EUROREGIO
BNAM2022
Joint Acoustics Conference
9th-11th of May, Aalborg, Denmark

Effect of Wireless Transmission Errors on Sound Zone Performance at Low Frequencies

Christian Sejer Pedersen[1,*], Martin Bo Møller[2], Jan Østergaard[1]

[1]Department of Electronic Systems, Aalborg University, Aalborg, Denmark.

[2]Bang & Olufsen A/S, Struer, Denmark.

*cp@es.aau.dk

Abstract
Sound zones can provide different audio content to several persons in the same room by creating separate zones of individual audio content with minimal interference between the zones. Optimal strategies for creating sound zones depend on the wavelength of the sound. For low frequencies it is advantageous to distribute a number of woofers in the room and use sound field control based on measured transfer functions. For such a distribution of woofers it is desirable to stream the audio signals wirelessly to the woofers, but wireless streaming to multiple receivers can be challenging since transmission artefacts may occur. As the sound field control depends on the superposition of the sound from all the woofers, any errors in the reproduced sound will not only affect the audio quality, but also have a negative effect on the sound zone generation. This study investigates relevant types of degradation that transmission errors can have on the sound zone performance for a specific sound zone setup with eight woofers. Errors such as transmission delays and data dropouts are simulated. It is demonstrated that all these transmission artefacts lead to decreased contrast between the sound zones which is mainly caused by increased leakage to the dark zone. A mitigation strategy where the affected woofer is excluded from the sound zone control shows promising results with slightly lower contrast than the error-free system, but with much better performance than the system affected by transmission errors.

Keywords: sound zones, sound field control, wireless transmission, low frequencies

1 Introduction

Sound zones can be used to provide different audio content to several persons in the same room [1]. It uses multiple loudspeakers and sound field control to create a bright zone, where a specific audio content is wanted and a dark zone, where that specific audio content is unwanted i.e. it should preferable be inaudible. This is often practically impossible so some leakage from the bright zone to the dark zone is expected. The leakage between the zones is quantified by the contrast, which is the level difference between the bright and dark zone. By creating bright and dark zones for the different audio content the personal sound zones can be superimposed on each other.

There are different solutions for creating sound zones and the most suitable solution depend on the wavelength of the reproduced sound. For high frequencies the directivity of drivers is suggested [2], while beamforming may be suitable at mid frequencies. At low frequencies the omnidirectional nature of low-frequency drivers and low absorption at room boundaries means that a suitable method is to distribute several woofers in the room and use sound field control to minimize the mean square error between desired sound pressure and the actual sound pressure in the two zones (also referred to as pressure matching [1]). The sound field control method is based on a feedforward control scheme, that assumes that all physical characteristics remain linear and time invariant. Any violation of these assumptions can lead to deviation between the target and the actual sound fields in the bright and the dark zone and therefore lower contrast. For example, physically moving one or more woofers would cause considerable performance loss [3] and it is important to keep the woofers in their linear amplitude range [4]. Even physical characteristics like temperature changes can potentially affect the contrast performance [5].

As the woofers are distributed in the room it is most convenient to transfer audio signals wirelessly – especially for domestic use, where cabling may give practical challenges. But low latency streaming of audio or video over a wireless network to multiple receivers can be challenging [6]. Due to multipath and Doppler effects, the channel conditions become time varying and the instantaneous transmission rates, jitter, and packet loss probabilities, can be hard to determine [7]. It is generally necessary to use buffering [8] [9] [10] and forward error control strategies [11] to ensure smooth playback at the receivers. In [12] it was furthermore shown that that a joint design of the buffer size and forward error correction codes can lead to a reduced delay in real-time streaming applications. A complementary approach to forward error correction is packet loss concealment [13] [14]. At the expense of a longer delay, interleaving techniques can be advantageous in order to deal with packet burst errors [15]. It is not the purpose of this study to simulate these complex errors, but rather to quantify the effect on the sound zones in simple scenarios of missing one woofer channel, having one woofer channel delayed, and having one woofer channel losing a packet of audio (data dropout) at different error rates. How these simple scenarios reflect real transmission errors depend very much on the system architecture and the transmission scheme and codec used.

This study uses a specific laboratory setup with eight woofers which is used in a larger research project [16] and simulate simple transmission errors to one of the woofers in order to evaluate the degradation in the performance of the sound zones. If the sound zone system detects transmission errors on a channel, then the mitigation solution of removing this channel from the sound zone generation is evaluated.

