Personal Sound Zones: Study on the Threshold of Acceptability in an Automotive Environment

Maximilian Schaab, Tobias Clauß, Jakob Bergner, Christoph Sladeczek
Fraunhofer-Institut für Digitale Medientechnologie, 98693 Ilmenau, Email: christoph.sladeczek@idmt.fraunhofer.de

Abstract

Personal sound zones aim to spatially limit the audibility of reproduced audio content. An ideal solution supplies multiple listeners in the same room with different audio signals without the need for headphones or constructional acoustic barriers. These systems however suffer from physical limitations, resulting in crosstalk beyond the borders of a sound zone into another. Especially when reproducing multiple sound zones in small rooms (e.g. within vehicle cabins), significant interferences from adjacent sound zones can diminish the perceptive listening experience. While physical metrics can quantify the magnitude of these interferences, e.g. by measuring the sound pressure level of desired and interfering sounds, the actual impairment of the subjective listening experience can usually not be quantified on that basis. The studies presented in this work show the existence of a program-dependent threshold of acceptability of interfering audio signals as well as indications of different factors influencing this threshold. Furthermore, the relation to auditory masking thresholds will be motivated suggesting the possibility of predicting thresholds of acceptability using masking effects. An ideal predictive model would eventually allow an accurate estimate on the listener's perception in such scenarios by means of physical metrics.

Introduction

In many situations it is desirable to spatially limit the audibility of reproduced audio content without the need for constructional acoustic barriers or headphones. This topic is addressed by systems for the creation of personal sound zones (PSZ). One exemplary application is the creation of multiple sound zones in the passengers cabin of a car. This way, the driver could e.g. exclusively listen to the sound of the navigation system while the passengers could each hear different radio programs that ideally don't interfere with each other. PSZ systems therefore aim at minimizing the crosstalk by maximizing the physical channel separation between adjacent sound zones. Such systems are usually realized by special forms of signal processing combined with loudspeaker arrays for the reproduction. The different approaches can roughly be assigned to two categories. While the methods of the first category try to optimize the effective sound radiation of the loudspeaker array (beamforming) [1], the methods of the second category focus on the analytic or data-based synthesis of a desired sound field at certain positions in space [2]. All approaches however suffer from physical limitations so that the audibility of reproduced audio content is usually not strictly bound to the desired spatial borders of a sound zone. Especially when reproducing multiple sound zones in small rooms, disturbing interferences from adjacent zones (interferer zones) will most likely impair the subjective listening experience in the own sound zone (target zone).

The physical performance of such systems can be quantified by means of different metrics. The extent of interference can be quantified e.g. by the *interferer-to-target ratio (ITR)* expressed in dB values. It represents the ratio of the measured interference sound pressure level (SPL) to the target signal SPL in the target zone. However, no reliable statements on the perceptive impairment can be made on the basis of a simple ITR.

Since the subjective listening experience should always be the main intention of innovative audio systems, focus has to be placed on the perceptual assessment of such systems. Using the level of interference (or the ITR respectively) as a physical metric and considering the mechanisms of the human auditory system, the listening experience can principally be partitioned into three perceptual regions (as shown in fig. 1).

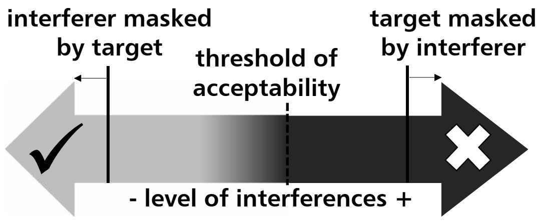


Figure 1: In the left region, the interferences are completely masked by the target signal. In the right region, the target signal is masked by the interferences. In the middle region, the threshold of acceptability separates the listening scenario in subjectively acceptable and not acceptable.

In the very left region, the interfering signal gets completely masked by the target program and is therefore not audible, resulting in a highly negative ITR. This listening scenario is defined as unimpaired. In the very right region, the interferences mask the target completely. Here, the listening scenario is defined as unsatisfactory. In the middle region, where both the target and the interference can be perceived, no statements on the perceived impression can be made on that basis. For this work it was therefore assumed that in this region, a program dependent threshold exists, which divides the perceived listening experience into subjectively acceptable and not acceptable (*threshold of acceptability*, referred to as ITR_A). The literature often refers to acceptability as an useful attribute regarding the perceptual assessment of audio systems [3, 4] or the impact of noise [5]. Based on work by Jumisko-Pyykkö et al. acceptability can be seen as the minimum requirement of a user towards a system regarding his expectations and needs [6]. This justifies the aforementioned assumption of a binary threshold. In this context the threshold of acceptability can thus be considered a minimum requirement with regard to the performance of PSZ systems indicating the minimum required physical channel separation to guarantee an acceptable listening experience. Knowledge about this threshold is therefore crucial for the development of such systems.

