DOI: 10.61782/fa.2023.0722

forum acusticum 2023

SOUND QUALITY EVALUATION OF PACKET LOSS CONCEALMENT FOR WIRELESS LOW-FREQUENCY SOUND ZONES

Christian Sejer Pedersen[1*] **Mo Zhou**[1] **Martin Bo Møller**[2]
Niels Evert Marius de Koeijer[2] **Jan Østergaard**[1]
[1] Department of Electronic Systems, Aalborg University, Denmark
[2] Bang & Olufsen A/S, Struer, Denmark

ABSTRACT

Personal sound zones based on sound field control are sensitive to errors in the control signals and both the sound quality and the leakage to other zones are affected. When transmitting the signals over wireless networks, packet losses can occur. For a specific sound zone setup with distributed woofers, a packet loss concealment method based on auto-regressive models has been shown to predict lost low-frequency woofer packets to such a degree, that the objective sound quality model PEAQ ODG estimates that the effect is inaudible at low packet loss rates. The purpose of this study is to evaluate packet loss concealment strategies more thoroughly for different wireless error conditions using several additional objective sound quality models (instantaneous PEAQ ODG, 2f-model and ViSQOLAudio) and assess how the music signal affects the performance. The objective evaluations show that both the severity of packet loss and effect of packet loss concealment greatly depend on the type of music and the sensitivity of the objective sound quality models varies considerably. Experiments are in preparation to verify the findings subjectively.

Keywords: *personal sound zones, packet loss concealment, objective sound quality models*

*[*]**Corresponding author**: cp@es.aau.dk*

1. INTRODUCTION

Sound field control based on several loudspeaker signals is sensitive to errors in loudspeaker signals as successful control is based on the superposition of contributions from all the loudspeakers. Personal sound zones with wireless transmission of woofer signals is an application where the transmission is not guaranteed to be lossless. It has been shown that a sound zone system is sensitive to both synchronization errors and packet losses which can greatly limit the contrast [1] between sound zones [2]. An objective sound quality model like PEAQ ODG [3] also predicts a significant reduction in sound quality with packet loss, which is mainly due to the transient behavior of the artifacts when the playback transitions between correct and lost packets. Introducing packet loss concealment (PLC) based on autoregressive (AR) models shows promising results both when it comes to contrast between the zones and in initial evaluations of objective sound quality from PEAQ ODG [4].

Many objective sound quality models exist and from the review presented in [6] several other methods seem to perform better than PEAQ ODG in several different contexts and especially the 2f-model in general seems to be state-of-the-art and outperform the others.

The aim of this study is to expand on the initial evaluations presented in [4] by more thorough evaluation of the sound quality performance of PLC methods for different conditions of packet loss in a low frequency sound zone system based on distributed woofers. This is done using several additional objective sound quality models like the instantaneous PEAQ ODG as proposed by [5], the 2f-model [6] [7] and ViSQOLAudio [8] and in addition the influence of music characteristics on the performance is assessed. It should be mentioned that no objective sound quality model has been verified for the situation of low-frequency sound zone setups and the present study is a first stage of this verification, while subjective evaluations are in preparation, both for verification and for assessing which objective sound quality model best predicts the subjective sound quality. This is useful for further advancements in e.g. evaluating robust filters for generating sound zones [9] and refinement to PLC methods.

2. PERSONAL SOUND ZONE SYSTEM

2.1 Physical setup and sound zone generation

A two-zone personal sound zone setup with eight distributed 10''-woofers is implemented in a well damped room (T20 ranging from 0.6 s at 50 Hz to 0.2 s at 400 Hz) as shown in Figure 1. The two sound zones are fixed in position at the left and right side of a three-person sofa, resulting in approx. 1 meter distance between centers of the sound zones. The low-frequency system is part of a full-range sound zone system, but the soundbar for generating the higher frequencies > 200 Hz is not active in the present study as it is intended to evaluate the low-frequencies under worst case conditions and the soundbar audio can potentially provide masking of the errors.

