Audio Engineering Society

Conference Paper 7

Presented at the International Conference on Acoustics & Sound Reinforcement
2024 January 22–26, Le Mans, France

On the factors influencing groove fidelity in immersive live music events

Thomas Mouterde[1], Nicolas Epain[1], Samuel Moulin[1], and Etienne Corteel[1]

[1]*L-Acoustics, 13 rue Levacher Cintrat, 91460 Marcoussis, France*

Correspondence should be addressed to Thomas Mouterde (thomas.mouterde@l-acoustics.com)

ABSTRACT

Spatial audio is employed more and more often in large-scale live music events. In events of this kind, loudspeakers can be widely spaced apart, which may result in large time differences of arrival between certain sources. These timing differences may in turn affect the perceived rhythmic quality of music, or groove, as the synchronization between instruments is modified. This paper presents the results of a perceptual experiment that investigated how different factors, such as the nature of the instrument or the musical genre, impact the perceived groove modification resulting from sound propagation time differences. The results indicate that different instruments can show more or less sensitivity to time shifts, even in the same musical excerpt. Based on these findings, we derive mixing and sound system design guidelines that aim at preserving an optimal musical quality for the majority of the audience.

1 Introduction

Spatial audio is increasingly used in large-scale live music events. Sound systems employed for such *immersive sound* events typically rely on five or more loudspeakers spanning the entire width of the performance area [1]. Compared to traditional stereo systems, they offer better spatialization accuracy and audio-visual consistency. Moreover, using spatial extensions on the sides and rear of the audience provides additional creative freedom when mixing.

However, in live sound, the dimensions of the audience and performance areas can be so large that the time it takes for sounds emitted by two different loudspeakers to propagate to a given point in the audience may be significantly different. In previous work [2], the authors demonstrated that this propagation time difference could reach 10 to 15 ms at the mixing position, and even 20 to 40 ms on the side of the audience, for a frontal system spanning a 20-meter wide stage.

When audio objects are panned in different directions using sound systems of this kind, propagation time differences result in time shifts between instruments. This may modify the perceived rhythmic characteristics, or *groove*, of musical pieces, as instrument synchronization is critical in this regard [3, 4, 5]. Hence, panning audio objects over a large-scale immersive sound system may alter the perception of musical groove at certain positions in the audience.

In [2], the authors presented a first study that aimed to assess the impact of propagation time differences occurring in immersive sound live events on groove perception. The paper presented the results of a perceptual experiment, which showed how the perceived groove quality decreases when inter-instrument time shifts increase. The results also indicated that the perceived quality degradation did not depend on sound spatialization.

Building upon this previous work, the present paper aims to explore the impact of the following factors on the perceived groove quality:

- nature of the instrument, for a given excerpt,
- musical complexity,
- musical genre, and
- time shift direction (early or late).

Regarding the latter, a sound that arrives early on one side of the audience arrives late on the other side, thus the impact on groove perception can be different.

The remainder of the paper is organized as follows. Section 2 describes the design of the perceptual experiment. Section 3 presents the test results, showing how the different factors above affect the perceived groove quality. Lastly, in Section 4, results are discussed and mixing and system design guidelines are derived.

2 Experimental design

In this section, we describe the perceptual experiment.

2.1 Conditions and stimuli

The test stimuli were generated using the same three musical excerpts as in [2]. These excerpts were selected as examples of three different musical genres, and were extracted from pieces of music for which we have multi-track recordings:

- Excerpt 1, a 16 s excerpt from "Dance with you", by La Reserve, referred to as *Funky* in the following (tempo: 124 BPM);
- Excerpt 2, a 23 s excerpt from "Coming home to you", by La Reserve, referred to as *Ballad* in the following (tempo: 86 BPM);
- Excerpt 3, a 15 s excerpt from "Terrain" by Halina Rice, referred to as *EDM* in the following (tempo: 125 BPM).

Each audio stimulus was synthesized by time-shifting one harmonic instrument (bass, guitar, or keyboard) ahead (early degradation), or behind (late degradation) with respect to the rhythmic instrument(s). Six audio tracks, *i.e.* combinations of instruments, were generated from the three excerpts († indicates the time-shifted instrument):

1. *Funky 1*: drums and bass†;
2. *Funky 2*: drums and guitar†;
3. *Funky 3*: drums, bass, keyboard, and guitar†;
4. *Ballad 1*: drums and arpeggio guitar†;
5. *Ballad 2*: drums and keyboard†;
6. *EDM*: rhythm patterns and synthetic bass†.

Note that, depending on the audio excerpt, the instruments may have not been played strictly on the beat. This is because the musicians played according to their own feelings of the musical groove. Hence, any time shift present between instrument onsets in the original recordings must be considered as the artistic intention of the musicians. However, the objective of the test is not to qualify the groove of the original music pieces but to estimate how this particular groove is affected by additional time shifts.

