW- $p = 0.465$, $R^2c = 0.40$ Bass strength ~ mode* + gain*** + mode:filter*** SW- $p = 0.455$, $R^2c = 0.46$	Natural ~ gain*** + content** ... SW- $p = 0.0001$, $R^2c = 0.45$ Distance ~ gain*** + content* SW- $p = 0.069$, $R^2c = 0.19$ Understandable ~ mode* + gain*** + content*** + mode:content** SW- $p = 0.011$, $R^2c = 0.52$
Distraction ~ Mode*** + gain*** + Content*** + mode:content*** SW- $p = 0.737$, $R^2c = 0.40$	Distraction ~ mode*** + gain*** + content*** + mode:content** SW- $p = 0.441$, $R^2c = 0.39$	Distraction ~ mode*** + gain*** + content*** + mode:content** SW- $p = 0.367$, $R^2c = 0.55$
Liking ~ Mode*** + content*** + Mode:content*** SW- $p = 0.220$, $R^2c = 0.54$	Liking ~ mode*** + gain*** + content*** + mode:content*** SW- $p = 0.833$, $R^2c = 0.39$	Liking ~ mode*** + gain*** + content*** + mode:content** SW- $p = 0.234$, $R^2c = 0.59$

Note: Attributes annoyance and confusion are excluded because of their high correlation with distraction. Type III ANOVA significance with Satterwaite's method are shown with an asterisk (*) for p -values below 0.05. In cases with higher significance levels, this is indicated by a double asterisk (** $p < 0.01$) and a triple asterisk (***) $p < 0.001$). Normality of residuals of the reduced models was tested with a Shapiro-Wilks test; SW- p values are shown in the table together with explained variance calculated as conditional R^2 values. For non-normal model residuals, estimates of skewness (acceptable ± 3) and kurtosis (acceptable ± 10) were found acceptable and therefore not included here. The intercept and error term are implicit in each model.

SW- p = Shapiro-Wilks p -values.

In contrast, the reproduction mode, comparing wide-angle audio and sound zones, had a clear significant influence on ratings of liking and distraction across all use cases. Playing audio in sound zones generally increased liking and decreased distraction across the three listening task cases compared with wide-angle reproduction (Fig. 7).

The impact of the reproduction mode interacted with the audio content (i.e., mode:content) for both distraction and liking in all use cases. The interaction effect is shown in Fig. 8 for rest scenarios. Here, sound zones ("SZ") in general increased liking of audio-on-audio scenarios significantly ($p < 0.0001$) and decreased distraction ($p < 0.001$) compared with wide-angle ("NoSZ") reproduction. Scenarios with pop/rock music or radio sports as the interferer in the wide-angle mode had the lowest liking and highest distraction compared with calm music. However, when playing audio in sound zones, the liking and distraction was similar regardless of the interferer. As a result, the largest effect of playing audio in sound zones on the liking and distraction of rest scenarios was seen when the interfering audio was pop/rock music or radio sports. The scenarios with no interferer (i.e., silence) showed no differences in ratings between the two reproduction modes.

The interaction effect between reproduction mode and audio content is largely driven by the differences between the silence condition and audio-on-audio scenarios with

interfering audio content. Although the silence condition can indicate the level of distraction caused by the hospital soundscape (see next section), the difference between ratings of scenarios with and without an interferer can indicate the distraction caused by the interfering audio.

In sound zone scenarios, perceived distraction increases in the presence of interferers by 12.0–14.7 score points compared with no interferer and is not significantly different for the three audio contents (Fig. 8). In comparison, in scenarios with wide-angle audio, the added distraction by interferers depends on the audio content, with a higher distraction when the interferer is pop/rock music or radio sports than when the interferer is calm music. In the same way, the difference in ratings of liking in scenarios with and without interfering audio may indicate how an interferer influences the affection of the assessor. Liking is significantly lower when the interfering audio is radio sports or pop/rock music compared with silence or calm music played in wide-angle mode, whereas liking is more similar across all audio genres when reproduced in sound zones (Fig. 8).