2 Method

2.1 Laboratory setup

The sound zone setup is placed in a laboratory room (dimensions 7.0 m x 8.12 m and a height of 3.0 m) at Aalborg University. The room has heavily damped walls and ceiling and a wooden floor with a carpet (reverberation time T20 ranging from 0.6 s at 50 Hz to 0.2 s at 400 Hz). The sound zone setup consists of eight woofers distributed in the room and a soundbar for the mid to high frequencies placed in front of the sound zones at a distance of 2.911 m as shown in Figure 1. The two sound zones are at a left (Zone A) and right (Zone B) seat of a sofa.

The woofers are 10'' woofers in custom made MDF cabinets and are designed to be used for frequencies below 500 Hz enabling different cross-over frequencies depending on the setup. The connections to each woofer are cabled, but for future implementations closer to a commercial system the transmission is expected to be using a wireless system of transferring the audio to the woofers.

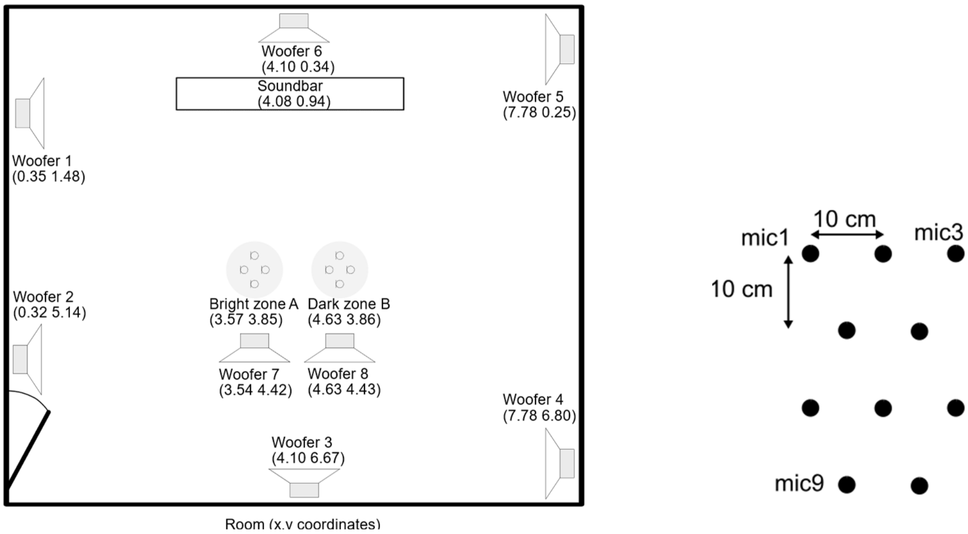

Figure 1: Diagram of the room seen from above. The two zones A and B and the eight woofers 1-8 are shown. The center coordinates in meters are shown in brackets. Right figure shows microphone positions for the transfer function measurements seen from above. Microphone 1-3 are facing the sound bar and two heights (1.051 m and 1.164) are used.

2.2 Sound zone control method

The sound zone control method used in this study is a time domain method that attempts to match a reference target in the bright and the dark zone while minimizing the mean square error between the reference target and the actual sound [17]. Additionally, the method shapes the envelope of the filters in order to reduce pre- and post-ringing as this is expected to produce better audio quality [18]. It is specifically the method proposed by [18] that is used in this study.

The block diagram including sound zone control filters, w_k, wireless transmission, t_k, transfer functions between woofers and microphone positions, h_k, and their resulting pressures in the microphone positions in the bright zone, p_b, and dark zone, p_d, can be seen in Figure 2. For the laboratory setup the impulse responses of woofer 7 with a modelling delay, $w_k^{(eq)}$, is the reference target function for Zone A due to its close proximity to this zone. The reference target for the dark zone is zeroes i.e., no sound is wanted in the dark zone.