It may be noted that Tracks 2, 4, and 6 are the same as the ones used in [2]. However, in this previous study, the bass or guitar was only shifted ahead of rhythmic instruments. In the present study, stimuli were generated by shifting instruments both ahead of and behind rhythmic instruments. In addition, including Tracks 1 and 5 allows us to compare the impact of time shifts on two different instruments within the same musical excerpt: a guitar versus a bass comparing Tracks *Funky 1* and *Funky 2*; a guitar versus a keyboard comparing Tracks *Ballad 1* and *Ballad 2*. The purpose of Track 3 (*Funky 3*) is to investigate how musical complexity impacts the perceived degradation when time offsets are inserted. Indeed, Track 2 only consists of drums and a guitar while Track 3 also includes a bass and a keyboard.

Four levels of time offset were used: 15, 30, 45, and 60 ms. These durations correspond to 1/32, 2/32, 3/32, and 4/32 beats for excerpts 1 and 3 (tracks 1, 2, 3, and 6) with a tempo of around 125 bpm. However, unlike in [2], time offsets have been applied regardless of the musical excerpt tempo. Note that the relative timing of the elements that form the rhythmic part was not modified.

In summary, the test stimuli were generated to test the following independent variables: the music track (different musical genres, instrument combinations, and complexity levels), the direction of the degradation (early or late shift), and the time offset level. The test was divided into 12 successive trials, corresponding to every combination of track (6) and direction of degradation (2), with each trial comparing four different time

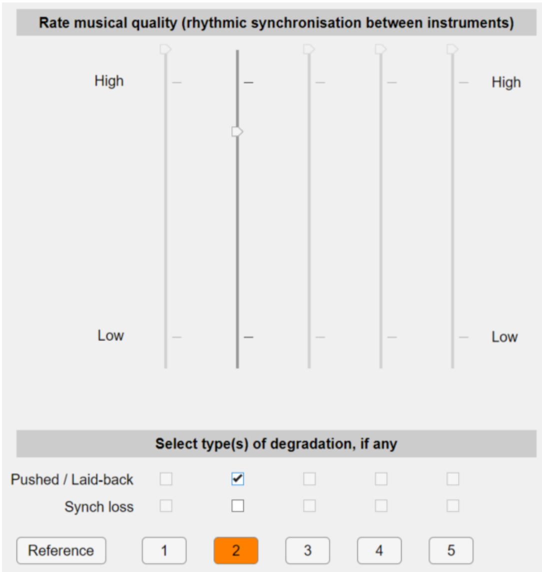

Fig. 1: The graphical user interface of the application used for the perceptual test.

offset values. The order in which trials were presented was randomized. Please note that all stimuli are available online [1].

2.2 Test methodology

The same method, inspired by the MUltiple Stimuli with Hidden Reference and Anchor (MUSHRA) methodology, and interface were used as in [2]. The test interface is shown in Fig. 1. Participants were presented with an explicit reference, which consisted of the track as initially played by the musicians. They were then instructed to compare five stimuli to this reference: the four degraded versions with different time-offset levels, and a hidden reference (HRef). The time offset values of 15, 30, and 45 ms correspond to those encountered with large-scale frontal spatial-audio sound systems, as demonstrated in [2]. Stimuli with a time offset of 60 ms were used as a low-quality anchor: the duration of these offsets was larger than that simulated for large-scale sound systems and was expected to result in obvious quality degradations when compared to the reference. The stimuli under test were randomly labeled with numbers ranging from 1 to 5 for every trial.

[1] https://l-acoustics.github.io/grooveperception.github.io/

Participants were asked to rate the rhythmic synchronization between instruments for each of the 5 conditions (HRef, 15, 30, 45, and 60 ms offset) compared to the explicit reference. The quality rating was done using a continuous scale ranging from low to high quality, with the highest quality meaning that there was no perceived degradation. For each trial, participants could freely listen to the different stimuli, and go back and forth between conditions as many times as needed.

In addition to the quality rating, participants were asked to classify any perceived degradation within the two following categories:

- *Pushed / Laid-back*: a modification of the groove, of the instrument leading;
- *Synch. loss*: a loss of synchronization between instruments.

Note that the participants were instructed not to tick any box if they perceived no degradation, but the presence of a hidden reference among the stimuli under test was not explicitly expressed.

2.3 Test procedure

The experiment was conducted individually using headphones (Sennheiser HD650) in a quiet meeting room. Participants interacted with a Matlab application running on a laptop equipped with an RME Digiface AVB audio interface. The sound level was set by the organizers to be loud enough to hear the details of the music, but remain comfortable in the event of a one-hour test session. In addition, listeners could slightly adjust the sound level to their taste.