3.1.4 Influence of the Hospital Soundscape

In scenarios without an interfering audio (i.e., silence) it was possible to evaluate the influence of the hospital soundscape on the listening experience of the target stimu-

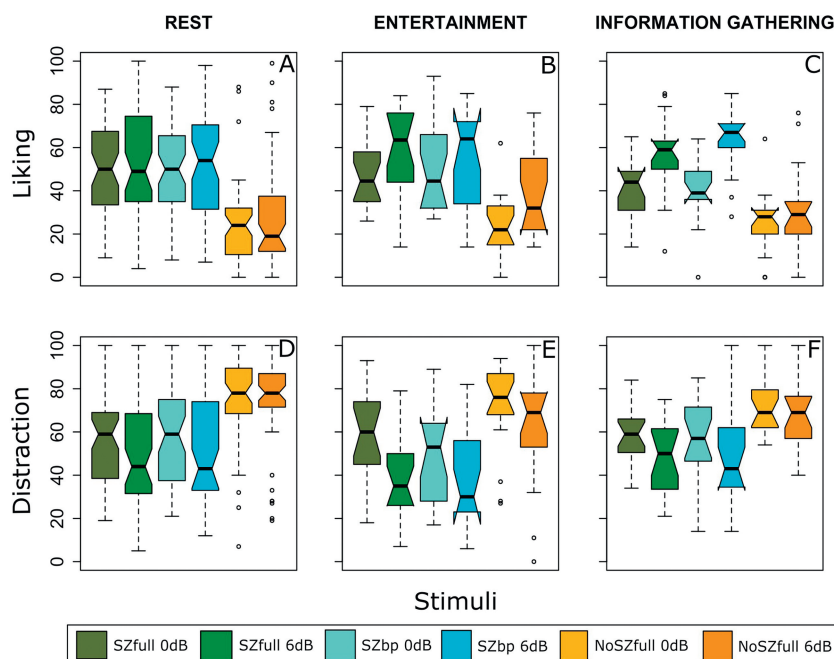


Fig. 7. Boxplots of ratings of liking (A to C, top) and distraction (D to F, bottom) across all three listening tasks (columns) for the six stimuli (see colorbar) represented on each page in the user interface. The stimuli are combinations of two reproduction modes (SZ; NoSZ), two filters (full = full bandwidth; bp = band-pass filter) and two gain settings (0 or 6-dB gain added to the original reproduction level). Within each listening task, ratings are averaged across the three audio scenarios that included an interfering audio (i.e., the silence in SZ B scenario is not included). Boxplots show the minimum, lower quartile, median (horizontal line) with the 95% CI (notches), upper quartile, and maximum.

lus. For rest scenarios as an example (Fig. 8), the perceived distraction of the hospital soundscape alone was around 50, regardless of the reproduction mode of the target stimulus. Despite variations in the sounds reproduced in the hospital soundscape, which contained both low-amplitude broadband sound and transient alarms, the CIs around the means were relatively small, indicating a high degree of agreement between the assessors regarding the distraction of the soundscape.

Within rest scenarios, two audio stimuli were used as target audio, with one of them being composed specifically to mask ICU alarms (Table 1) that were reproduced with the simulated hospital soundscape. Yet, this study did not find any significant differences in ratings/perception between this target audio and another track with calm music (see SEC. 2.1).

The level of distraction caused by the hospital soundscape was also evaluated across all three use cases. This was to test whether one type of target audio content would be a better masker of the soundscape and thereby decrease the perceived distraction. No significant differences ($p = 0.2$) were found between use cases, thereby indicating that none of the audio genres tested here were better than others at masking the hospital soundscape.

3.1.5 Evaluating Liking as a Proxy for Audio Quality of Experience

The perceived liking of the sound environment (as a proxy for audio QoE) was evaluated with linear regression models using the significant perceptual attributes as the

predictors. Model results for liking for the three listening tasks are seen in Table 4. All reduced models were found to include the attribute distraction as a significant predictor. For information gathering scenarios (right column), distraction is the only attribute predicting liking scores (22% explained variance), indicating a decrease of -0.45 score points in liking when the distraction score increases by 1. For entertainment scenarios, bass strength is also a significant predictor of liking, resulting in a total of 50% explained variance and negligible variance inflation caused by the interaction between distraction and bass strength [variance inflation factors (VIF) = 1.08]. For rest scenarios, the perception of natural is predicting liking with a total of 22% explained variance and no correlation between distraction and natural (VIF = 1.01).

4 DISCUSSION

4.1 Perceptual Dimensions of Sound Zones

The perceptual attribute distraction was found to be highly correlated with the two other interference attributes annoyance and confusion (Sec. 3.1.2). Because of their high multicollinearity and the availability of a real-time model for distraction [25], only distraction was kept in the further analysis and is therefore in focus when discussing the experienced interference caused by the interfering audio and the hospital soundscape.

Table 4. Regression analysis of perceived liking in each of the three listening tasks (columns).