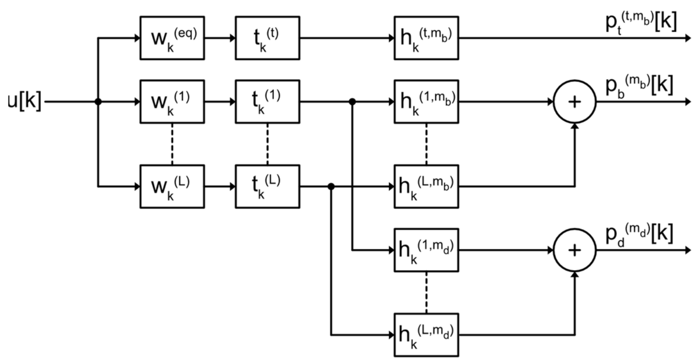

Figure 2: Block diagram of the signal flow from input u[k] to target pressure in the bright zone $p_t^{(t,m_b)}[k]$, reproduced bright zone pressure $p_b^{(m_b)}[k]$ and dark zone pressure $p_d^{(m_d)}[k]$. w_k are the sound zone filters, h_k are the transfer functions from each woofer to each microphone positions in the bright zone, m_d, and microphone positions in the dark zone m_d. The diagram is based on [19] with the addition of the wireless transmission blocks t_k for each woofer 1 to L.

The frequency range of the sound field control focus on the frequencies from 35 to 500 Hz, which is the range of the woofers. The sampling frequency is chosen to 1200 Hz in order to have the most efficient setup for the used frequency range and an integer relation to the 48 kHz used for the playback system and a FIR filter length of 100 taps is chosen to have adequate performance while getting efficient real time processing for the sound zone playback system.

2.3 Woofer transfer function measurements

Transfer functions are measured from each woofer to each 20 microphone positions in the bright and dark zones using the logarithmic swept sine technique [20]. The microphones positions shown in right Figure 1 are ten positions in two different heights of 1.051 m and 1.164 m. This particular layout was chosen in order to have a reasonable approximation of the continuous sound field below 450 Hz over a range of seated ear positions in the zones. The measurements are performed at 48 kHz sampling rate and down-sampled to 1.2 kHz in order to reduce the required processing power both with regard to filter calculation and the actual processing of the audio playback. The complete set of transfer functions are measured twice – one set of for sound zone filter generation and another set for evaluating the performance of the sound field control in order to prevent what is referred to as the “inverse crime” [18].

2.4 Simulation of transmission errors

The transmission errors are simulated both directly with the transfer functions, but also with audio signals processed through the sound zone filters for preparing the audio streams for each woofer and then introducing the errors in t_k on blocks of audio streams to the 20 microphones positions in the bright and dark zone using the scheme shown in Figure 2. The source signal for the plots is gaussian white noise, while for informal listening the source signal is a mono signal from different music signals with bass content.

2.4.1 Simulation of one woofer not playing

The scenario of having one woofer not playing i.e. if it is not turned on or the wireless connection to it is lost is simulated by removing the contribution from that woofer in the summation of the transfer functions (see Figure 2) by effectively setting the impulse response for the affected channel to all microphone positions to zero. For time-domain testing with audio signals the audio stream for the affected channel is omitted in the superposition at the microphone positions.

2.4.2 Simulation of delay in one woofer signal

The scenario of having a synchronization issue resulting in a delay on one woofer is simulated by shifting the contribution from that woofer by integer samples in the summation of the transfer functions (see Figure 2). As the sampling frequency in the simulations is 1200 Hz each sample shift corresponds to a delay of approx. 0.83 ms. For time-domain testing with audio signals the audio stream for the affected channel is delayed by integer samples in t_k .

2.4.3 Simulation of random packet loss on one woofer channel

The scenario of packet loss of audio on one woofer channel is done by preparing the source signal streams and then by random (from uniform probability distribution, independent from previous and future packets) introducing an empty block (packet loss) resulting in data dropout at a predefined error rate in the blocks of streams for the selected woofer channel. For the chosen sampling frequency of 1200 Hz, the block size has been chosen to 24 samples which correspond to 20 ms of signal and simulations of 5%, 10% and 15% packet loss are evaluated.

2.5 Mitigation by removing affected woofer channel if error is detected

If the transmission to a woofer is affected by errors, and there is some way of detecting this, then it is possible to mitigating the error. If a stable delay is detected, then the sound zone system should mitigate this by getting all woofer signals in sync, and the effect of the error will disappear. But packet loss or a missing woofer needs a more drastic solution of removing the affected woofer from the sound zone generation. This requires a new set of control filters based on the seven other woofers. It is expected that this will reduce the performance as compared to an error-free eight woofer solution as there are fewer woofers to control the sound, but it is expected that this solution will perform better than no mitigation.

2.6 Method of evaluating of sound zone performance and transmission error effects

The evaluation is based on the method described in [21] and the performance metrics used in this study is contrast and normalized mean square error as defined in the following.