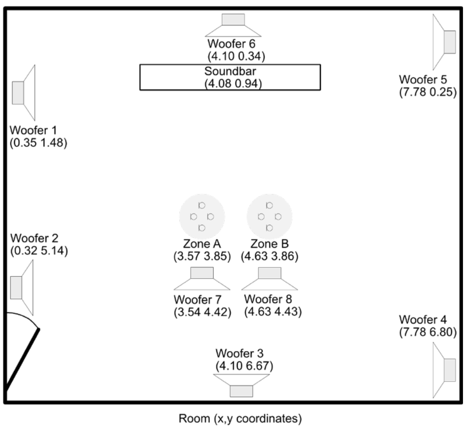

Figure 1: Diagram of sound zone setup with distributed woofers.

The sound field control for the low-frequency part (<200 Hz) of the sound zone system is based on pressure matching in the time domain using the method from [10] where the tradeoff between maximizing acoustic contrast while minimizing the mean square error between a reference target and the actual sound is controlled. In addition, shaping of the envelope of the control filters is introduced in order to reduce pre- and post-ringing [11] as these effects can reduce sound quality. The filter design is a feedforward paradigm based on measured transfer functions from each woofer to 20 microphone positions (two heights of 10 microphone positions covering approx. 20 x 30 x 10 cm in an interleaved pattern) in each sound zone and uses regularization in order to control the effort of each woofer. Each set of filters creates a bright zone, where audio is wanted and a dark zone where audio is unwanted. By superposition two separate sound zones are created. The filters are 100 tap FIR filters calculated at a sampling frequency of 1200 Hz, and the signals to the woofers are downsampled to this frequency. An example of the audio separation between the zones can be seen in Figure 2.

From a practical and aesthetical point of view, the woofers are intended to be wireless, but in order to have full control of the experimental conditions they are wired and have perfect synchronization as this is very important for the performance of the sound field control [2]. Packet loss conditions are simulated in MATLAB [12].

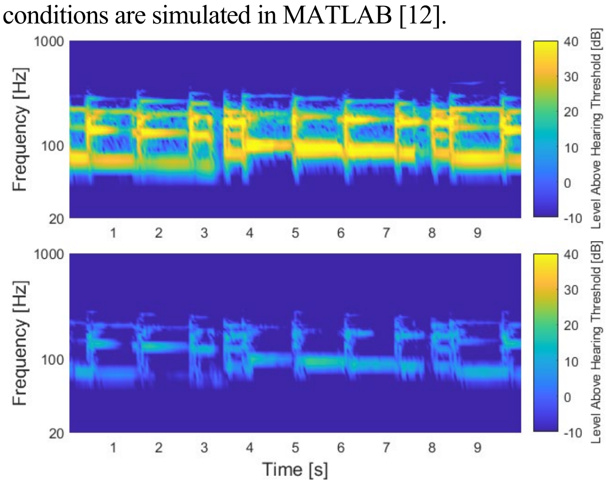

Figure 2: Threshold weighted spectrogram of the audio in the two sound zones when playing *Led Zeppelin – Dazed and Confused* in zone B without packet loss. Top is zone B, bottom is zone A. The contrast from zone B to the zone A is 18.3 dB.

2.2 Wireless packet loss simulation

Wireless packet loss errors are simulated by splitting the signals for the woofers into packets (packet size of 24 samples corresponding to 20 ms at 1200 Hz) and dropping packets according to the error type. Two types of errors, independent and identical distributed (i.i.d.) packet loss and bursty packet loss patterns derived from a two-level Markov model [13] are simulated and evaluated separately. I.i.d. packet loss is simulated for a single woofer

(woofer 7) for a wide range of packet loss rates and a worst-case scenario of bursty packet loss to all woofers (with individual bursty packet loss patterns) is also simulated for a few packet loss rates from 1 to 5%. Figure 3 dotted lines shows the effect of packet loss on the contrast between the zones.

2.3 Packet loss concealment

Two different types of low latency packet loss concealment (PLC) are being evaluated, transient mitigation and autoregressive (AR) model PLC [4].

2.3.1 Transient mitigation PLC

The transient mitigation PLC is reducing the audibility of the transient artifacts of packet loss by fading in/out to/from lost packets using half Hanning windows. Optimal window sizes were found to be fade out of 8 samples and fade in of 6 samples [4] as a compromise between reducing transients while not decreasing the contrast too much (see Figure 3).