Attributes	Rest	Entertainment	Information Gathering
Full model	Liking $\sim$ loudness + natural + distraction	Liking $\sim$ loudness + canny + clean + bass strength + distraction	Liking $\sim$ loudness + natural + understandable + distance + distraction
Reduced model	Liking $\sim$ distraction + natural $R^2c = 0.20$, $p = 0.613$, VIF = 1.01 <i>Estimate/SE/p-value</i>	Liking $\sim$ distraction + bass strength $R^2c = 0.50$, $p = <0.001$, VIF = 1.08 <i>Estimate/SE/p-value</i>	Liking $\sim$ distraction $R^2c = 0.22$, $p < 0.001$ <i>Estimate/SE/p-value</i>
Intercept	60.58/5.04/***	60.41/10.76/***	69.52/6.61/***
Distraction	−0.42/0.05/***	−0.65/0.13/***	−0.45/0.12/***
Natural	0.16/0.06/*	...	...
Bass strength	...	0.37/0.14/***	...

Note. Models are calculated with 'lm' in R and reduced with backward elimination using a 0.05 elimination criterion with the 'step' function. Conditional R^2 values are shown for the reduced models together with Shapiro-Wilks test p -values and VIF when more than one attribute is predicting liking. For models with non-normal residuals, estimates of skewness (acceptable $+/-3$) and kurtosis (acceptable $+/-10$) were found acceptable and therefore not included here. Significance levels are shown for p -values below 0.05 as and asterisk (*). In cases with higher significance levels this is indicated by a double asterisk (** $p < 0.01$) and a triple asterisk (***) ($p < 0.001$). The intercept and error term are implicit in each model.

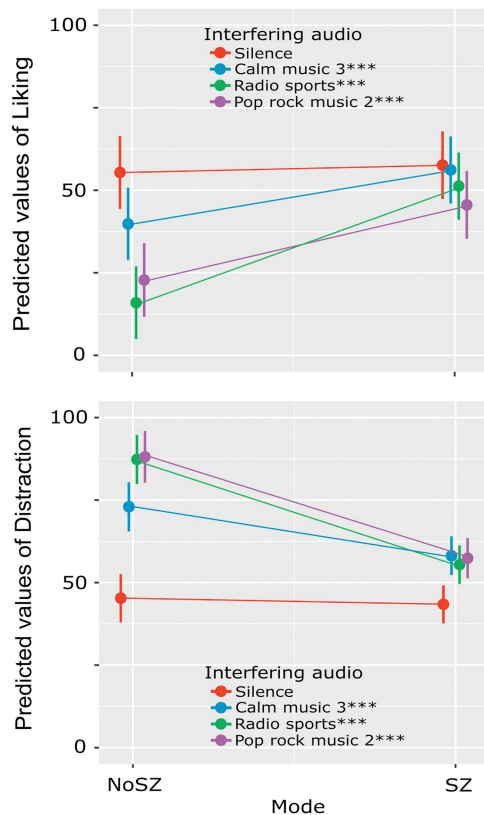


Fig. 8. The predicted values of liking (top) and distraction (bottom) in rest scenarios for the interaction between reproduction mode (left/right) and audio content (colors) as seen in Table 3. The values are shown as estimated marginal means with 95% CIs for each of the four audio contents played as the interferer with p -value significance of the interaction <0.05 (*), <0.01 (**), and <0.001 (***).

4.2 Sound Zones Versus Wide Angle

The results of this study confirm that reproduction in sound zones improves liking, as a proxy for the audio QoE, for all three use cases compared with wide-angle reproduc-

tion of two audio streams (Fig. 7). When audio-on-audio scenarios occurred in sound zones, the perceived distraction caused by interfering audio was reduced, whereas the liking of the sound environment increased compared with audio played in wide angle. This is likely to be a result of the improved TIR experienced with sound zones and thus an improved ability to focus/concentrate on the target sound as the exposure to the interfering audio is reduced. Directional cues can be expected to be more blurred in a reverberant room, such as a real hospital room, which can make it more difficult to distinguish target and interfering sound. This is particularly true for wide-angle audio, which will thereby enhance the benefits of a sound zone system in scenarios where speech intelligibility is important. However, when only a target audio is present, the benefits of having sound zones is not observed. This is to be expected because the level of distraction (in this case with the hospital soundscape being the distractor) and experienced liking is similar for wide-angle and sound zone scenarios (Fig. 7). Here, it must be noted that sound zones are intended for areas with multiple listeners, so when only one audio content is played in a sound zone, listeners in the surroundings may still benefit from the reduced leakage of sound into the shared space, but that situation was not tested here.

The wide-angle reproduction of audio was reproduced with individual drivers on each end of the health care system array with a distance of 1.97 m from each other. Placing the individual sound sources at a longer distance from each other and/or closer to the intended listening positions could have improved the TIR. However, to avoid any influence of individual loudspeaker driver characteristics on the perceptual differences between wide-angle audio and sound zones, it was decided to use the same drive units for both modes. Also, this setup was considered a viable use case in a health care environment because it is not uncommon to have multiple audio sources (e.g., TVs) within close range in hospital rooms with multiple patients. In one-patient rooms, multiple audio streams may also exist within

close range because these rooms can contain patients, relatives, and hospital staff that may all have different desires and needs for the audio environment.