The experiment started with the reading of the test instructions. A discussion with the tester followed to ensure instructions were properly understood. Next, the participants went through a two-step protocol to familiarize themselves with the test interface and task:

1. participants were presented with examples of timing degradations;
2. participants took a short pre-test, with the same interface as for the actual test.

The familiarization phase used an excerpt of "End of the Road" by La Reserve, which combines drums and a bass guitar. The timing degradation examples were generated by presenting the bass 30 ms late, as an example of the "pushed/laid-back" degradation, and 60 ms late, as an example of "synch. loss". Then, participants took two test trials using the same music excerpt, with the bass early and late compared to the drums. The following time offset values were used: 30 ms, 45 ms, and 60 ms. The familiarization phase lasted between 5 and 10 minutes, after which the participant was invited to confirm that everything was clear before starting the actual test. Participants were invited to take one or two short breaks during the test if needed.

3 Perceptual test results

A panel of 19 people (3 females and 16 males, aged 22 to 52) participated in the test. All testers reported normal hearing. None of the testers participated in the test presented in [2] and the test described in the present paper took place before any presentation of the earlier study. Most of the participants were used to performing critical listening, as the majority are musicians and/or have a background in sound engineering. However, listener profile was not found to have a significant influence on the results in previous studies [5, 2] and is thus not discussed in the following. The test lasted between 20 and 60 minutes, with an average duration of 34 minutes.

3.1 Analysis of variance

A Kolmogorov-Smirnoff test (*kstest* function in Matlab) indicated that quality ratings were normally distributed for every stimulus. Therefore, parametric methods could be used to analyze the test results. An analysis of variance (ANOVA) was performed on quality ratings with the following factors: test participant ($N = 19$), track ($N = 6$), direction of degradation ($N = 2$), and time offset ($N = 5$). The participant factor was treated as random while the other factors were treated as fixed. Main factor effects were analyzed, as well as first-order interactions. The analysis was done using the *anovan* Matlab function.

The ANOVA indicated that every factor had a significant effect on the quality rating. However, the effect size of the time offset factor ($F(4,1135) = 306.25$, $p < 0.001$) was much greater than that of the track ($F(5,1134) = 60.77$, $p < 0.001$), the direction of degradation ($F(1,1138) = 50.25$, $p < 0.001$), and interactions. This result suggests that, in our test, the time offset is the main factor in the perception of musical groove degradation. In the following, we examine the role of other parameters.

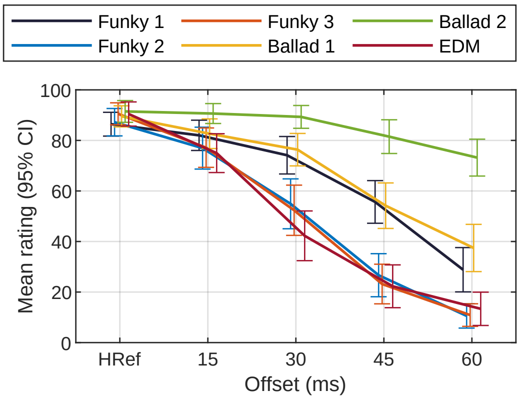

Fig. 2: Mean quality rating and 95% confidence intervals obtained for the 6 tracks, as a function of the time offset (merging early and late offset directions).

3.2 Influence of the track

Let us start by analyzing the influence of the track. For this analysis, we merged quality ratings obtained for the early and late conditions. Figure 2 shows the mean quality ratings and 95% confidence intervals (CIs) as a function of the time offset for the six tracks.

First, hidden references are generally well identified as the corresponding mean ratings are around 90-100 for every stimulus, with CIs comprised between 80 and 100. As the time offset increases, differences between tracks can clearly be observed, even for the smallest time value (15 ms). Broadly speaking, *Ballad 2*, which corresponds to the Ballad with the keyboard shifted, seems less sensitive to time offsets than other tracks. Note that, for this track, the ratings obtained for the low-quality anchor (60 ms offsets) are for the most part above 60, which cannot be qualified as low quality. On the contrary, Tracks *Funky 2*, *Funky 3*, and *EDM* are the most critical. This difference in the susceptibility of tracks to time offsets is observed for different tracks generated from the same musical excerpt: quality ratings are better for *Funky 1* (drums and bass) than *Funky 2* (drums and guitar) and for *Ballad 2* (drums and keyboard) than *Ballad 1* (drums and guitar).

