Perception of Audio Quality and Audio-on-Audio Interference in Sound Zones

Hermannsen, Line; Bech, Soren

Published in:
2022 14th International Conference on Quality of Multimedia Experience, QoMEX 2022

DOI (link to publication from Publisher):
10.1109/QoMEX55416.2022.9900892

Publication date:
2022

Document Version
Accepted author manuscript, peer reviewed version

2022 14th International Conference on Quality of Multimedia Experience (QoMEX)

Perception of Audio Quality and Audio-on-Audio Interference in Sound Zones

Line Hermannsen*†, Søren Bech*‡
* Bang & Olufsen A/S, Struer, Denmark, lihe@bang-olufsen.dk | sbe@bang-olufsen.dk
† Department of Computer Science and ‡ Department of Electronic Systems, Aalborg University, Aalborg, Denmark

Abstract **— Audio playback from multiple media devices simultaneously can cause chaotic soundscapes. Sound zones have the potential to minimize audio-on-audio interference by directing sound to intended users. Yet to create good audio experiences, sound zone systems should be adapted to fit different user scenarios and needs. Evaluation of Quality of Experience (QoE) can help to understand and meet user needs, but first requires knowledge of the key characteristics that are relevant to describe the perception of sound zones. Here, we evaluate potential perceptual attributes of a newly developed sound zone system in a listening test with naïve assessors. In total, 8 out of 17 attributes were found to be significantly relevant to distinguish between presented stimuli. Seven of the attributes had a high understandability (score of 89-95/100), while one attribute was poorly understood by half the assessors, suggesting a need for an improved attribute definition. These findings deliver important knowledge on the perceptual attributes that are relevant for evaluating the QoE of sound zone systems, which will allow for optimization and adaptation of sound zones to improve future listening experiences.**

Keywords - Sound Zone System, Quality of Experience, Audio-on-Audio Interference, Perceptual Attributes

I. Introduction

Noise is unwanted sound that can interfere with the ability to detect signals of interest, and exposure to noise can cause deleterious effects on our health and well-being [1]. Following the expanding use of personal media devices, unwanted sound in our daily lives is increasingly due to audio-on-audio interference resulting from individual preferences for media content. Headphones offer a solution to minimize noise exposure and enable a better signal-to-noise ratio (SNR) of a desired signal. However, a disadvantage of headphones is that potential signals of interest from the surroundings are more difficult to detect, i.e., the listener isolates themself from the environment around them, which is often not desirable. Fortunately, there is a way to create individual listening experiences without headphones, namely sound zones. Sound zones are created by utilizing constructive and destructive interference between sounds from multiple sources in an array to direct sound to a specific area in a room [2]. A good separation between sound zones is key to enable a sufficient SNR for each listener. Yet optimal listening experiences not only require a good SNR, but also a high audio quality of the reproduced sound. High audio quality is often linked to reproduction of relevant low-frequency sounds (bass) [3] made up of long wavelengths, which are challenging to control in a sound zone system. This pinpoints one of the main challenges for sound zone systems: the trade-off between minimizing interference (i.e., having a good separation) and improving audio quality. To optimize sound zones, it is therefore necessary to first understand which key characteristics influence the perceived Quality of Experience (QoE) of sound zone systems in different scenarios. Perceptual differences between audio stimuli can be evaluated using consensus vocabulary made up of descriptive words called 'attributes' [4]. Here, we conduct a listening test to find perceptual attributes that are relevant for describing audio experiences in sound zones in a domestic environment. The attribute list resulting from this study will be used in further studies within the ISOBEL project[1] to understand how to quantify and thereby optimize QoE of a sound zone system.