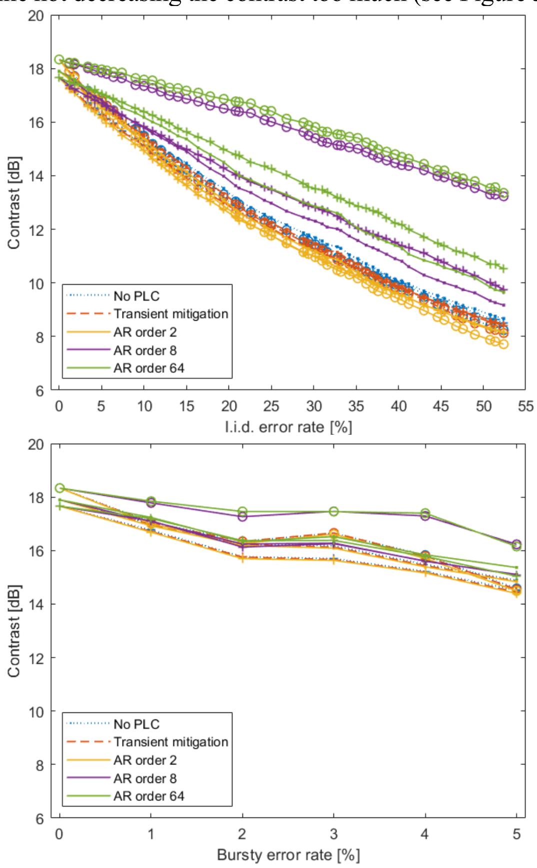

Figure 3: Contrast from bright to dark zone for different packet loss and PLC conditions for the tracks: o = *Dazed*, + = *Classical*, · = *Rock* (see section 3.5).

2.3.2 Autoregressive model PLC

The AR model PLC is attempting to reconstruct the lost packets by extrapolation with a linear prediction filter using an AR model [14], [4]. Based on previous samples it predicts the lost packet and additional samples for merging with the next correct packet. In the initial work on the AR model [4] a merge overlap size of 8 samples was found to be the best compromise between mitigating transients without decreasing contrast too much. The two main parameters of the AR model are order and training size, N, i.e. how many previous samples to base the AR model on. It was found in [4] that an order of 64, and N size between 360-400 samples seems optimal regarding contrast and objective sound quality. In the results presented here, N=384 samples corresponding to 320 ms is used, but the order is varied to evaluate its effect. As shown in Figure 3 the AR model PLC is predicting the lost packets to such a degree that contrast is considerably improved compared to no PLC.

3. OBJECTIVE SOUND QUALITY MODELS

Using transfer function measurements of the sound zone system to two omnidirectional microphone positions with 20 cm distance corresponding approx. to a left and right ear of a listener it is possible to evaluate the audio in the zones under different simulated conditions of packet loss and PLC. Numerous objective sound quality models exist and an overview and evaluation can be found in [6]. Generally, related to packet loss and PLC of music signals PEAQ ODG [3] has previously been used in the literature e.g. [14], and other potential candidates are investigated. All the models require a reference, which is the audio in the zone without packet loss. As required by all the used models, the audio is upsampled to 48 kHz. The reduction of the artifacts due to the inherent lowpass filtering in the resampling is justified because we intend to transmit the woofer signals at 1200 Hz sampling frequency.

It should be mentioned that no objective sound quality model has ever been verified for the context of low-frequency sound zones or for packet loss and PLC in this context. Therefore, this will be part of the second stage of the current study and is in preparation at the time of writing.

3.1 Perceived Audio Quality Objective Difference Grade (PEAQ ODG)

PEAQ ODG [3] exists in a basic and advanced version, where the basic version is based on an FFT-based model of the human hearing, while the advanced is based on a filter bank. The code used here is based on the MATLAB implementation of the PEAQ basic version from [15]. In general, PEAQ splits the audio into frames, and frame by frame compare an internal hearing model representation of the degraded audio frame with that of the reference frame. It then calculates several different output variables and an overall Objective Difference Grade (ODG) ranging from 0 "imperceptible" to -4 "very annoying". In the case of binaural signals, as used here, most model output variables are calculated separately for left and right channel before averaging, and only the combined output variables are reported.

3.2 Instantaneous PEAQ ODG

According to [5] the mean of instantaneous PEAQ ODG and especially the mean of the lowest 18% of the instantaneous PEAQ ODG is a better predictor of sound quality in cases of packet loss as compared to the overall PEAQ ODG. The rationale for this is that it is the severity of the audio degradations that dictate the overall sound quality.