3.3 Influence of the instrument

Early and late time offset directions are now separated to investigate the influence of the shifted instrument

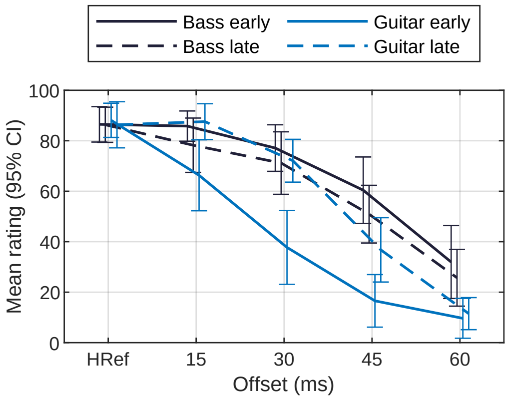

Fig. 3: Mean rating and 95% confidence intervals for Tracks *Funky 1* and *Funky 2*, as a function of the time offset and offset direction.

within the same musical excerpt. Figure 3 compares the mean rating and CIs obtained for Tracks *Funky 1* (guitar shifted) and *Funky 2* (bass shifted), as a function of the time offset.

It can be observed that ratings obtained for *Funky 2* (bass) do not depend on the offset direction. Average ratings seem slightly lower for late offsets but CIs overlap almost entirely, therefore the effect is not statistically significant. On the contrary, for *Funky 1* (guitar), ratings obtained for the early condition are significantly lower than those obtained for the late condition with offsets of 15, 30, and 45 ms. Note that, with 15 ms time offsets, only the track with the guitar shifted early is significantly different from the hidden reference (t-test $p < 0.001$).

The difference between the ratings obtained for the guitar and bass, and between the ratings obtained for early and late offsets, can also be observed by looking at the "Synch. Loss" and "Pushed/Laid-back" degradations reported by the participants, as illustrated in Figure 4. First, more than 50% of the participants perceived a loss of synchronization when the guitar was shifted 30 ms ahead of the drums. By contrast, less than 20% of the participants perceived a loss of synchronization with the other tracks: for those tracks, stimuli were perceived as pushed/laid-back in the majority and ratings were relatively high (around 70/100). With a time offset of 45 ms, about 90% of the listeners reported a loss of synchronization with the guitar shifted early and the corresponding ratings were poor, while for the

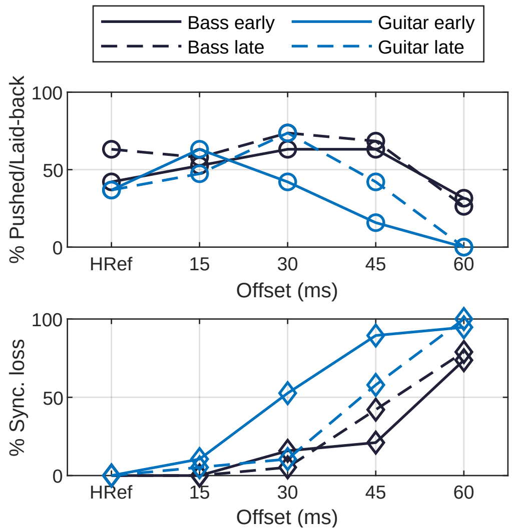

Fig. 4: Percentage of "Pushed/Laid-back" (circle) and "Synch. loss" (diamond) reports for Tracks *Funky 1* and *Funky 2*, as a function of the time offset and offset direction.

guitar shifted late only 60% of the listeners reported a synchronization loss. With the same offset value, early and late bass stimuli were reported as pushed/laid-back by the majority of listeners, and most quality ratings were above 50/100.

The influence of the instrument can also be analyzed by comparing the ratings obtained for Tracks *Ballad 1* (arpeggio guitar) and *Ballad 2* (keyboard), as illustrated in Figure 5. Track *Ballad 2* seems far less sensitive to time offsets than Track *Ballad 1*. In the case where the guitar is shifted, quality ratings decrease as the offset increases, reaching 50/100 with a time offset of 45 ms. With a 60 ms time offset, ratings are below 50/100, and the majority of listeners reported a synchronization loss. In the case where the keyboard is shifted, ratings decrease only for offsets longer than 45 ms and are not significantly different from the hidden reference at 60 ms (t-test: $p = 0.0538$). Note that, when the keyboard was shifted, almost no one reported a loss of synchronization, but the percentage of pushed/laid-back reports slightly increased with the offset. No significant difference can be observed between the ratings obtained for early and late conditions with the *Ballad* tracks.

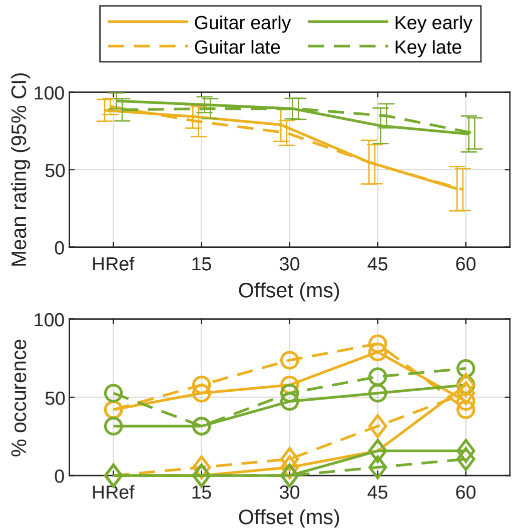

Fig. 5: Quality ratings (mean values and 95% CIs, top), and percentages of "Pushed/Laid-back" and "Synch. loss" (circles and diamonds, resp., bottom) for Tracks *Ballad 1* and *Ballad 2*, as a function of the time offset.