978-1-6654-8794-8/22/$31.00 ©2022 IEEE

II. Methods

A. Experimental facilities

The listening test was conducted in a 60 m^2 sound-insulated laboratory at the Department of Electronic Systems, Aalborg University, Denmark. The background noise level in the laboratory when the sound system was active, but not playing any audio, was 50 dB(A) (measured over 16 minutes) recorded with a Head and Torso Simulator (HATS, Brüel & Kjær) with built-in ear microphones (-28.7 dBV sensitivity at 1 kHz/94 dB SPL) placed in the target sound zone at the assessor's position.

B. Sound system

A custom-built soundbar, consisting of 14 midrange 5" drivers and 45 fullrange 1.5" drivers, was driven individually through ten IcePower 150ASH7 class D amplifiers connected to three MOTU 24 I/O soundcards with external processing on a stationary PC running MATLAB version 2016a. The sound zone system played back directional sound in two sound zones, and/or sound played out at wide angles of ±90º in the room. The two sound zones were created with delay and sum beamforming and sent out at angles of ±9.9º relative to the center of the soundbar. All drivers were active in all configurations.

C. Listening test

One assessor participated in the listening test at a time. The assessor was seated in a sofa 3 m in front of the soundbar, which focused the target sound zone towards the assessor (Figure 1). The user interface (UI) for the perceptual evaluations followed the experimental design by [5] and were generated with a custom-built MaxMSP patch mirrored to a tablet via the app Mira. On each page in the UI, the attribute in question was presented together with a description of it at the top. Assessors were then asked to rate "*How well do you understand the attribute?"* on a horizontal intensity scale (0-100) showing only the low anchor description at a value of 10 ("*I really do not understand what the attribute means")*

[1] ISOBEL, Interactive SOund zones for BEtter Living, www.isobel.dk.

and the high anchor description at a value of 90 ("*I completely understand what the attribute means*"). The attribute relevance was evaluated by presenting five audio stimuli (A-E) that the assessor listened to and then answered the question "*Can the differences in audio be described with the given attribute?*" (Yes/No). Since the tolerated level of interference is linked to the intended listening task of the listener [6], the test included three user scenarios with audio stimuli relating to either 'Rest', 'Entertainment' or 'Information Gathering'. The participant was introduced to the given scenario by a short text (translated from Danish): "*Imagine that you are in a hospital bed and need to rest after treatment*" ('Rest')/*"Imagine that you are at home relaxing to music"*('Entertainment')/*"Imagine that you are at home and listening to understand and convey the information in the audio clip*" ('Information Gathering').

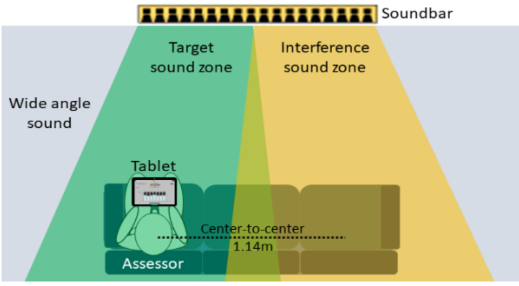

Figure 1. Schematic view of the listening test setup to evaluate perceptual attributes for audio quality and audio-on-audio interference of a sound zone system.

D. *Audio stimuli*

A total of six 55-60 s audio clips were used as target audio for the listener. The one-minute length of audio clips was chosen to allow for sufficient time to listen and compare different stimuli and evaluate them for the given attribute. Two target audio clips were presented within each of the three listening tasks. Both audio clips within a listening task were evaluated before continuing to the next listening task. The presentation order of listening tasks, as well as the presentation order of the two audio clips within a given listening task, were randomized for each participant.