Literature review

Research in the area of personal sound zones receives much attention. After Druyvesteyn and Aarts presented the concept in 1994 [7], many authors have published novel approaches for the development of PSZ systems. Especially in the fields of signal processing and optimization of technical parameters a lot of literature can be found (see e.g. [2, 8, 9]). In [9] different physical metrics are described, that can be considered for the technical evaluation of such systems. The acoustic contrast between two zones or the uniformness of a reproduced sound zone are exemplary metrics therein.
Despite the high importance, the perceptual assessment of PSZ systems has received little attention. Baykaner and Francombe each present different approaches for the perceptual assessment of simple audio-on-audio interference scenarios and model attributes of the listener perception in this context [3, 4]. Both authors consider listener acceptance as an important attribute in such scenarios. Besides acceptance, Francombe also considers other perceptual attributes such as listener distraction or annoyance caused by interfering audio programs. Baykaner incorporates the findings of his study into a predictive model which models the threshold of acceptability in such scenarios based on auditory masking effects.
Based on the relevance of these publications for the present work, certain aspects of the methods used will be adapted here.

Perceptual study

This work describes an empirical study which aims at gaining knowledge about some factors that influence the subjective threshold of acceptability within an PSZ scenario. Since it is part of an automotive project at Fraunhofer IDMT, the described study focuses on the assessment of a personal sound zones system installed directly inside a car cabin.

It seems obvious that the threshold of acceptability might be affected by numerous factors besides the actual physical extent of interference (ITR). Since it is a subjective threshold, the listener's preferences and tolerance towards disturbing interferences should play a major role. Besides the expected subjective variance, it is further assumed that the threshold gets strongly influenced by the reproduced programs. Depending on the spectrotemporal properties of the target and interfering audio signals, different thresholds of acceptability seem likely. Furthermore, the impact of road noise on the threshold will be examined.

The following section outlines the experimental methodology and discusses the results subsequently.

Design

Nine subjects were confronted with a PSZ system in the passenger cabin of a SUV-like car (see fig. 2) comprising a target zone playing a target signal and clearly audible interferences coming from an adjacent sound zone. All subjects were native speakers and are considered trained listeners. The car was parked in a empty garage with the engine turned off. The PSZ system comprised two loudspeaker arrays each including 20 speakers that were mounted inside the vehicle at the top of the B-Columns facing the passengers. Using a beam forming algorithm, two sound zones were reproduced over the arrays simultaneously. The system was optimized in a way, that the target zone was positioned at the front passengers seat and the interfering sound zone directly behind at the right back seat, both at ear level.

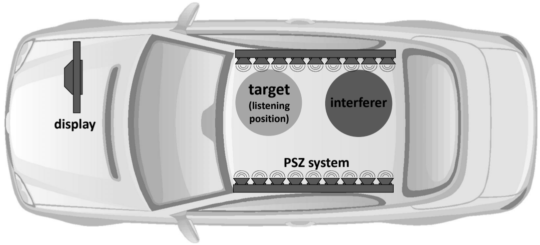


Figure 2: Schematic setup for the threshold of acceptability study.

The level of the interfering sound zone had to be adjusted by the subjects, who were seated at the front passengers seat. They were asked to imagine being passenger on a relaxed ride and to focus on the target signal. The task was to set the interferer signal gain in a way that they would rate the listening impression as still acceptable, while louder interference levels should lead to an unacceptable impression. No feedback about the set level was given except the auditory sensation itself. The resulting difference between interferer and target signal gain was considered the respective threshold of acceptability ITR_A. Since the reproduction system is focused on minimizing interferences in the target zone, it already creates a physical separation between both zones. This frequency dependent value can be approximated by 15 dB and is not included in the given ITR ratings.
The reproduction level of the target program was set to a fixed value of 70 dB LAeq with a time constant of 20 s based on [10] followed by minor subjective adjustments by the author. A display showing the test interface was mounted to the bonnet so that it could be seen by the subjects, showing basic playback options (Play, Pause, Stop, Loop), the current item number and a Next-button for storing the rating and going to the next item.