The code for calculating the instantaneous PEAQ ODG was further modified from [15] in order to extract the PEAQ ODG from each frame i.e. the instantaneous PEAQ ODG. The mean of the lowest 18% will be abbreviated $\mathrm{PEAQ_{inst}}$ $\mathrm{ODG_{m18\%}}$. An example of the instantaneous PEAQ ODG for a case with 5% i.i.d. packet loss (audio is *Led Zeppelin – Dazed and Confused*) can be seen in Figure 4 for no PLC and Figure 5 for an AR model PLC order 64.

3.3 2f-model

According to the review of sound quality models [6] the 2f-model is state-of-the-art, as it is successfully able to predict the subjective results in several different contexts. It is based on two model output variables, Averaged Modulation Difference (AvgModDiff1) and Average Distorted Blocks (ADB) from PEAQ. For the PEAQ implementation from [15] the model equation is given in Eqn. (1) [7]:

$$\mathrm{MMS_{est}} = \frac{56.1345}{1+(-0.0282\cdot\mathrm{AvgModDiff1}-0.8628)^2} - 27.1451\cdot\mathrm{ADB}+86.3515, \quad (1)$$

where $\mathrm{MMS_{est}}$ represents a Mean MUSHRA Score and must be limited to the range of 0-100.

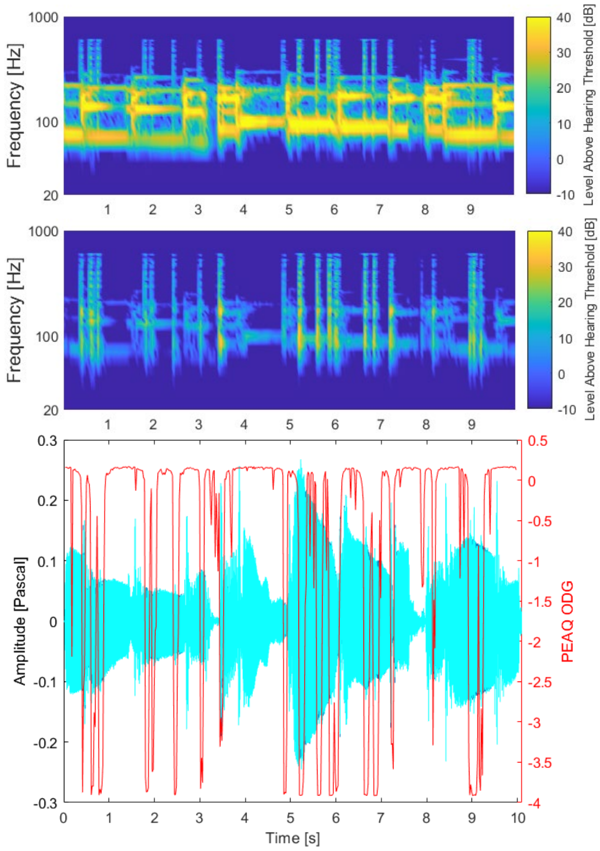

Figure 4: Top: spectrogram bright zone, middle: spectrogram dark zone, bottom: audio in bright zone zone (dark blue: reference audio, cyan: actual audio, red: instantaneous PEAQ ODG) with 5% i.i.d. packet loss woofer 7 with no PLC, PEAQ ODG:-3.56, $\mathrm{PEAQ_{inst}}$ $\mathrm{ODG_{m18\%}}$:-3.65, 2f-model:59.6, ViSQOLAudio:0.97, contrast:16.8 dB.