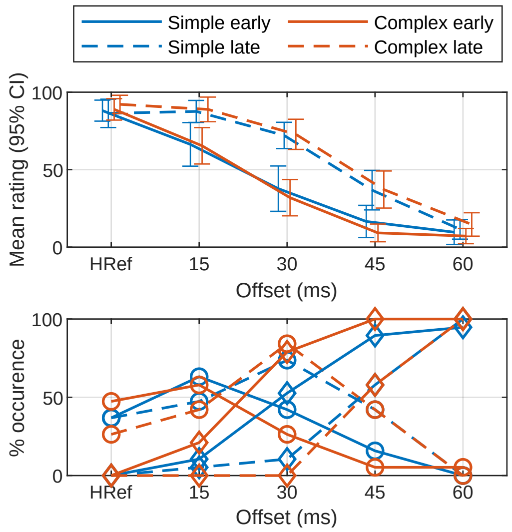

Fig. 6: Quality ratings (mean values and 95% CIs, top), and percentages of "Pushed/Laid-back" and "Synch. loss" (circles and diamonds, resp., bottom) for Tracks *Funky 2* (simple) and *Funky 3* (complex), as a function of the time offset.

3.4 Influence of the complexity

The influence of musical complexity can be examined by comparing Track *Funky 2* (drums and guitar) with Track *Funky 3*, which is the same as *Funky 2* with additional bass and keyboard. In both cases, the guitar was the only instrument that was shifted in time. Figure 6 compares the ratings obtained for these two tracks.

No difference can be observed between the ratings obtained for the simple and complex conditions (*Funky 2* and *Funky 3*, respectively). However, a slight difference in the reports of synchronization loss can be observed. With 30 ms time offsets, slightly more listeners perceived a loss of synchronization with the complex condition than with the simple condition.

3.5 Influence of the tempo

In [2], time offsets were computed as a function of the beat length, with the underlying assumption that, for a fixed offset value, listeners would perceive a stronger groove degradation with faster tempos. By contrast, in the test presented here, stimuli were generated using the same offset values, regardless of the Track's tempo. Figure 7 compares the ratings obtained for Tracks *Ballad 1* (slow) and *Funky 2* (fast). In both cases, the shifted instrument is a guitar. Note that, in the figure, the offset values are expressed in beats: offsets of 15, 30, 45, and 60 ms correspond to approximately 1/32, 2/32, 3/32, and 4/32 beats for *Funky 2* (124 BPM), and 1/48, 2/48, 3/48, and 4/48 for *Ballad 1* (84 BPM).

Ratings obtained for the slow condition (early or late offset) are higher than those obtained for the fast condition with early offsets but lower than those obtained for the fast condition with late offsets. However, despite the difference between the two ways of playing the guitar, ratings are closer when they are compared on the tempo-relative time scale. Hence, the tempo could be an important factor in the perception of musical groove degradation.

3.6 Influence of the music genre

Tracks *Funky 1* and *EDM* both consist of a rhythmic instrument and a bass, and their tempo is almost the same, around 125 BPM. Figure 8 compares the ratings and percentages of synchronization loss and "pushed/laid-back" reported by the users for the two tracks.

Ratings obtained for the early condition of *EDM* are significantly lower than those obtained with the other stimuli. The 15 ms early EDM stimulus is perceived as

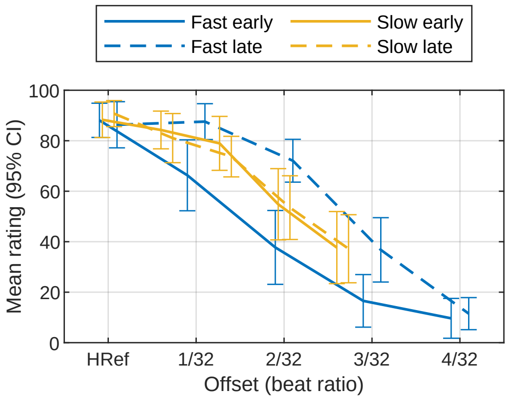

Fig. 7: Mean ratings and 95% confidence intervals as a function of the time offset for Tracks *Funky 2* and *Ballad 1*.