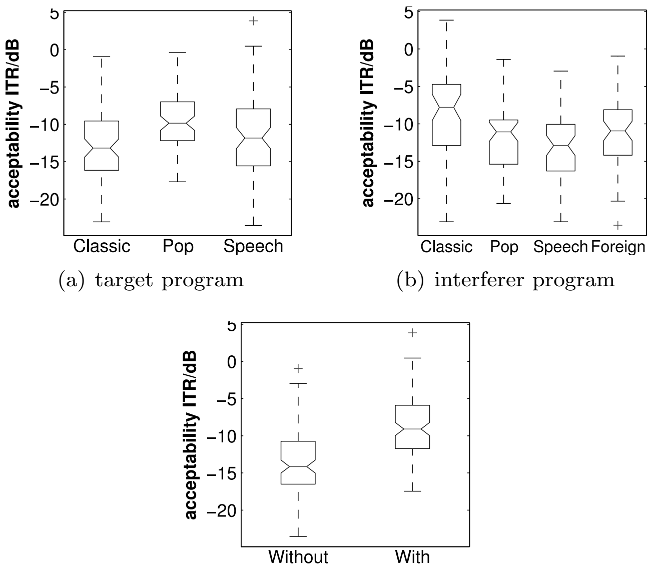
Regarding the test stimuli, certain criteria had to be met. On the one hand, sufficiently long real-world items (actual music or speech signals) were considered reasonable, so that the listener can experience a state of acceptability. On the other hand, the items had to be sufficiently short and dynamically static for single threshold values to be valid. Therefore, for the target and interferer programs 10 s excerpts from pop and classical music pieces, as well as native male speech items (German) were used. The interferer items additionally included a foreign female speech item (Sorbian). While the pop music item was exclusively instrumental comprising guitar, brass and drums, the classical item was dominated by strings. All items were normalized to equal loudness.
To account for the automotive aspect of the experiment, the influence of additional road noise was examined. It comprised real-world interior recordings of a driving van at 80 km/h and was reproduced over the built-in audio system of the test vehicle at a level of 60 dB LAeq [10]. This way, a sufficiently realistic and diffuse sound impression could be realized. The used road noise item was rather static and the spectral characteristics resembled those of pink noise. All test stimuli were once played without and once with additional road noise. In this full factorial design, the item combinations resulted in 3 target · 4 interferer · 2 road noise = 24 unique test stimuli.

Results

Figure 3 shows the resulting ITR_A ratings as boxplots. The lower the ITR value, the more level reduction of the interfering signal was necessary to get an acceptable listening scenario. Each figure (a-c) comprises all 24 ratings from each of the nine subjects (resulting in a total of 216 ratings) and is classified to the respective factors (target program, interferer program and road noise).

It can be seen in a) that interfering audio signals seem to be slightly more acceptable when the target program is spectrotemporally rather dense like the used pop item. Conversely, the level of the interfering signals had to be reduced more, when the target signal was classical music or speech. A repeated measures analysis of variance (RM-ANOVA) combined with a Bonferroni post-hoc test confirmed significant differences between the Pop target and Classic ($p = .027$) or Speech ($p = .005$) respectively (with $\alpha = .05$). The sparser the spectrotemporal properties of the target program, the easier interferences could be perceived. This resulted in more level reduction that is necessary to gain an acceptable impression. Especially the speech item showed short passages of silence following spoken words or sentences leading to clearly audible interferences as can be seen in Figure 4.

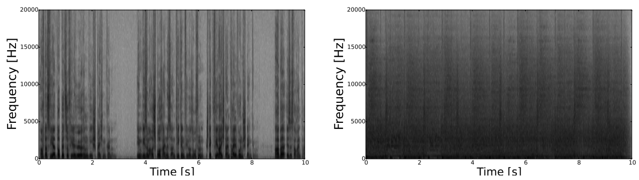
Regarding the interfering programs (fig. 3 b)), the classical music item was apparently more acceptable, indicated by higher ITR_A ratings. The statistical analysis showed significant different mean values between Classic and Pop ($p = .002$), Speech ($p < .001$) and Foreign ($p = .013$). Some subjects commented, that especially higher frequency components (e.g. hissing sounds) and



(a) target program (b) interferer program

(c) road noise

Figure 3: Threshold of acceptability ratings broken down to the three considered factors target program, interferer program and road noise. The center lines of the notches indicate the rated medians, the notches at the side of each box represent the 95 % confidence intervals. The upper and lower box limits represent the first and third quartile respectively.