4.3 Target-to-Interferer Ratio

The influence of the TIR was confirmed by the variance analyses (Table 3); a 6-dB increase in gain of the target audio had significant effects on perception of all attributes except ratings of liking within rest scenarios (Table 3 and Fig. 7). The lack of gain as a predictor of liking of calm music may be caused by nonvocal music requiring a lower TIR than pop/rock music and speech, where vocal information must to be decoded and understood. Still, the level of distraction when listening to calm music was significantly affected by changes in reproduction level. This highlights that TIR also matters in rest scenarios, but this study has a limited ability to assess this using only two reproduction levels. The next experiment within the project [26] was improved by including four levels of TIR in sound zone scenarios.

An opposite influence of TIR on perceived distraction was noted by one of the assessors in the postevaluation, who found it more distracting when the interfering audio was low compared with high. This could be caused by the assessor in a low TIR scenario needing to listen harder to recognize the interfering audio and distinguish the two audio streams, particularly in the presence of the hospital soundscape. In everyday life, the human auditory system and brain are constantly exposed to multiple audio streams simultaneously and are conducting auditory stream segregation to focus auditory attention and pinpoint signals of interest while keeping other audio streams in a “background listening” mode [27]. Segregating audio has been shown to be more challenging in environments with high noise load (i.e., the hospital soundscape reproduced here) because of an increased perceptual demand.

4.4 Filters and Masking

With the current setup, the removal of bass with a band-pass filter decreased the perceived bass strength of pop/rock music (Table 3). It was hypothesized that if the low-frequency components (below 150 Hz) in full bandwidth audio reproduction could improve masking of hospital noise compared with band-pass filtered audio [28], it would be noticeable as decreased distraction and elevated liking. However, changes in bandwidth were not found to significantly affect the perceived distraction or liking of any audio content (Table 3). This could be a result of the main distracting elements in the soundscape being alarms with peak frequencies at 1–3 kHz that overlapped with the same elements in both audio bandwidths. Also, the impact of the NoSZ scenarios on all pages of the listening test may have reduced the resolution of the rating scale, hiding the smaller differences between the two sound zone filters (full bandwidth and band-pass filtered audio). A track was composed to specifically mask ICU alarms but was found to not significantly change/improve the perceived distraction compared with the other tracks, as discussed in SEC. 3.1.4.

4.5 Attributes Predicting Audio Quality of Experience of Sound Zones

To better understand how different perceptual attributes contribute to the experienced audio QoE, the relevant attributes were evaluated with linear regressions for liking (proxy for audio QoE; Table 4). In all three listening tasks, distraction was found to be the main predictor of liking (note that annoyance and confusion were excluded before this analysis; see SEC. 3.1.2).

For rest scenarios with calm music as the target audio, attributes distraction and natural were found to predict liking. For the information gathering scenarios, where a podcast was chosen as the target stimulus, perceived distraction was found to be the only significant predictor of liking. The low explained variance of the models (both explaining 22% of variance) indicates that other perceptual attributes, which were not included here, could be relevant for estimating the audio QoE. This suggests that a higher resolution of the physical factors in the form of more levels, should be investigated to improve the model. For entertainment scenarios, with pop/rock music as the target audio, liking was predicted by attributes distraction and bass strength with an explained variance of 50%, highlighting the importance of bass for audio QoE of at least some audio genres.

The attribute loudness was removed through backward elimination in all three models for liking and was therefore not part of the reduced models. This suggests that loudness is not a relevant predictor of audio QoE in the selected scenarios, which is likely because of the use of comfortable listening levels determined by the optimal loudness test (see SEC. 2.4.2). In other scenarios and during normal use of sound systems, perceived loudness is expected to have a large impact on the listening experience.

Ratings of the attribute understandable for speech signals was not found to significantly predict the liking of a podcast, which is again likely owing to stimuli being reproduced at a level that resulted in a sufficient speech intelligibility regardless of the gain setting (0 or 6 dB). Also, assessors repeatedly heard the same 1-min clip and therefore had an increasing ability to understand the audio information. Oppositely, their liking may have been negatively influenced over time because of increasing annoyance with repetitions of the same audio clips, as noted by one assessor. Recent studies have investigated how audio perception evolves over time, using methods such as the temporal check-all-that-apply method [29]. In the study [26] that follows this work within the ISOBEL project, the authors further investigate how perception and affection may change over time.