For the time domain the contrast is following the notation of [19] and is defined as:

$$Contrast[k] = 10log_{10}\left(\frac{M_b^{-1}\sum_{m=1}^{M_b}\left|{p_b}^{(M_b)}[k]\right|^2}{M_d^{-1}\sum_{m=1}^{M_d}|{p_d}^{(M_d)}[k]|^2}\right), \tag{1}$$

where ${p_b}^{(M_b)}[k]$ is the sound pressure at the M_b microphone control points in the bright zone at time sample k and ${p_d}^{(M_d)}[k]$ is the sound pressure at the M_d microphone control points in the dark zone at time sample k. In the frequency domain the contrast is following the same notation but with the time sample k replaced by the angular frequency ω.

The normalized mean square error follows the notation from [19]:

$$nmse[k] = \left(\frac{\|p_b[k] - p_t[k]\|_2^2}{K^{-1}\sum_{k=1}^{K}\|p_t[k]\|_2^2}\right) \tag{2}$$

where $p_t[k]$ is the target pressure in the bright zone and K is the number of time-steps used to evaluate the normalized mean square error.

In addition to the quantitative results also informal listening was performed on one center microphone signal in the bright and dark zone in order to assess the qualitative effects of transmission errors on the sound quality.

3 Results

3.1 Evaluation of one woofer missing

The effect of a missing woofer in the sound zone generation can be seen in left Figure 3. This shows that the loss in contrast is depending much on which woofer is affected and especially woofer 7 and 8, which are close to the sound zones, are important for maintaining contrast between the zones.

It can be seen by comparing right top and bottom plots in Figure 3 that the loss in contrast is mainly due to the increased level in the dark zone as the level in the bright zone is much less affected by the missing woofer 7. This illustrates how sensitive the destructive interference in the dark zone is to violations of the linear time invariance assumptions.

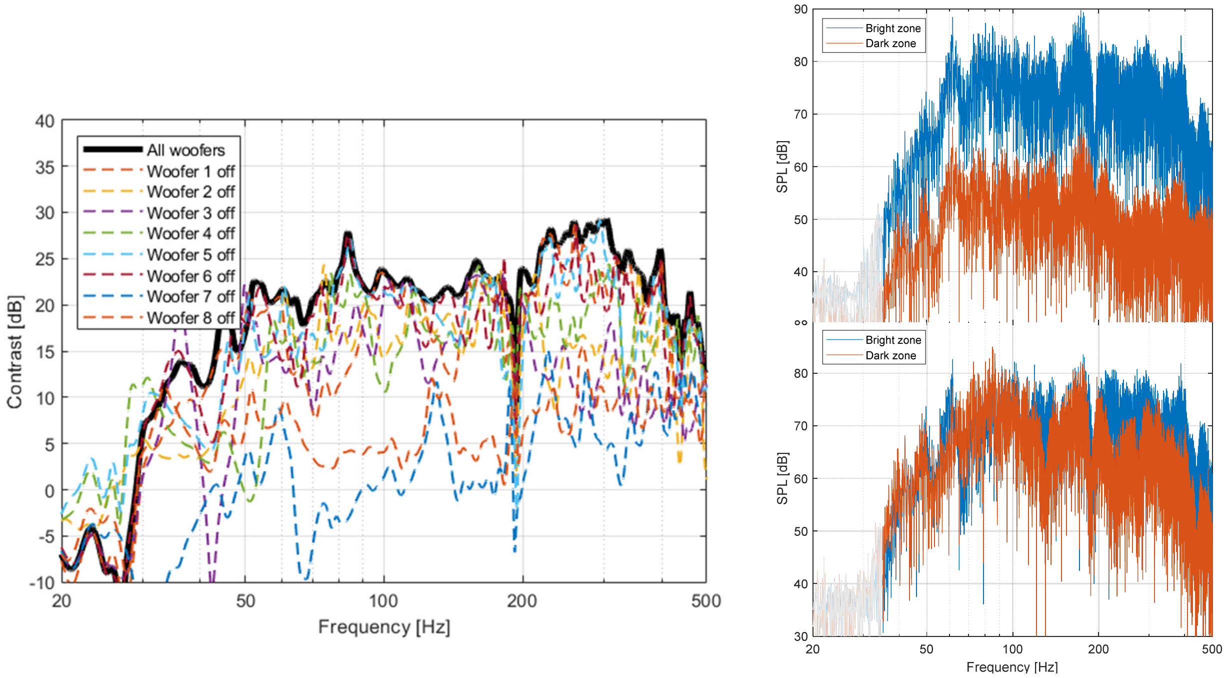

Figure 3: Left: frequency domain mean contrast (Equation (1)) between Zone A and Zone B when all woofers are playing (no error) and the effect of individual woofers not playing. Right: Frequency response of a 10 s gaussian white noise signal in the bright and dark zone; top figure is when there is no error and bottom figure is when woofer 7 is not playing.

