Audio Engineering Society

Conference Paper 17

Presented at the International Conference on Spatial and Immersive Audio
2023 August 23–25, Huddersfield, UK

On the perception of musical groove in large-scale events with immersive sound

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

Immersive audio is increasingly used in large-scale live music events. The dimensions of the audience area impose that propagation times from several loudspeakers to a given audience position can be significantly different. This may be perceived by listeners as a loss of time synchronization between sound sources, which in turn affects the perception of musical groove. In this paper, we first investigate the range of propagation time differences that can occur with large-scale loudspeaker deployments. The results of a listening test confirm that time differences may degrade the rhythmic characteristics. The degradations may depend on the musical content but not on the spatialization. Mixing guidelines and methodologies are finally proposed to overcome the potential issues.

1 Introduction

Over the past decade, immersive audio has become more common in entertainment, notably in live events. Following this trend, live sound slowly evolves, from using channel-based mixing and left/right sound systems to object-based mixing and systems that now often span the entire performance area. Compared to the former approach, the latter offers better localization accuracy and audio-visual consistency. Such immersive sound systems are often completed by lateral extension sources, which widen the panorama, and surround and overhead speakers are sometimes used to provide 360-degree or 3D reproduction, respectively. Thus, immersive audio leads to complex sound systems, comprised of numerous loudspeakers distributed across venues.

In small venues, the distance between loudspeakers is relatively small, therefore they are mostly aligned in time. In large venues such as stadiums or arenas, however, loudspeakers may be spaced dozens of meters apart from each other. Hence, the distances that separate the speakers from a given point in the audience may be significantly different. Differences in distance imply disparities in the time it takes for sound waves to propagate from the loudspeakers to a listener, which may result in two distinct issues.

The first issue occurs when a given sound object is panned across several loudspeakers. In this case, propagation time differences may result in spectral coloration or, when the time difference is sufficiently large, may be perceived as an echo. This issue has been addressed in [1] and implies constraints in the loudspeaker system design. The second issue occurs when different audio objects are panned to speakers that are distant from each other. In this instance, propagation time differences modify the synchronization between sound ob-

jects (see Figure 1) and may introduce microtiming in music contents, hence altering the perceived "groove".

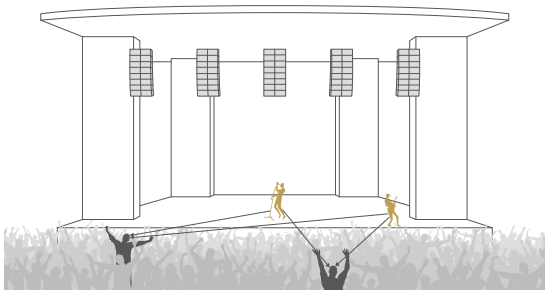

Fig. 1: Illustration of the propagation time difference between two sound sources at different positions in a large audience.

In this study, we investigate how immersive sound systems may affect the perception of musical groove in the context of large-scale public events. In the first part, we determine the range of time differences that can occur in events of this kind and compare delay-based with amplitude-based algorithms. These values are compared to that associated with the notion of microtiming in music. In the second part, we report the results of a listening test, which assessed how such time differences between instruments affect the perception of music. Lastly, we discuss how to account for potential issues early in the creation of the spatial mix to preserve the groove.

2 Background

In this section, we briefly survey the notion of groove in music and describe how a large-scale sound system can impact the perception of the groove at different positions in the audience.

2.1 Microtiming and the notion of groove

In the context of music, *microtiming* consists in time offsets, on the order of milliseconds, introduced by a performer around an interpretation that would perfectly follow the score, rhythmically speaking. Such time offsets can either be negative (i.e. sound events are played early), leading to a *pushed* interpretation, or positive (sound events are played late), which results in a *laid-back* or behind-the-bar style. Microtimings are often considered as an essential element of the *groove*, which can be described as the pleasant feeling of being drawn into dancing along with the music. In jazz music particularly, Keil [2] claims that the groove originates from microtiming within the drum beat, between the bass and drums, or between the rhythmic section and the soloists. In [3], the author reports numerous studies stating that, in jazz music, time offsets between the bass and drums of less than 20 ms are not perceived as an error in timing. The author is not confident, however, that microtiming helps create the groove.