(a) Speech item (b) Pop item

Figure 4: Spectral representation of the used Speech and Pop target items (Time vs. Frequency). Darker areas in the spectrograms indicate passages of high energy.

transients in the interferer (mainly present in the pop and speech items) had a negative impact on the listening experience. This explains, why the pop and speech interferer items show comparatively low ITR_A ratings, indicating that they were less acceptable. Furthermore, a small trend can be observed regarding the two speech items (native and foreign speech). For the native speech interferer ("Speech") a slightly lower ITR_A value can be observed compared to the "Foreign" speech item. This gives rise to the assumption, that interfering speech programs have less negative impact on the listener acceptance when the semantic aspect of the speech is missing. When the subjects could understand the meaning of the interfering speech program, slightly more level reduction was necessary to achieve an acceptable impression. However, no significant differences between the native and foreign speech item were found. In [3] more detailed experiments towards the impact of speech intelligibility on the threshold of acceptability can be found.
Figure 3 c) reveals another interesting observation considering the automotive aspect of the experiment. The

ITR_A ratings were about 5 dB higher with road noise present than without. The ANOVA showed highly significant differences ($p < .001$). This means, that the level of the interfering signals had to be reduced less, when additional road noise was reproduced. This in turn indicates, that in an automotive context PSZ systems benefit from road noise in terms of masking interfering signals making the necessary channel separation less critical.
The lower 95 % percentile of all ratings is, in combination with the approximated 15 dB physical channel separation of the array, at -38 dB ITR. This means that in 95 % of the considered cases an absolute level difference between target and interfering signal of 38 dB would lead to an acceptable listening scenario. A similar value was found in [4].

Conclusion

This work outlined an empirical study which aimed at gaining knowledge about the threshold of acceptability regarding personal sound zones in an automotive environment. As expected, the results suggest that the necessary level difference to achieve an acceptable listening scenario depends strongly on the reproduced audio content. The spectrotemporal properties of target and interferer signal seem to strongly influence the listener tolerance towards audible interferences. The semantic content of speech programs might have an impact on the threshold of acceptability, as well. In an automotive environment, PSZ systems also benefit from road noise in terms of masking interfering signals.
Since the threshold of acceptability is bound by the auditory masking thresholds (see fig. 1) it seems likely that a deeper relation exists between them. A study on the masking thresholds in an equivalent scenario could reveal a relation between both thresholds that could support the development of predictive models. These models could eventually allow the algorithmic prediction of the threshold of acceptability in such scenarios as shown in [3, 11]. This could in turn help the developers of PSZ systems to easily assess the quality of their products in terms of auditory perception and reduce the need for time-consuming and expensive listening tests. This work lays a foundation for the development of such predictive models by exploring the impact of different factors on the listener acceptance.

References

[1] Van Veen, B.D. and Buckley K.M.: Beamforming: A versatile approach to spatial filtering, IEEE ASSP Magazine, vol. 5, no. 2, pp. 4–24, 1988.

[2] Olik, M.; Francombe, J.; Coleman, P.; Jackson, P.J.B.; Olsen, M.; Møller, M.; Mason, R.; and Bech, S.: A Comparative Performance Study of Sound Zoning Methods in a Reflective Environment, 2013.

[3] Baykaner, K.R.: Predicting the Perceptual Acceptability of Auditory Interference for the Optimisation of Sound Zones. Dissertation, University of Surrey, August 2014.

[4] Francombe, J.: Perceptual Evaluation of Audio-on-Audio Interference in a Personal Sound Zone System. Dissertation, University of Surrey, Guildford, UK, August 2014.

[5] Bishop, D.E.: Judgments of the relative and absolute acceptability of aircraft noise, *The Journal of the Acoustical Society of America*, vol. 40, no. 1, pp. 108–122, 1966.

[6] Jumisko-Pyykkö, S.; Malamal Vadakital, V.K. and Hannuksela, M.M.: Acceptance threshold: A bidimensional research method for user-oriented quality evaluation studies, *International Journal of Digital Multimedia Broadcasting*, vol. 2008, no. 1, pp. 1–20, 2008.

[7] Druyvesteyn, W. F. and Aarts, R.M.: Personal sound, *The Journal of the Acoustical Society of America*, vol. 96, no. 5, p. 3281, 1994.

[8] Coleman, P.; Jackson, P.; Olik, M. and Pedersen, J.A.; eds.: Optimizing the Planarity of Sound Zones, 2013.

[9] Møller, M.B. and Olsen, M.: Sound Zones. Master's thesis, Technical University of Denmark, Lyngby, 2011.

[10] Benjamin, E. and Crockett, B.: Preferred listening levels in the automotive environment," in *Audio Engineering Society 119th Convention*, 2005.

[11] Schaab, M.: Psychoakustische Bewertung von privaten Hörzonen. Master's thesis, Technical University of Ilmenau, Germany, 2017.