3.2 Evaluation of one woofer channel delayed

The effect of having one woofer channel delayed at different integer sample delays for either woofer 5 (best case scenario) and woofer 7 (worst case scenario) can be seen in Figure 4. As expected, the degradation in contrast performance increases with increased delay and as can be expected from the results in Figure 3 it also depends on which woofer is affected.

Left Figure 5 shows the mean time domain contrast depending on delay and affected woofer. In general, the mean contrast drops from zero to two samples delay and is then more or less stable with larger delays, but depends on which woofer is affected.

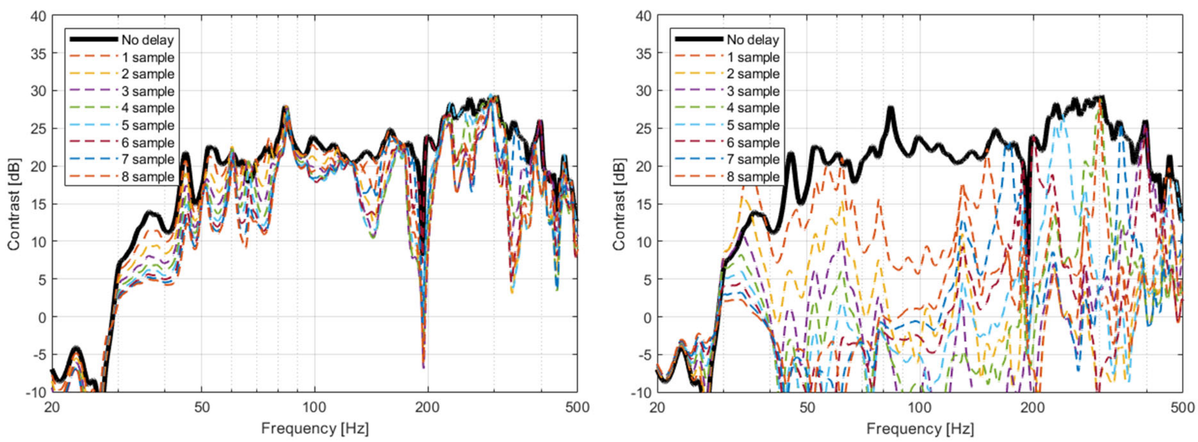

Figure 4: Left is the effect of delay of woofer 5 (far from the sound zones), and right is the effect of delay of woofer 7 on the frequency domain contrast (Equation (1)).

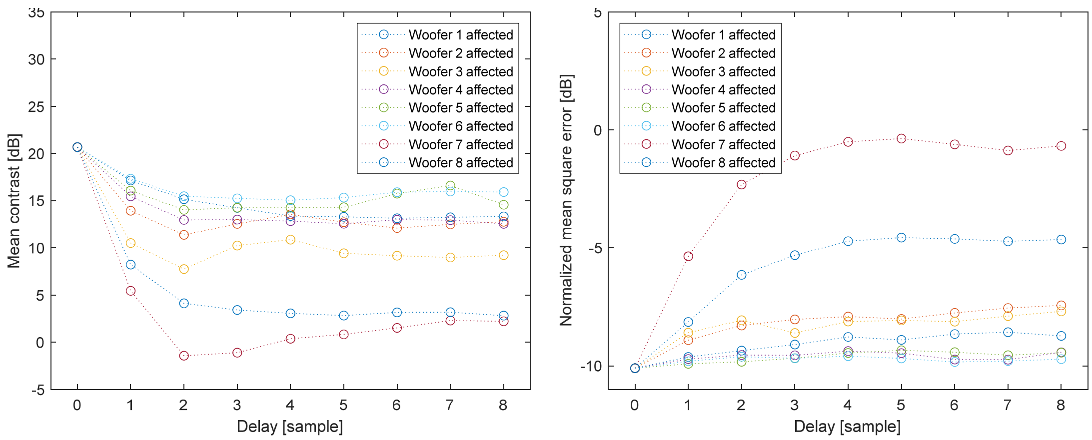

Figure 5: Left shows mean time domain contrast (Equation (1)) and right shows normalized mean square error (Equation (2)) depending on integer sample delay on specific woofer signal (10 s gaussian white noise at 1200 Hz sampling frequency).