None of the models (Table 4) included any spatial characteristics because distance (information gathering) was eliminated in the backward elimination and width (entertainment) had previously been excluded because it was not sufficiently understandable and had a low explained variance for physical factors [13]. The beamforming technology of sound zones allows for dynamic sound that can change beam width and position in the room to adapt to the listener(s) changing needs/desires. These dynamic properties were not included in the present study, where the listener position was fixed to a hospital bed, which was evaluated as

an ecologically valid use case for a health care environment, but may have affected the relevance of spatial characteristics on liking. Also, the assessors were not given any visual cues about the sound sources (they were hidden behind curtains), which could have affected their perception because auditory and visual senses are normally working tightly together to analyze the environment [30].

4.6 Sound Zones in Health Care Environments

The listening tests reported here were conducted in a laboratory setup with a simulated soundscape to improve the ecological validity of audio perception and affection in sound zones. In scenarios without interfering audio (i.e., silence in SZ B), the experienced interference caused by the hospital soundscape could be assessed and was showed to result in considerable distraction (Fig. 8). The distraction caused by the hospital soundscape alone was similar for sound zone or wide-angle scenarios.

The level of distraction was always evaluated for scenarios that included the target audio and a potential distractor (hospital soundscape or interfering audio) because it requires a signal of interest that one can be distracted from hearing. It was therefore not viable to evaluate the distraction of the soundscape alone without a target audio, which could have helped to understand if and to what degree different target audio contents were able to mask the soundscape. No difference in distraction was found for the different audio material within the three use cases when presented in silence scenarios, indicating that their ability to mask the hospital soundscape was similar.

The user experience study by [15], which used a similar experimental setup as this present study, can help to shed light on the benefits of playing audio in a simulated hospital environment. Here, assessors increased their perception of calmness and reassurance¹⁴ in any case of audio reproduction, including audio-on-audio, compared with only listening to the hospital soundscape. This therefore suggests that audio-on-audio scenarios reproduced in sound zones is preferable compared with absence of audio in a health care environment. The experience depended on the audio content, with higher ratings of calm when calm music was played alone in the target sound zone compared with a speech signal.

Since the hospital soundscape was always active in this present study, distraction caused by the interfering audio alone could not be evaluated. It is suggested that the difference between perceived distraction in scenarios with and without interferers may represent the distraction caused by the interfering audio (Fig. 8). Yet, the distraction may not have been additive, with the interfering audio adding additional distraction on top of the existing hospital soundscape. Instead the interfering audio may have aided the target audio in masking the soundscape and thereby been experienced as the major distractor, or vice versa, the soundscape may have partially masked the interfering audio and limited its ability to distract the assessor.

It is seen that the total level of distraction caused by the combination of hospital soundscape and interfering audio is smaller when audio content is played in sound zones compared with wide-angle audio (Fig. 7). This is in agreement with the positive effects of the reproduction mode, where sound zones reduce distraction.

4.7 Influence of Sensitization or Habituation

The hospital soundscape used here was produced from recordings of sounds of real hospital equipment at an ICU, with randomization of the noise reproduction during listening test implemented to reach similar effects across all attributes and stimuli. Hospital alarms and medical equipment emit frequent high-level sounds that are among the most disturbing noises in ICUs [5]. The relatively high exposure rate of noise from alarms and machinery in the soundscape used here may have been overestimating the noise exposure at a real hospital, resulting in elevated distraction. Also, the task of having to continuously assess your perception will increase the awareness of the interfering audio and limit the ability of assessors to become immersed [31] in the target audio.

On the other hand, the laboratory assessors may have been less affected by the hospital soundscape than real patients would be because the noise from alarms and other hospital equipment were not linked to the current health status of themselves or people close by and none of the assessors have had traumatic hospital experiences. Multiple assessors noted that they were less distracted by the hospital soundscape over time, whereas patients that would be repeatedly alerted by noise from alarms and other sources over a long-term exposure could instead become sensitized to the noise. The uncertainty of the duration and lack of control of the noise exposure could cause further sensitization in real patients, whereas a test participant always had the option of stopping the experiment and thereby escaping the noise exposure.

5 CONCLUSION

The first research goal in this study was to compare the perceived audio QoE for audio-on-audio scenarios reproduced by a sound zone setup (SZ) versus a wide-angle (NoSZ) setup. The results show that a sound zone setup can improve the listening experiences in scenarios where two audio streams are played simultaneously and evaluated for three different use cases (rest, entertainment, and information gathering) in a simulated hospital setting. When reproducing audio in sound zones, assessors experienced less distraction and greater liking (proxy for audio QoE) compared with wide-angle reproduction.