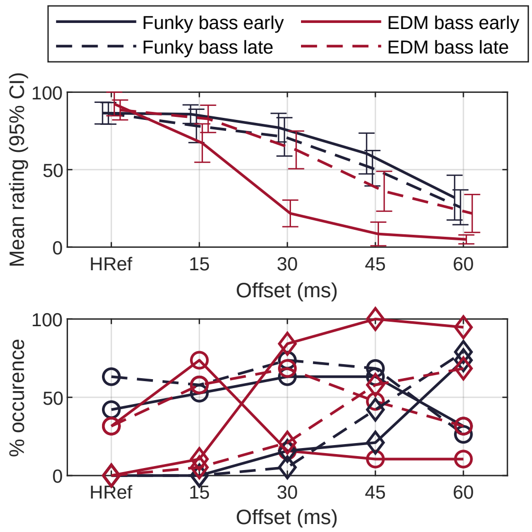

Fig. 8: Quality ratings (mean values and 95% CIs, top), and percentages of "Pushed/Laid-back" and "Synch. loss" (circles and diamonds, resp., bottom) for Tracks *Funky 1* and *EDM*, as a function of the time offset.

significantly different from the hidden reference (t-test: $p < 0.01$). Then, as the offset increases, ratings drop to very low values, from around 25/100 for a 30 ms offset to nearly 0/100 for a 60 ms offset. This drop in ratings is correlated with a sharp increase in the percentage of reported synchronization loss. On the other hand, there is no significant influence of early/late condition on the perceived quality for Track *Funky 1*.

4 Discussion

4.1 Main results

Time-shifting instruments with respect to each other clearly impacts the way we perceive the rhythmic characteristics, or groove, of musical pieces. However, the amplitude of this impact seems to depend on various parameters. The nature of the shifted instrument, the style in which this instrument is played, or the music genre appear to be major factors, as shown in Figs. 3, 5, and 8. The tempo of the music also seems to be an important parameter, as shown in Section 3.5

Early and late time offsets lead to very different perceived groove degradations with some of the tested stimuli. In the tested stimuli, early time offsets generally lead to lower quality ratings than late offsets. However, it could be the opposite with other stimuli, other music styles, or instrument combinations. For instance, late offsets of the bass lead to greater groove degradations than early offsets in Track *Funky 1*. A similar observation was done in [5]. Nevertheless, we have to consider the worst-case scenario, because an instrument that is late on one side of the audience is early on the opposite side.

The results of the present study are consistent with those presented in [2]. For instance, a change is perceived in the groove with offsets as short as 15 ms with the *Funky* guitar or the *EDM* bass. Also, similar to what was observed in the previous study, reports of a loss in instrument synchronization are correlated with a drop in the quality rating, whereas reports of a "pushed/laid-back" groove are more difficult to interpret and are not necessarily correlated with specific rating levels.

4.2 Mixing guidelines for frontal system

The test results provide new insights on the issue of mixing for immersive live events, which allows us to build up on the mixing guidelines proposed in [2]. The impact of time offsets on the perceived groove quality depends on different factors, which must be taken into account when panning sources over loudspeakers. First, let us consider the same large-scale frontal loudspeaker deployment as in [2] (see Figure 9). This system consists of 5 main speakers, numbered 2 to 6, which span

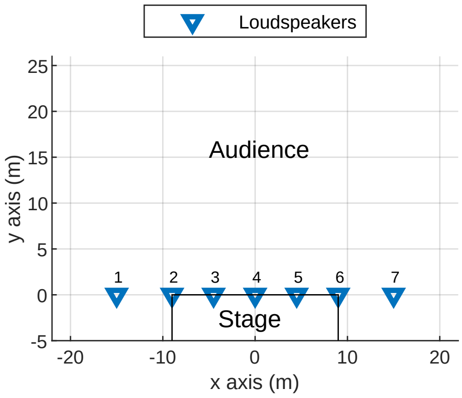

Fig. 9: The geometrical configuration of a large-scale loudspeaker deployment used for a live event with immersive sound.

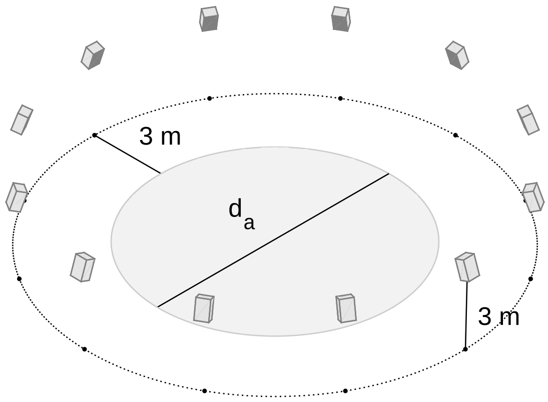

Fig. 10: 360° loudspeaker setup.