The degradation in normalized mean square error between the reference target audio and the resulting audio in different situations of a delayed woofer signal can be seen in right Figure 5. This shows that the normalized mean square error increases with increased delay until approx. four samples delay. As expected, delay in woofer 7 affect the performance in the bright zone the most, as it is closest to the zone and furthermore, it is the transfer functions for this woofer that is used as the desired reference in the sound field control.

3.3 Evaluation of packet loss at different error rates based on gaussian white noise signal

Mean time domain contrast under different situation can be seen in left Figure 6. It is seen that increased packet loss results in decreased mean contrast. The degradation in normalized mean square error between the reference target audio and the resulting audio in different situations can be seen in right Figure 6. It is clear that especially woofer 7 and to some degree woofer 8 are important for reproducing the desired signal in the bright zone. If other woofers are affected the error is minimal.

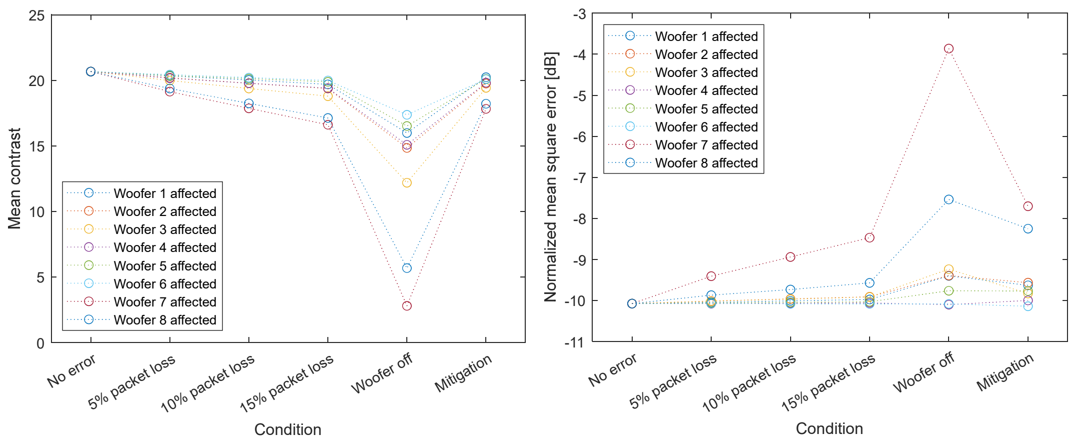

Figure 6: Left shows mean time domain contrast (Equation (1)) and right shows time domain normalized mean square error (Equation (2)) for 10 s gaussian white noise depending on error and affected woofer.

3.4 Evaluation of mitigation based on omitting affected woofer

Sound zone control filters based on the seven remaining woofers (as described in section 2.5) are calculated for each combination and their frequency response in the bright zone the contrast performance compared to the eight woofers situation can be seen in Figure 7 and the mean time domain contrast and normalized mean square error is also shown in Figure 6. As expected, the overall contrast is a little lower, but it depends very much on which woofer is off. Woofer 7 and 8 are the most important woofers in the setup due to their close proximity to the sound zones, and missing one of these affects the performance the most. As the reference target for the bright zone is the response of woofer 7 it is evident that the deviation in the frequency response in the target zone is also highest if woofer 7 is missing.