The second research question was to quantify how changes in TIR for the investigated use cases influence the perception and affection of the stimuli. The results show that improving the TIR by 6 dB has a clear positive effect when listening to pop/rock music (entertainment) or a podcast (information gathering) but not when listening to calm music (rest). This study used an estimated optimal

¹⁴ The semantic scales were calm-agitated and reassured-worried.

loudness for the reproduction level of stimuli and contained only two levels of TIR, which was likely influencing the relevance of some attributes. The attribute understandable was predicted by the gain but not found to be a significant predictor of liking, perhaps because the speech intelligibility was always sufficient for decoding the stimuli. Future studies should include a better resolution of TIR with more levels to investigate how TIR influences audio perception and affection of sound zones.

The third research question was to determine how removal of bass components in a sound zone setup will influence audio QoE. The results obtained here show that the perceived bass strength is negatively affected by bass removal, whereas there are no significant effects on distraction and liking of the two filter settings (Table 3). Still, the liking of the entertainment scenario was found to depend on both perceived distraction and bass strength (Table 4). Because of the limited number of conditions examined, more testing is needed to evaluate the trade-off between improving audio quality with full bandwidth audio and improving the sound zone separation with high-pass filtered audio.

Additionally, it can be concluded that the three attributes annoyance, confusion, and distraction, which focused on evaluating the perceived interference by the interfering audio and hospital soundscape, were found to be highly correlated. Owing to the availability of a real-time model for prediction of distraction [25], this attribute was chosen for further analysis in the following study (part 3) [26]. In all three use cases tested here, liking was found to be mainly predicted by distraction in addition to natural (rest) and bass strength (entertainment) (Table 4). Part 3 will further evaluate how modeling of distraction based on physical factors can be used to predict liking.

This study shows that the use of sound zones can help to improve the acoustic environment in a simulated hospital setting. Further studies of sound zones in real hospital environments are needed to better understand patient needs and how sound zones should be adapted to maximize the acoustic benefits in different scenarios. The results of this study can help to focus efforts and adapt future health care systems. The findings were used to improve the experimental protocol for listening tests in simulated and real homes within the project [26]. This series of studies within the ISOBEL project help to uncover how to optimize dynamic sound zones to minimize perceived distraction and improve the audio QoE in domestic and health care environments.

6 ACKNOWLEDGMENT

This project was funded by the Innovation Fund Denmark (Project Number 9069-00038B) and the Interactive Sound Zones for Better Living (ISOBEL) project partners. Thanks to Dr. Martin Bo Møller (Bang & Olufsen) for creating the adaptive filters for emulation of the health care prototype and for support with the setup, Kate Bosen (Bang & Olufsen) for software development and support, Dr. Geoff Martin (Bang & Olufsen) for expert evaluations of the

sound zone scenarios, Claus V. Skipper (Aalborg University) for support with the experimental setup, Wavecare ApS for the audio material, and Dr. Jon Francombe (Bang & Olufsen) for valuable discussions on data analysis and for reviewing a previous edition of the manuscript. Special thanks to the participants.

7 REFERENCES

- [1] World Health Organization (WHO), "Chapter 11: Environmental Noise," in *Compendium of WHO and Other UN Guidance on Health and Environment* (World Health Organization, Geneva, Switzerland, 2022). https://cdn.who.int/media/docs/default-source/who-compendium-on-health-and-environment/who-compendium_noise_01042022.pdf.
- [2] B. Van Rompaey, M. J. Schuurmans, L. M. Shortridge-Baggett, S. Truijen, and L. Bossaert, "Risk Factors for Intensive Care Delirium: A Systematic Review," *Intensive Crit. Care Nurs.*, vol. 24, no. 2, pp. 98–107 (2008 Apr.). <https://doi.org/10.1016/j.iccn.2007.08.005>.
- [3] M. Basner, W. Babisch, A. Davis, et al., "Auditory and Non-Auditory Effects of Noise on Health," *Lancet*, vol. 383, no. 9925, pp. 1325–1332 (2014 Apr.). [https://doi.org/10.1016/S0140-6736\(13\)61613-X](https://doi.org/10.1016/S0140-6736(13)61613-X).
- [4] G. Loupa, "Influence of Noise on Patient Recovery," *Curr. Pollut. Rep.*, vol. 6, no. 1, pp. 1–7 (2020 Mar.). <https://doi.org/10.1007/s40726-019-00134-3>.
- [5] N. Akansel and S. Kaymakçi, "Effects of Intensive Care Unit Noise on Patients: A Study on Coronary Artery Bypass Graft Surgery Patients," *J. Clin. Nurs.*, vol. 17, no. 12, pp. 1581–1590 (2008 Jun.). <https://doi.org/10.1111/j.1365-2702.2007.02144.x>.
- [6] M. J. Park, J. H. Yoo, B. W. Cho, K. T. Kim, W.-C. Jeong, and M. Ha, "Noise in Hospital Rooms and Sleep Disturbance in Hospitalized Medical Patients," *Environ. Health Toxicol.*, vol. 29, p. e2014006 (2014 Aug.). <https://doi.org/10.5620/eh.2014.29.e2014006>.
- [7] Global Council on Brain Health, "Music on Our Minds: The Rich Potential of Music to Promote Brain Health and Mental Well-Being." https://www.aarp.org/content/dam/aarp/health/brain_health/2020/06/gcbh-music-report-english.doi.10.26419-2Fpia.00103.001.pdf (accessed 3 May 2022).
- [8] B. Truax, *Acoustic Communication* (Ablex Pub. Corp., Norwood, NJ, 1984). <https://doi.org/10.5040/9798216955412>.
- [9] M. Lawrence, R. Ramirez, and P. Bauer, "Communicating With Unconscious Patients: An Overview," *Dimens. Crit. Care Nurs.*, vol. 42, no. 1, pp. 3–11 (2023 Jan.-Feb.). <https://doi.org/10.1097/DCC.0000000000000561>.
- [10] E. Persson and S. Määttä, "To Provide Care and Be Cared for in a Multiple-Bed Hospital Room," *Scand. J. Caring Sci.*, vol. 26, no. 4, pp. 663–670 (2012 Dec.). <https://doi.org/10.1111/j.1471-6712.2012.00976.x>.
- [11] J. Hu, C. Chen, M. Wu, et al., "Assessing Consciousness in Acute Coma Using Name-Evoked Responses," *Brain Res. Bull.*, vol. 218, p. 111091 (2024 Nov.). <https://doi.org/10.1016/j.brainresbull.2024.111091>.