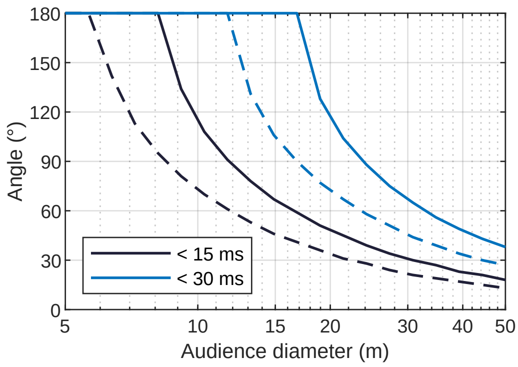

Fig. 11: Maximum speaker angular distance such that the propagation time difference is less than 15 or 30 ms for 75% (continuous line) or 95% (dashed line) of the audience.

the entire width of an 18 m wide stage, and two spatial extension speakers, numbered 1 and 7.

Assuming that the rhythmic components (drums) are positioned at the center (Speaker 4):

- the *EDM* bass or the *Funky* guitar could only be located in the 3 center loudspeakers (Speakers 3 to 5) because changes in the groove can be perceived with time offsets as short as 10 to 15 ms with these sources;
- the *Funky* bass and the arpeggio guitar (*Ballad*) could be panned anywhere along the performance area (loudspeakers 2 to 6); at these positions time offsets are lower than 20 ms, resulting in almost imperceptible degradations with these sources;
- the keyboard in *Ballad* could be positioned anywhere over the system (Speakers 1 to 7) as, even with 45 ms time offsets, almost no difference was perceived with this source.

4.3 Mixing guidelines for 360° system

We now consider the case of a 360° loudspeaker system, whereby speakers are positioned all around the audience. Setups of this kind can create timing issues in that their purpose is to make it possible to pan objects anywhere around the circle. As shown in [6], with a 50 m diameter circular deployment, the propagation time difference between loudspeakers can reach 150 ms on the border of the audience area.

To begin, we ran numerical simulations to determine how the dimensions of a 360° loudspeaker system impacted the difference in propagation times between speakers, for any position in the audience. In our simulations, the speakers were distributed around a circle, as shown in Figure 10. The audience was assumed to be located inside a circle of diameter d_a, with a listening height of 1.6 m. The speakers were located 3 m outside the audience area at a height of 3 m. For different values of d_a, we calculated the propagation times between the speakers and a large number of points in the audience. For each audience diameter, we then estimated the maximum speaker angular distances such that, over a given percentage of the audience area, propagation time differences were below 15 or 30 ms, respectively.

The simulation results are shown in Figure 11. In the case of an audience area with a diameter below about 7 m, the sound mixer can freely pan audio objects anywhere around the audience. Indeed, with a system of this kind, the propagation time difference is below 15 ms for the vast majority of the audience even when sources are panned in diametrically opposed directions. Hence, our perceptual test results indicate that there should be no perceivable degradation of the groove. Such a setup could fit a 150-person standing audience, for instance, or could be used as a pre-production studio setup.

In the case of larger setups, the angle between two rhythmically critical instruments (such as the drums and the guitar in Track *Funky 2*) must be reduced to ensure that the quality of the groove remains constant across the audience. Note that the maximum opening angle drops very quickly as the audience area diameter increases. With an audience diameter of 12 m, rhythmic instruments should be located within a 90° angle to make sure that 75% of the audience perceive no degradation. With an audience diameter of 20 m, this angle again drops to 45°. On the other hand, less critical instruments, such as the arpeggio guitar in Track *Ballad 1*, can be panned anywhere around the circle with a 12 m audience diameter and up to 120° from the rhythmic instruments with a 20 m audience diameter.

4.4 Constraints with 360° system

The purpose of a 360° loudspeaker deployment is the freedom to pan objects anywhere around the audience. Often, setups of this kind consist of 12 sources, which are regularly distributed around a circle. Consequently, in principle, each of the 12 sources must be able to reproduce any audio object and have a full-range frequency response (usually with subwoofers) with sufficient sound pressure level (SPL) capacity. Nevertheless, beyond a certain audience area size, our simulations show that the main rhythmic instruments should be located within a relatively narrow opening angle. With a 20 m audience area diameter, this angle is about 45°, which means that any two successive sources of the 12 must be able to reproduce the entire rhythmic section.

Therefore, in the case of large audiences, it may be more efficient to favor one direction, where the rhythmic instruments are panned. This direction hence becomes the frontal direction and must be equipped with enough full-range speakers to reach the required SPL. On the side and at the back of the audience, sources could be less powerful. The angular definition of the frontal system is generally higher than that of the typical 360° setup: the frontal system proposed in Figure 9 includes 5 sources in a 40° angle. Beyond that, in the field, it is a common practice to reinforce the 3 center loudspeakers of the frontal system (loudspeakers 3 to 5 in Figure 9) so that they can reproduce the core of the rhythmic instruments with sufficient headroom, particularly at low frequencies. In this sense, frontal systems with extensions and surrounds seem to better fit the large-scale immersive sound use case.