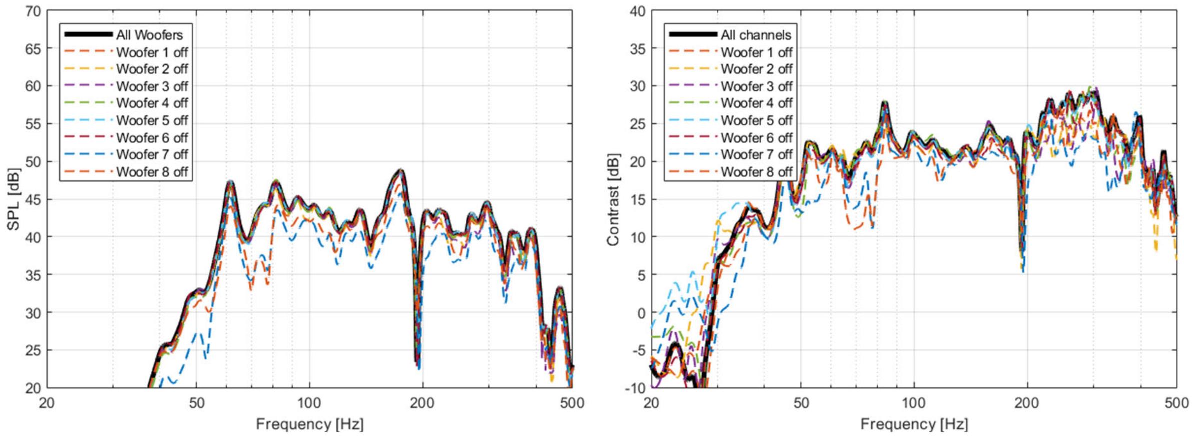

Figure 7: Left is the frequency response in bright zone and right is the frequency domain contrast (Equation (1)) for the mitigation by removing the affected woofer and using new filters.

3.5 Informal listening to simulated sound in bright and dark zone

From informal listening to the simulated sound in both the bright and the dark zone it is clear that the effect of having a woofer off gives increased leakage to the dark zone as the level is much higher. Packet loss gives audible transients, so it is expected to be more "distracting" than what can be expected by just looking at the mean contrast (Figure 6). Here the mitigation solution provides an audio experience similar to the "no error" scenario, but with a slight increased leakage in the dark zone.

4 Discussion

The transmission error manifest itself especially as noticeable leakage to the dark zone as destructive interference is very sensitive to deviations from the "perfect" signal. The reduction in contrast due to transmissions errors is frequency dependent due to the fact that the different woofers are contributing differently at different frequency ranges, i.e. one woofer may be more important in a specific frequency range. Therefore, the contrast performance of the sound zones is depends very much on which woofer is affected as could be expected from the investigated layout of the woofer placement. The woofers closer to the sound zones are more important for the sound zone performance. This opens the question if another reference target could give better performance. Possibly a reference target with a smoother frequency characteristic would give a more "transparent" sound reproduction inside the target zone.

The mitigation solution requires that the transmission error is detected, but the mitigation is effective in maintaining acceptable performance of the sound zones. For a specific setup the performance of the mitigation depends on how many woofers that are running error-free as the performance gets lower with the lower number of woofers available for the sound zone generation. In order to take any situation into account it is necessary to calculate sound zone control filters for all possible permutations of working woofers i.e., there could be any number and combination of woofers with transmission errors.

The setup investigated here consists of eight woofers, but for setups with less woofers, e.g. four woofers it is expected that the performance degradation due to transmission errors will be more severe as there are fewer woofers contributing appropriately to the sound zone generation. As a consequence, it is expected that a mitigation solution with only three woofers will show a larger performance degradation as compared to this study. Accordingly, setups with more than eight woofers will probably be less affected than what is seen here and the mitigation is expected to have less performance penalty.

As the transmission errors mainly causes increased leakage to the dark zone, the main consequence is that person(s) in the sound zones will be more distracted from the audio material from the other zone [22]. How distracting the error is in an actual setup with two sound zones with independent audio material depends very much on the audio material as this may provide masking if it is in the same frequency range [22]. But it is also seen that the distraction model proposed by [22] fails to predict the distraction when the music characteristics cannot be well described by the energy-based features used in the model. Informal listening revealed that especially random packet loss errors are audible due to transients. This opens up for future research in the audibility of transmission errors and mitigation solutions.

5 Conclusion

For feedforward sound zone systems any violation of the assumed linear time invariance will affect the performance. Missing the sound from a woofer or getting errors in the sound from one woofer gives significant increases in leakage to the dark zone seen as decreased contrast ranging from 3.3 to 17.9 dB depending on the affected woofer . This increased leakage can effectively be mitigated by removing this channel from the sound zone control system and by using another set of control filters where this woofer is excluded from the filter calculation. The mitigation only decreases the contrast 0.5 to 2.9 dB depending on the woofer. However, this requires some kind of error checking system in order to identify which woofer is affected and by which type of errors and a large set of precalculated filters in order to effectively deal with the possible combinations of errors. The results shown in this study are for a specific setup, but similar effects are expected to be general also for other sound zone setups based on distributed woofers.

Acknowledgements

This work is partly funded by the Innovation Fund Denmark (IFD) under File No 9069-00038B in the project: Interactive Sound Zones for Better Living (ISOBEL) [16].

References

[1] T. Betlehem, W. Zhang, M. A. Poletti and T. D. Abhayapala, "Personal Sound Zones: Delivering interface-free audio to multiple listeners," *IEEE SIGNAL PROCESSING MAGAZINE,* pp. 81-91, March 2015.