- [12] W. F. Druyvesteyn and R. M. Aarts, "Personal Sound," *J. Acoust. Soc. Am.*, vol. 96, no. 5 Supplement, p. 3281 (1994 Nov.). <https://doi.org/10.1121/1.410932>.
- [13] L. Hermannsen and S. Bech, "Perception of Audio Quality and Audio-on-Audio Interference in Sound Zones," in *Proc. 14th Intl. Conf. Quality of Multimedia Experience (QoMEX)*, pp. 1–4 (Lippstadt, Germany) (2022 Sep.). <https://doi.org/10.1109/QoMEX55416.2022.9900892>.
- [14] L. Hermannsen and S. Bech, "Perceptual Evaluation of Dynamic Sound Zones, Part 1: Audio Quality of Experience in a Simulated Home Environment," *J. Audio Eng. Soc.*, vol. 73, no. 10, pp. 601–618 (2025 Oct.). <https://doi.org/10.17743/jaes.2022.0214>.
- [15] K. F. Skov, P. Axel Nielsen, and J. Kjeldskov, "Tuning Shared Hospital Spaces: Sound Zones in Healthcare," in *Proc. 18th Intl. Audio Mostly Conf.*, pp. 63–70 (Edinburgh, UK) (2023 Aug-Sep.). <https://doi.org/10.1145/3616195.3616198>.
- [16] K. F. Skov, "Intermediate Care Unit Ambience (3rd Order Ambisonics)," (2024 Feb.).
- [17] J. Francombe, R. Mason, M. Dewhirst, and S. Bech, "Elicitation of Attributes for the Evaluation of Audio-on-Audio Interference," *J. Acoust. Soc. Am.*, vol. 136, no. 5, pp. 2630–2641 (2014 Nov.). <https://doi.org/10.1121/1.4898053>.
- [18] V. Pulkki and M. Karjalainen, *Communication Acoustics: An Introduction to Speech, Audio and Psychoacoustics* (John Wiley & Sons, Chichester, UK, 2015). <https://doi.org/10.1002/9781119825449>.
- [19] ITU-R, "Algorithms to Measure Audio Programme Loudness and True-Peak Audio Level," *Recommendation ITU-R BS.1770-5* (2023 Nov.).
- [20] L. Farina, "Simultaneous Measurement of Impulse Response and Distortion With a Swept-Sine Technique," in *Proc. 108th Conv. Audio Eng. Soc.* (2000 Feb.), paper 5093. <https://aes.org/publications/elibrary-page/?id=10211>.
- [21] V. Hansen, and G. Munch, "Making Recordings for Simulation Tests in the Archimedes Project," *J. Audio Eng. Soc.*, vol. 39, no. 10, pp. 768–774 (1991 Oct.). <https://aes.org/publications/elibrary-page/?id=5961>.
- [22] T. Pedersen and N. Zacharov, "The Development of a Sound Wheel for Reproduced Sound," in *Proc. 138th Conv. Audio Eng. Soc.* (2015 May), paper 9310. <https://aes.org/publications/elibrary-page/?id=17734>.
- [23] H. T. Lawless and H. Heymann, *Sensory Evaluation of Food: Principles and Practices*, Food Science Text Series (Springer, New York, NY, 2010), 2nd ed. <https://doi.org/10.1007/978-1-4419-6488-5>.
- [24] A. Kuznetsova, R. H. Christensen, C. Bavay, and P. B. Brockhoff, "Automated Mixed ANOVA Modeling of Sensory and Consumer Data," *Food Qual. Prefer.*, vol. 40, pp. 31–38 (2015 Mar.). <https://doi.org/10.1016/j.foodqual.2014.08.004>.
- [25] J. Rämö, S. Bech, and S. H. Jensen, "Validating a Real-Time Perceptual Model Predicting Distraction Caused by Audio-on-Audio Interference," *J. Acoust. Soc. Am.*, vol. 144, no. 1, pp. 153–163 (2018 Jul.). <https://doi.org/10.1121/1.5045321>.
- [26] L. Hermannsen and S. Bech, "Perceptual Evaluation of Dynamic Sound Zones, Part 3 - Audio Quality of Experience in Real and Simulated Home Environments," *J. Audio Eng. Soc.* (under review) (May 2026).
- [27] A. S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound* (The MIT Press, Cambridge, MA, 1990). <https://doi.org/10.7551/mitpress/1486.001.0001>.
- [28] H. Fastl and E. Zwicker, *Psychoacoustics: Facts and Models*, Springer Series in Information Sciences (Springer-Verlag, Berlin, Germany, 2007), 2nd ed. <https://doi.org/10.1007/978-3-540-68888-4>.
- [29] J. C. Castura, L. Antúnez, A. Giménez, and G. Ares, "Temporal Check-All-That-Apply (TCATA): A Novel Dynamic Method for Characterizing Products," *Food Qual. Prefer.*, vol. 47, pp. 79–90 (2016 Jan.). <http://dx.doi.org/10.1016/j.foodqual.2015.06.017>.
- [30] B. Rivet, W. Wang, S. M. Naqvi, and J. A. Chambers, "Audiovisual Speech Source Separation: An Overview of Key Methodologies," *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 125–134 (2014 May). <https://doi.org/10.1109/msp.2013.2296173>.
- [31] S. Agrawal, S. Bech, K. De Moor, and S. Forchhammer, "Influence of Changes in Audio Spatialization on Immersion in Audiovisual Experiences," *J. Audio Eng. Soc.*, vol. 70, no. 10, pp. 810–823 (2022 Oct.). <https://doi.org/10.17743/jaes.2022.0034>.