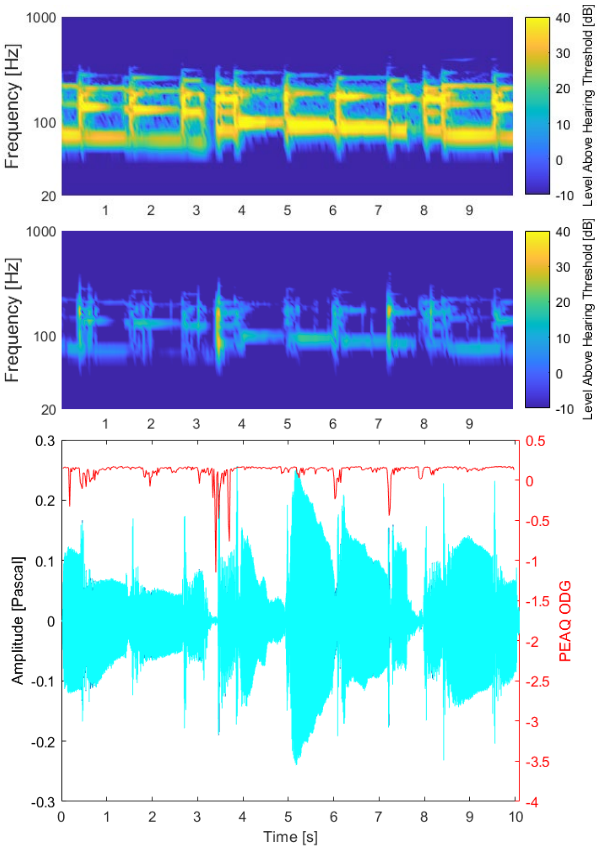

Figure 5: Top: spectrogram bright zone, middle: spectrogram dark zone: bottom: audio in bright zone (dark blue: reference audio, cyan: actual audio, red: instantaneous PEAQ ODG) with 5% i.i.d. packet loss woofer 7 using AR model PLC (order 64), PEAQ ODG:0.06, PEAQ$_{inst}$ ODG$_{m18\%}$:-0.02, 2f-model:98.7, ViSQOLAudio:0.99, contrast:18.0 dB.

3.4 ViSQOLAudio

ViSQOLAudio [8] is based on differences in internal hearing representation using spectrograms. This difference is expressed on a similarity scale from 0 to 1, and for the available MATLAB implementation from [8] there is no transformation to a sound quality metric. This transformation has been implemented in a later version 3 of ViSQOLAudio [16], but according to the review by [6] it seems in general to perform slightly worse than the original version [8] and therefore it was not considered for the present study. The MATLAB implementation expects only one channel, so the left and right ear signal is mixed before calling the function.

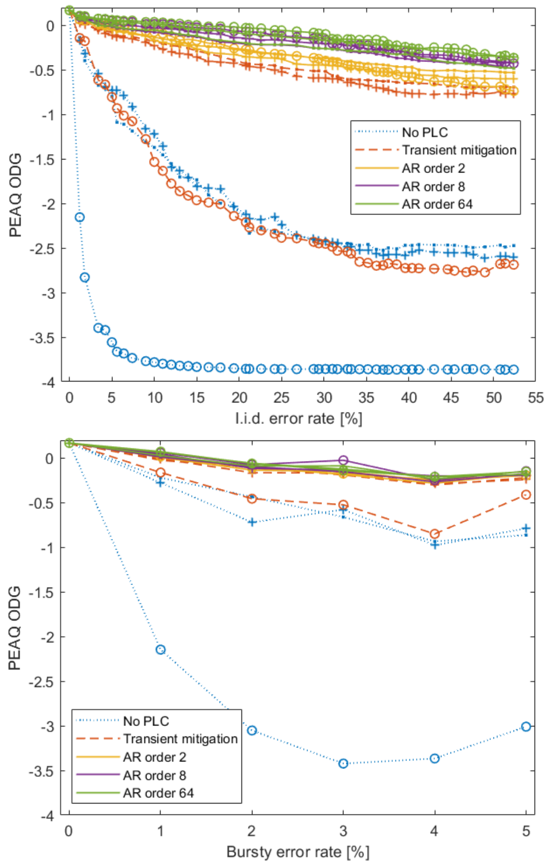

Figure 6: Sound quality predictions from PEAQ ODG for different packet loss and PLC conditions for the tracks: o = *Dazed*, + = *Classical*, · = *Rock*.

3.5 Comparison of objective sound quality model outputs

Three 10 s music tracks (starting 5 s into the tracks) are used to evaluate the different sound quality model outputs for different conditions of packet loss and PLC: *Led Zeppelin - Dazed and Confused* (*Dazed*), *The Killers – On Top* (*Rock*), *Brahms - 21 Hungarian Dance No. 18 in D Major*, (*Classical*). The outputs from the objective sound quality models can be seen in Figure 6 to Figure 9 where top figures shows conditions with i.i.d. packet loss to woofer 7 and bottom figures show bursty packet loss to all eight woofers.