4.5 Limitations of our findings

The results of this study provide new insight into understanding the impact of musical tempo and complexity on perceived groove degradation. However, due to practical constraints on the duration of the listening test, we have not collected enough data to conclude on these matters. The tempo seems to influence groove quality, as slower tracks appear less sensitive to time offsets between instruments. However, the link between tempo and perceived quality is still not precisely established. On the contrary, the complexity of musical excerpts seems to have very little influence on the perceived groove quality, but this factor was only tested for one excerpt in our test.

Some of the differences observed between tracks could be explained by the analysis of the original stimuli. For instance, the quality of Track *Funky 1* is perceived worse when the bass is shifted behind than when it is shifted ahead, in opposition to most of the other stimuli. Examining the recordings, the bass is generally played slightly late (up to 30 ms) with respect to the drums, therefore small early offsets tend to bring the bass closer to the beat. This could also explain why the results obtained for Tracks *Funky 1* and *EDM* are different, to some extent: in *EDM*, the bass is played strictly on the beat. Hence, the differences observed in the perception of the different tracks could be related to their respective rhythmic characteristics, which include tempo and rhythm accuracy, rather than the music genre in itself.

Further, the ecological validity of the study can be questioned. It is possible that, in a live concert situation, listeners would be more tolerant of changes in instrument synchronization than indicated by the results of our test. Nevertheless, it could be argued that perceivable changes in the musical groove go against the artist's intention. In addition, in a real situation, level differences between spaced speakers would be perceived and a real room would generate reflections and reverberation. Both elements could influence groove perception.

Lastly, analyzing the signal characteristics of the different instruments could help in understanding some of the differences observed in the perceived groove degradation. In particular, the envelope (attack strength, decay time, sustain duration, etc.) would be an interesting parameter to consider to determine objective criteria that govern how sensitive different instruments or tracks are to time offsets. For example, in Track *Ballad 2*, keyboard notes have a long sustain, which may explain why time offsets up to 45 ms barely affect the perception of the groove for this track.

In the absence of rules that predict whether an instrument will pose a problem for the preservation of the musical groove within an audience, the authors recommend simulating how the mix will sound at different positions in the audience. This is important for frontal systems but even more so for 360° deployments, as suggested in [7]. In addition to propagation time differences, such simulation can also emulate level differences between objects, which were not tested in this study.

5 Conclusion

This study builds upon a previous paper, which investigated how propagation time differences occurring within large-scale immersive sound systems could modify the perception of musical groove. This study demonstrated that, in the case of large frontal setups, propagation time differences could reach several dozen ms, which modifies significantly the perceived rhythm characteristics of musical pieces.

In the present study, we have run a perceptual experiment with additional stimuli, such that we could investigate the influence of different factors: the nature of the instrument, the music genre, the music complexity, and the time offset direction. Our results show that, for a fixed offset and with the same music excerpt, the perceived groove degradation strongly depended on the type of instrument. With certain tracks, a difference between the impact of early and late time offsets could also be observed, the groove being generally more degraded by early offsets than late ones. Lastly, the rhythmic characteristics of the track (tempo, music genre, or rhythmic accuracy) must be considered.

The potentially severe groove degradations perceived when shifting instruments with respect to each other impose some constraints when mixing for a large audience. To preserve the musical quality, core rhythmic instruments should not be spread over speakers that are too distant from each other. This applies to frontal systems, but even more so to 360° loudspeaker setups, for which propagation time differences can be even larger.

In order to complete this study, a large-scale perceptual test, using a wide range of stimuli, should be carried out. Mixing guidelines could then be elaborated by relating the results of the perceptual experiment to objective signal characteristics.

References

[1] Kennedy, D., "Perception of Stereo vs. Immersive Live-Sound Systems," *Front of House Magazine*, February, pp. 34–35, 2020.

[2] Mouterde, T., Epain, N., Moulin, S., and Corteel, E., "On the perception of musical groove in large-scale events with immersive sound." in *International Conference on Spatial and Immersive Audio*, Audio Engineering Society, 2023.

[3] Butterfield, M., "Participatory discrepancies and the perception of beats in jazz," *Music perception*, 27(3), pp. 157–176, 2010.

[4] Frühauf, J., Kopiez, R., and Platz, F., "Music on the timing grid: The influence of microtiming on the perceived groove quality of a simple drum pattern performance," *Musicae Scientiae*, 17(2), pp. 246–260, 2013.

[5] Matsushita, S. and Nomura, S., "The asymmetrical influence of timing asynchrony of bass guitar and drum sounds on groove," *Music Perception: An Interdisciplinary Journal*, 34(2), pp. 123–131, 2016.

[6] Corteel, E., Le Nost, G., and Roskam, F., "3D audio for live sound," in *3D Audio*, pp. 19–42, Routledge, 2021.

[7] L-Acoustics, "L-ISA Preproduction, training module," 2022.