The target audio was either calm instrumental music ('Rest'[2]), pop/rock or classical music ('Entertainment'[3]) or focused on a speech signal in a podcast or radio news ('Information Gathering'[4]). Each target audio clip was played back in five different configurations, denoted stimuli A-E. The assignment of letter A-E to the five stimuli of a target audio clip was randomized on each UI page. Three of the stimuli contained only the target audio (i.e., no interfering audio) and were played either (1) in a wide angle without directed sound; (2) in a wide angle and with directed sound in sound zone A; or (3) in a wide angle and with directed sound in both sound zones, adding a 6 dB gain to the target audio. A 6 dB gain was chosen as this reflects a doubling in the physical loudness [9], i.e., an audible change in loudness and in the SNR of the target and interfering audio. For the two remaining stimuli, an interfering audio clip was introduced in the sound zone next to the assessor, which either belonged to (4) the same listening task; or (5) a different listening task than the target audio. This was done to reflect realistic use case scenarios and to create scenarios with differences in audio in the two sound zones by different combinations of vocal music, instrumental music and speech signals. Six interference audio clips, each with a length of 55-60 s, were used. Whenever a new stimulus button was pressed in the UI, a continuous listening experience was created by changing the configuration whilst maintaining the same playback position in the audio.

The playback levels of stimuli were determined for each listening task in a small experiment prior to the listening test with 6 assessors (one woman, five men, age 25-30), asking them to set their preferred volume if listening for 30 minutes to an audio clip. Assessors were presented with each of the 12 audio clips and could adjust the volume with a custom-built controller. Preferred levels were found to be significantly lower (pairwise comparisons using Wilcoxon rank sum test, $p<0.001$) with 3 dB for 'Rest' audio clips (mean of 52.5 dB rms) compared to 'Entertainment' and 'Information Gathering' audio clips (mean of 55.5 dB rms). The sound pressure levels of transmitted audio were therefore set at 52.5 and 55.5 dB (+ 6 dB gain for configuration 3 stimuli) in the listening test, calibrated at the assessor position before each test with a Brüel & Kjær Analyzer 2270.

E. *Perceptual attributes*

Relevant attributes for audio characteristics and audio-on-audio interference were included based on the literature [4,7,10]. The attribute *Understandable* ("*It is easy to understand, what is being said in the audio clip*") was added to cover speech intelligibility, resulting in an initial list of 52 attributes. Next, the attributes were evaluated by two expert listeners (tonmeisters) and the attribute list was further shortened for each listening task before the test. The final attribute list included 16 attributes for 'Rest', 14 attributes for 'Entertainment' and 5 attributes for 'Information Gathering'. Since two target audio clips were used for each of the three listening tasks, each attribute was presented twice for a given task. The listening tests were conducted in Danish and the attributes and their descriptions that were originally in English were therefore translated to Danish. The order of attributes was randomized in the test software. When an attribute was first presented, it had to be evaluated for both target audio clips within a listening task (i.e., two UI pages), before moving to the next attribute. This was done to minimize the number of times that the assessor would have to adapt to a different attribute. Prior to the listening test, the assessor was familiarized with the setup and UI in a short session playing 10 s segment of each of the 12 target and interference audio clips.

F. *Assessors*

Twelve naïve[5] assessors (four women, eight men, age 21-25) participated in the listening tests during January and February 2021. All assessors had their hearing tested[6] prior to the listening test and were evaluated to have normal hearing (< 20 dB hearing level criterion). Assessors gave informed consent and were paid for their participation.

[2] All music for 'Rest' were composed by Wavecare Aps to facilitate rest.
[3] All music used for 'Entertainment' were from [7]. [4] 'Information Gathering' audio were from the Danish Broadcasting Corporation (DR) radio archives (www.dr.dk) and from the Archimedes project [8]. [5] Naïve = no technical audio training and no prior listening test experience.

[6] Hearing tests were conducted with an Orbiter 922 (Madsen Electronics) presenting six pure tones in the frequency range 250-4000 Hz at sound pressure levels of 0-40 dB in increments of 5-10 dB.