The music tracks have been selected based on their different characteristics: i.e. *Dazed* is bass heavy with slow changing tones and not much energy in the frequency range of the transient artifacts, *Rock* is less bass heavy and contain more transients while *Classical* have more fast changing tones.

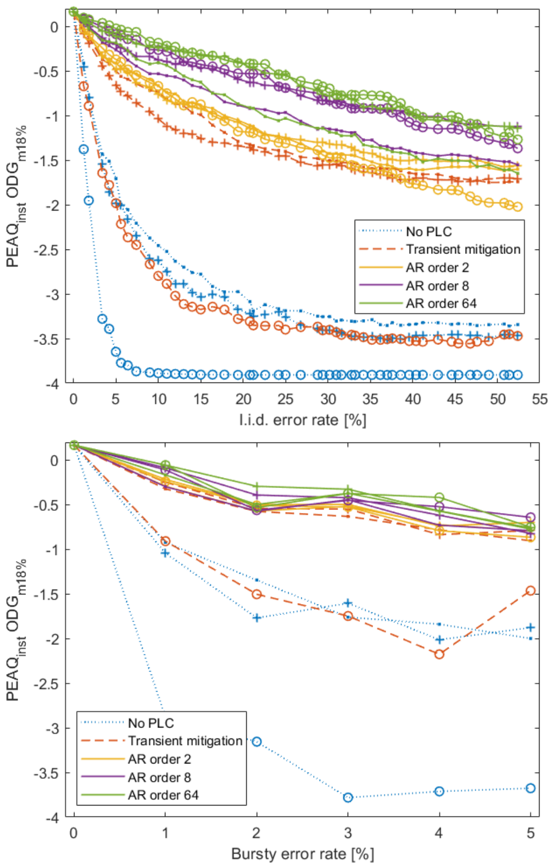

Figure 7: Objective sound quality from PEAQ_{inst} $\mathrm{ODG}_{m18\%}$ for different packet loss and PLC conditions for the tracks: o = *Dazed*, + = *Classical*, · = *Rock*.

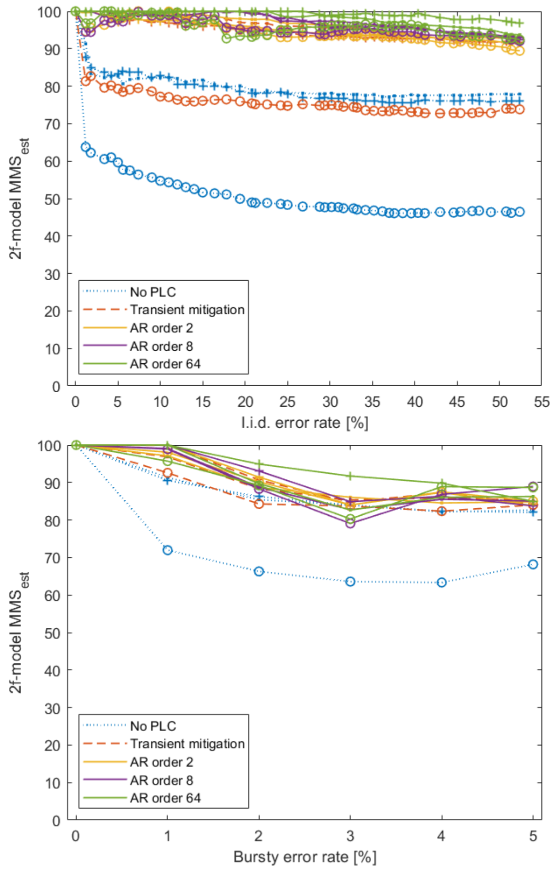

Figure 8: Objective sound quality from 2f-model for different packet loss and PLC conditions for the tracks: o = *Dazed*, + = *Classical*, · = *Rock*.

4. DISCUSSION

4.1 Dependency on audio material

As seen in Figure 6 to Figure 9 both the sound quality degradation due to packet loss, and the effectiveness of the PLC depends significantly on the audio material. Among the three tested audio tracks, *Dazed* shows both the lowest quality for packet loss, but also the highest improvement by the AR model PLC, while *Rock* is not affected as much by packet loss, and the AR model PLC is not able to improve the sound quality to the same degree.