THE AUTHORS



Line Hermannsen



Kasper Fangel Skov



Søren Bech

Line Hermannsen has an M.Sc. and Ph.D. from the Department of Bioscience, Aarhus University, Denmark (Marine Bioacoustics Lab and Marine Mammal Research) in underwater noise pollution and its effects on marine mammals, including acoustic impacts on mammalian hearing. From 2020 to 2023 she was a postdoctoral fellow at Bang & Olufsen Research and affiliated with the Department of Computer Science at Aalborg University, Denmark, working in the Interactive Sound Zones for Better Living (ISOBEL) project to investigate perception and optimization of sound zones that can create better acoustic environments. Since 2023, she has been working within marine biology at the Technical University of Denmark (DTU) (Section of Coastal Ecology, DTU Aqua, National Institute of Aquatic Resources, Denmark), focusing on coastal ecology, biodiversity, human-sea interactions and bioacoustics.

Kasper Fangel Skov holds an M.Sc. from the Department of Digital Design and Information Studies, Aarhus University, and a Ph.D. from the Department of Computer Science, Aalborg University, Denmark. During his Ph.D. studies, he conducted user experience research on sound

zone technology as part of the Interactive Sound Zones for Better Living (ISOBEL) project. Since 2024, he has been an assistant lecturer at Sonic College, University College South Denmark (UC SYD), Denmark.

Søren Bech retired by the end of 2024 as Director of Research at Bang & Olufsen and is now Director of Research at Systematic. He is also a professor of audio perception at Aalborg University. He received an M.Sc. and Ph.D. from the Department of Acoustic Technology of the Technical University of Denmark. Dr. Bech has authored more than 52 peer-reviewed journal articles and more than 100 conference contributions. He has been principal investigator (PI) of 19 funded (European Union and national) international collaborative research projects. He is a fellow of the Acoustical Society of America and Audio Engineering Society (AES), past governor and vice-President of the AES, and associate technical editor of the AES journal. His research interest includes psychoacoustics, in particular human perception of reproduced sound in smaller rooms. Other interests include experimental procedures and statistical analysis of data from sensory analysis of audio and video quality.