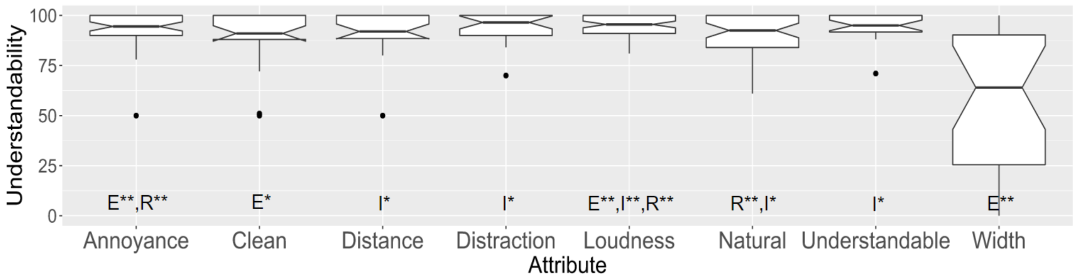

Figure 2. Box plots of perceived understandability (where 100 is the highest possible understandability) across all assessors and listening tasks for the attributes that were found to be significantly relevant for at least one listening task. The listening task(s) (E = Entertainment, I = Information Gathering, R = Rest) where an attribute was found to be significantly relevant (*<0.05, **<0.01) are shown below each box plot. Box plot notches indicate the 95% confidence interval of the median. Black dots are outliers.

III. Results

A. Attribute relevance

The relevance of presented attributes for perceptual evaluation of sound zones was analyzed with one-sided binomial tests (*binom.test*) in R version 1.4.1103. A significant result indicates that an attribute was selected as relevant significantly more often than at chance frequency (Figure 2). The tests used a significance level of 0.05 and a Bonferroni-Holm correction of p-values (*p.adjust*) to minimize type I errors (false positives). For the listening task 'Entertainment',4 of the presented 14 attributes were found to be significantly relevant for distinguishing between the presented audio stimuli: *Loudness, Clean, Width* and *Annoyance*. The remaining attributes *Balance, Boxy, Canny, Distance, Distraction, Distorted, Envelopment, Full, Tinny, Natural* were not found to be significantly relevant for the presented 'Entertainment' stimuli. For audio stimuli within the task 'Information Gathering', all five attributes *Loudness, Natural, Distance, Understandable* and *Distraction* were found to be significantly relevant. For 'Rest', only 3 out of 16 attributes were found to be significantly relevant: *Loudness, Natural* and *Annoyance.* Attributes *Calming, Chaotic, Clean, Distraction, Distorted* were not perceived as relevant significantly more often than at random for 'Rest' stimuli. Neither were the timbral attributes *Boxy*, *Canny*, *Full* and *Tinny,* or the spatial attributes *Balance, Distance, Envelopment* and *Width [4,10].*

B. Attribute understandability

Across all assessors, 7 of the 8 significantly relevant attributes were rated with an understandability between 50 and 100, with means of 89-95 (Figure 2), where a score of 100 reflects a full understanding of the given attribute. The attribute *Width* stood out by having understandability ratings ranging between 0 and 100, with five assessors scoring below 50 (with a mean score of 22) and one assessor rating *Width* as not at all understandable (i.e., a score of 0) within the 'Rest' task.

IV. Discussion and conclusion

This present study found that attributes related to loudness (*Perceived loudness*), transparency (*Clean* and *Natural*) and interference (*Distraction* and *Annoyance*) were significantly relevant in perceptual evaluation of target and interference audio played in two sound zones across three listening tasks: 'Rest', 'Entertainment' and 'Information Gathering'. The same two interference attributes, *Distraction* and *Annoyance,* were also previously found to be highly relevant for audio-on-audio scenarios [7], highlighting the importance of a good separation between sound zones to increase SNR and thereby improve the listening experience. A high SNR is particularly important when you need to process and understand a speech signal. Accordingly, attributes related to SNR (*Understandable* and *Distance*[7]) were found to be relevant for the perception of audio stimuli focused on speech ('Information Gathering'). Furthermore, the requirement of a higher SNR to decode speech signals could explain why the attribute *Distraction* is only found significantly relevant for 'Information Gathering' stimuli. Interfering audio introduced during the more demanding task of understanding speech could cause distraction and prevent the assessor from performing the intended task, whereas the task of listening to music would likely be less inhibited by an interferer, and the perceived interference would instead be expressed with an annoyance response, as seen for 'Rest' and 'Entertainment' stimuli. None of the attributes related to timbre (*Tinny, Canny, Boxy, Full* in [4,10]) were found to be significantly relevant for distinguishing stimuli in any of the listening tasks, reflecting that these characteristics were perceptually similar (or even absent) across the different stimuli. Out of four attributes related to the spatial characteristics of sound, only the attribute *Width* was found to be relevant for 'Entertainment' stimuli. However, since the understandability of *Width* was found to be relatively low, despite its definition in [10] being a result of attribute elicitation and consensus among both naïve and expert listeners, we recommend that the description of this attribute is reformulated to increase understandability in future studies. The remaining significantly relevant attributes for the given stimuli were rated to have a high understandability by all assessors and require no reformulation.