The relatively large impact of packet loss on *Dazed* can be explained by the higher levels of the transients and less spectral content in the frequency range of the transient artifacts. This makes the transients "stand out" more and they are therefore more severe as compared to the other music signals (compare Figure 4 with Figure 10). On the other hand, the slow changing nature of the music signal of *Dazed* is a better condition for the AR model prediction of the lost packets as compared to e.g. *Rock*, with its fast and "unpredictable" variations.

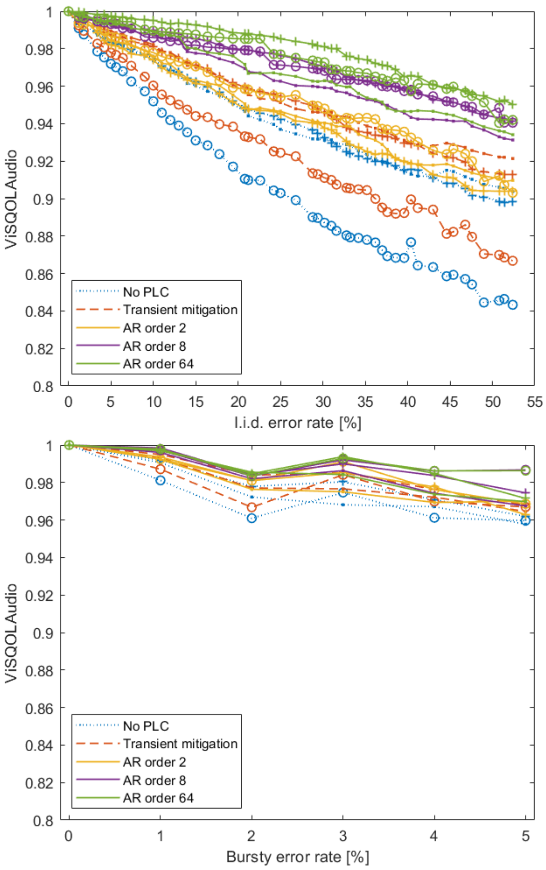

Figure 9: Objective sound quality from ViSQOLAudio for different packet loss and PLC conditions for the tracks: o = *Dazed*, + = *Classical*, · = *Rock*.

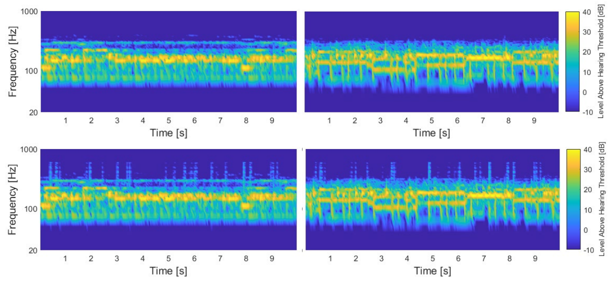

Figure 10: Spectrograms of *Classical*, left column, and *Rock*, right column. Top is no error and bottom is 5% i.i.d. packet loss.

4.2 Comparison of objective sound quality models

It is seen in Figure 6 to Figure 9 that the different objective sound quality models generally show great improvement with the AR model PLC especially for i.i.d. packet loss, which is also supported by informal listening. But they differ considerably in their predictions of the packet loss influence on sound quality. This is evident by comparing the condition of no PLC where ViSQOLAudio only goes below 0.9 for few cases of very high i.i.d. packet loss (Figure 9) while the PEAQ ODG and especially $\text{PEAQ}_{\text{inst}}$ $\text{ODG}_{\text{m18\%}}$ quickly goes towards -4 (very annoying) for the same conditions (Figure 6 and Figure 7).

In general ViSQOLAudio does not seem sensitive to the artifacts in this context, while $\text{PEAQ}_{\text{inst}}$ $\text{ODG}_{\text{m18\%}}$ is the most sensitive and the other two models are somewhere inbetween. Therefore, the results of future subjective tests are very important in order to both verify the improvement gained by the AR model PLC and to find the best predictor to be used for further advancement with robust filters [9] and in transmission schemes, sending redundant information, like Multiple descriptions [17] and finally refinement of PLC methods if everything else fails.