[2] W. F. Druyvesteyn and J. Garas, "Personal Sound," *J. Audio Eng. Soc,* vol. 45, no. 9, pp. 685-701, 1997.

[3] P. Coleman, P. J. B. Jackson, M. Olik, M. Møller, M. Olsen and M. A. Pedersen, "Acoustic contrast, planarity and robustness of sound zone methods using a circular loudspeaker array," *J. Acoust. Soc. Am.,* vol. 135, no. 4, p. 1929–1940, April 2014.

[4] X. Ma, P. J. Hegarty, J. A. Pedersen and J. J. Larsen, "Impact of loudspeaker nonlinear distortion on personal sound zones," *J. Acoust. Soc. Am.,* vol. 143, no. 1, p. 51–59, January 2018.

[5] M. Olsen and M. B. Møller, "Sound zones: on the effect of ambient temperature variations in feed-forward systems," in *AES 142th Convention*, Berlin, Germany, 2017.

[6] S. i. Hahm, P. Kang, H. Bang and H. j. Yeon, "Dynamic media buffer control scheme for seamless streaming in wireless local area networks," in *IEEE Wireless Communications and Networking Conference Workshops*, 2016.

[7] Rappaport, T. S., Wireless communications: principles and practice, vol. 2, New Jersey: Prentice Hall PTR, 1996.

[8] T. Kim and M. H. Ammar, "Receiver buffer requirement for video streaming over TCP," in *Electronic Imaging 2006*, San Jose, California, United States, 2006.

[9] K. J. Ma, M. L. A. Huang and R. Bartos, "Video rate adaptation in mobile devices via http progressive download of stitched media files," *Communications Letters, IEEE,* vol. 15, no. 3, p. 320–322, 2011.

[10] T. Kim, N. Avadhanam and S. Subramanian, "Dimensioning receiver buffer requirement for unidirectional vbr video streaming over TCP," in *Image Processing, 2006 IEEE International Conference on*, 2006.

[11] J. Hansen, J. Østergaard, J. Kudahl and J. H. Madsen, "Sequential Use of Block Codes and Convolutional Codes in a Real-Time Multi-Hop Network," in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, 2018.

[12] J. H. Sørensen, P. Popovski and J. Østergaard, "Delay Minimization in Real-Time Communications With Joint Buffering and Coding," *IEEE Communications Letters,* vol. 21, no. 1, pp. 52-55, Jan 2017.

[13] C. Perkins, O. Hodson and V. Hardman, "A survey of packet loss recovery techniques for streaming audio," *IEEE Netw,* vol. 12, no. 5, pp. 40-48, 1998.

[14] X. Lu, H. He and H. Tan, "A low complexity packet loss recovery method for audio transmission," in *Proc. 2nd Int. Conf. on Computer Science and Electronics Engineering (ICCSEE 2013)*, 2013.

[15] A. Botta and A. Pescapé, "IP packet interleaving for UDP Bursty losses," *J. Syst. Softw.,* vol. 109, p. 177–191, 2015.

[16] ISOBEL, "Interactive Sound Zones for Better Living," 2022. [Online]. Available: https://isobel.dk/.

[17] M. F. S. Gálvez, S. J. Elliott and J. Cheer, "Time Domain Optimization of Filters Used in a Loudspeaker Array for Personal Audio," *IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING,* vol. 23, no. 11, pp. 1869-1878, 2015.

[18] M. B. Møller and M. Olsen, "Sound zones: on envelope shaping of fir filters," in *ISCV24*, London, 2017.

[19] M. B. Møller and J. Østergaard, "A Moving Horizon Framework for Sound Zones," *IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING,* vol. 28, pp. 256-265, 2020.

[20] S. Müller and P. Massarani, "Transfer-Function Measurement with Sweeps," *Journal of Audio Engineering Society,* vol. 49, no. 6, pp. 443-471, 2001.

[21] M. B. Møller and M. Olsen, "Sound Zones: On Performance Prediction of Contrast Control Methods," in *Proc. Audio Engineering Society Conference: 2016 AES International Conference on Sound Field Control*, Guildford, United Kingdom, 2016.

[22] J. Francombe, R. Mason, M. Dewhirst and S. Bech, "A Model of Distraction in an Audio-on-Audio Interference Situation with Music Program Material," *J. Audio Eng. Soc.,* vol. 63, no. 1/2, pp. 63-77, January/February 2015.