This study presents the perceptual attributes relevant for the perception of audio quality and audio-on-audio interference in sound zones in a domestic environment. This study is the first in a series of studies to investigate the perceptual attributes of sound zones and assess how these attributes can be used to predict QoE in different scenarios, which will enable optimization of personal listening experiences in sound zones.

Acknowledgements

This work is part of the ISOBEL project funded by the Innovation Fund Denmark. We thank M.B. Møller, P. Nielsen, A.S.S. Larsen and C.V. Skipper for help with the experimental work, and J. Francombe for proofreading the manuscript. A special thanks to the participants.

[7] The perceived distance of audio is tightly linked to SNR, as SNR changes with changing distances between the listener and the sound source.

References

[1] World Health Organization, WHO (2018). *Environmental noise guidelines for the European region*. Report, pp. 1-181.

[2] Druyvesteyn, W. F., Aarts, R. M., Asbury, A. J., Gelat, P., and A. Ruxton (1994). *Personal sound*. Proceedings of the Institute of Acoustics, vol. 16, pp. 571-585.

[3] Freeman, J. & Lessiter, J. (2001). *Here, There and Everywhere: The Effects of Multichannel Audio on Presence.* Proceedings of the 2001 International Conference on Auditory Display, Espoo, Finland, July 29-August 1, 2001, pp. 231-234.

[4] Pedersen, T. H. and N. Zacharov (2015). *The development of a Sound Wheel for Reproduced Sound*. Audio Engineering Society Convention Paper 9310, pp 1-13.

[5] Hicks, L., Moulin, S. and S. Bech (2018). *Sensory Profiling of High-End Loudspeakers Using Rapid Methods-Part 3: Check-All-That-Apply with Naïve Assessors.* Journal of the Audio Engineering Society, vol. 66, no. 5, pp. 329-342.

[6] Francombe, J. (2014). *Perceptual Evaluation of Audio-on-Audio Interference in a Personal Sound Zone System.* PhD thesis, University of Surrey, Guildford, United Kingdom, pp. 24-25. https://openresearch.surrey.ac.uk/esploro/outputs/99511757502346.

[7] Francombe, J., R. Mason, M. Dewhirst and S. Bech (2014). *Elicitation of attributes for the evaluation of audio-on-audio interference.* The Journal of the Acoustical Society of America 136 (5), pp. 1-12.

[8] Bech, S. (1989). *Electroacoustical simulation of listening room acoustics for project ARCHIMEDES.* The journal of the Acoustic Society of America 86, S2.

[9] Gray, L. (2000). *Background Science*, *Properties of Sound.* Journal of Perinatology 20, S5-S10.

[10] Pedersen, T. H. (2015). *Karakteristik af lyden fra audioprodukter*. Report from DELTA, SenseLab, Denmark, pp. 1-29. https://forcetechnology.com/-/media/force-technology-media/pdf-files/unnumbered/senselab/teknotat-beskrivelse-af-lyden-fra-audioprodukter-dk.pdf?la=en.