5. CONCLUSION

Objective sound quality models predict that packet loss in low-frequency sound zones can have a significant degrading effect on the sound quality, and AR model PLC is capable of reducing this degradation considerably. But different sound quality models are shown to be more or less sensitive to the artifacts. Additionally, the degradation and effectiveness of the AR model PLC depend significantly on the music signal. Informal listening generally supports the objective effects of packet loss and AR model PLC, and a formal listening experiment is in preparation in order to verify these findings. It remains to be seen what the subjective sound quality impact of packet loss and PLC is, and which objective sound quality models best correlate with the results.

6. ACKNOWLEDGMENTS

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) [18].

7. REFERENCES

[1] 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.

[2] C. S. Pedersen, M. B. Møller and J. Østergaard, "Effect of Wireless Transmission Errors on Sound Zone Performance at Low Frequencies," in *EUROREGIO BNAM2022 Joint Acoustic Conference*, Aalborg, Denmark, 2022.

[3] Rec. ITU-R BS.1387-1, "Method for objective measurement of perceived audio quality," ITU, 2001.

[4] C. S. Pedersen, M. Zhou, M. B. Møller, N. E. M. d. Koeijer and J. Østergaard, "AR model for low latency packet loss concealment for wireless sound zones at low frequencies," in *AES EUROPE - 154 Convention*, Helsinki, Finland, 2023.

[5] J. E. Voldhaug, E. Hellerud and U. P. Svensson, "Evaluation of Packet Loss Distortion in Audio Signals," in *Audio Engineering Society 120th Convention Paper 6855*, Paris, France, 2006.

[6] M. Torcoli, T. Kastner and J. Herre, "Objective Measures of Perceptual Audio Quality Reviewed: An Evaluation of Their Application Domain Dependence," *IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING,* vol. 29, pp. 1530-1541, 2021.

[7] T. Kastner and J. Herre, "Subjective evaluation of blind audio source separation database: SEBASS-DB," Oct 2019. [Online]. Available: https://www.audiolabs-erlangen.de/resources/2019-WASPAA-SEBASS/.

[8] A. Hines, E. Gillen, D. Kelly, J. Skoglund, A. Kokaram and N. Harte, "ViSQOLAudio: An Objective Audio Quality Metric for Low Bitrate Codecs," *The Journal of the Acoustical Society of America,* vol. 137, no. 6, pp. EL449-EL455, 2015.

[9] M. Zhou, M. Møller, C. S. Pedersen and J. Østergaard, "ROBUST FIR FILTERS FOR WIRELESS LOW-FREQUENCY SOUND ZONES," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Rhodes Island, Greece, 2023.

[10] 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.

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

[12] MATLAB, *version 9.14.0.2239454 (R2023a) Update 1,* Natick, Massachusetts: The Mathworks, Inc., 2023.

[13] M. Ellis, D. P. Pezaros, T. Kypraios and C. Perkins, "A two-level Markov model for packet loss in UDP/IP-based real-time video applications targeting residential users," *Computer Networks,* vol. 70, p. 384–399, 2014.

[14] M. Fink and M. Z. U. Holters, "COMPARISON OF VARIOUS PREDICTORS FOR AUDIO EXTRAPOLATION," in *Proc. of the 16th Int. Conference on Digital Audio Effects (DAFx-13)*, Maynooth, Ireland, 2013.

[15] P. Kabal, "PQEvalAudio 1.0 Matlab toolbox," Dept. Electrical & Computer Engineering, McGill University, 2004. [Online]. Available: http://www-mmsp.ece.mcgill.ca/Documents/Software/index.html.

[16] M. Chinen, F. S. C. Lim, J. Skoglund, N. Gureev, F. O'Gorman and A. Hines, "ViSQOL v3: An Open Source Production Ready Objective Speech and Audio Metric," in *Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, 2020.

[17] J. Østergaard, C. S. Pedersen, M. Zhou and M. Møller, "Multiple Description Audio Coding for Wireless Low-Frequency Sound Zones," in *(DCC 2023) 2023 Data Compression Conference*, Snowbird, UT, USA, 2023.

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

[19] 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.

[20] T. Gueham and F. Merazka, "Packet loss concealment method based on interpolation in packet voice coding," *Computer Standards & Interfaces,* vol. 85, 2023.

[21] Rec. ITU-R BS.1383-2, "Guidance for the selection of the most appropriate ITU-R Recommendation(s) for subjective assessment of sound quality," ITU, 2019.

[22] Rec. ITU-R BS.1534-3, "Method for the subjective assessment of intermediate quality level of audio systems," ITU, 2015.