In [4], the groove's "dual nature" is described: the groove results from the interaction between a particular rhythmic structure and the musicians' interpretation of this rhythmic structure. In some musical styles, jazz music, for instance, microtimings are intensively used, while in some other music styles, such as rock music, the instruments are mostly snapped on the grid. The study in [4] is focused on microtiming in rock and pop music and investigates the influence of micro rhythmic deviation in a drum pattern. The bass drum or the snare drum is shifted ahead or behind the hi-hat. Deviations of 15 and 25 ms are tested and compared to the quantized version of the drum beat snapped on the grid. The results show that the quantized version is perceived as having the best quality. In addition, an early degradation is perceived as being worse than a late degradation.

In [5], the influence of microtiming between a drum pattern and a bass is studied. The bass is shifted by times ranging from -62.5 to 62.5 ms against the drum, which corresponds to offsets of -4/32 to 4/32 beat lengths (500 ms beat length with a tempo of 120 bpm). Delays of ± 15.63 ms (1/32 beat) are not presented as they were assumed to be imperceptible. It is shown that a slightly early bass can be perceived as "as good" as the synchronous bass and that there is no microtiming that significantly improves the groove. This indicates that the instruments can be synchronized with a certain flexibility, and that small time offsets do not necessarily affect how the groove is perceived.

In summary, the literature draws a strong link between the perception of musical groove and the timing of the instruments, but there remain questions regarding the role played by the musical style, tempo, spatialization, and offset duration in this regard. Nevertheless, it is clear that time offsets induced by a sound system between instruments could impact how the groove is perceived.

2.2 Immersive sound systems at various scales

An overview of the available 3D audio techniques for live sound is proposed in [6]. Most of the technical solutions for live sound rely on a minimum of 5 full-range systems spanning the width of the stage. They are optionally complemented by extensions over the width, surround, and/or overhead loudspeakers. Compared to the traditional stereo system, the objective of these immersive systems is to improve the localization accuracy through a large part of the audience and the dimensions of the spatial mix. Such spatialization solutions have been available for a long time in other entertainment sectors (movie theaters for instance) but need to be adapted for live sound, especially because of the size of the venues where the systems are deployed.

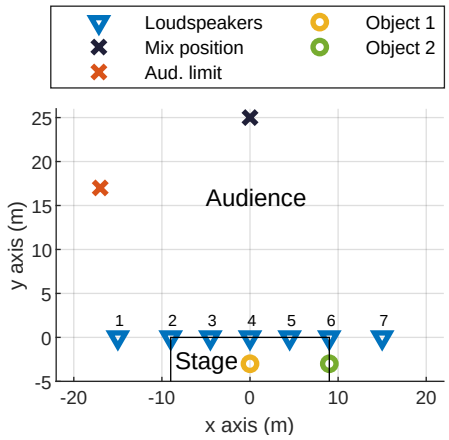

Fig. 2: Geometrical configuration of a large-scale loudspeaker deployment used for a live event with immersive sound.

In this paper, we focus on frontal sound systems with higher channel counts than a basic Left Right system. Such systems have the ability to favor audio visual localization coherence and spatial unmasking of spatially separated sound objects, thus enhancing immersion in a traditional show with performers onstage [6]. In the following, a large-scale configuration of seven loudspeakers is considered. The system consists of five loudspeakers that span the entire width of an 18 m wide performance area, and two additional "extension" speakers that widen the soundscape on the left and right sides of the stage, as illustrated in Figure 2. We define the x-axis along the venue width and the y-axis along the venue depth, with positive y toward the audience. The speakers are denoted Speaker 1 to 7 and are located at $x = -15, -9, -4.5, 0, 4.5, 9$ and 15 m, respectively. Two listening positions are considered in the audience: the mixing position, in the axis of the system, 25 m from the stage, and the audience limit position, on the left side of the audience, at the limit of the area covered by the system (-6 dB). From the mixing position, the angular width of the stage is 40°. The sound system is used to spatialize objects that can be placed anywhere along the loudspeakers, as shown in Figure 2.

Let us now consider a studio equipped with a scaled-down version of the loudspeaker system described above. A studio of this kind may be used when preparing the production of a live show in the venue described above, for instance. The loudspeaker system is scaled down to a total width of 8.6 m, with speakers located at positions -4.3, -2.5, -1.25, 0, 1.25, 2.5, and 4.3 m along the x-axis. The mixing position is located 6.9 m from the sound system to obtain the same angular width as experienced in the live event setup.

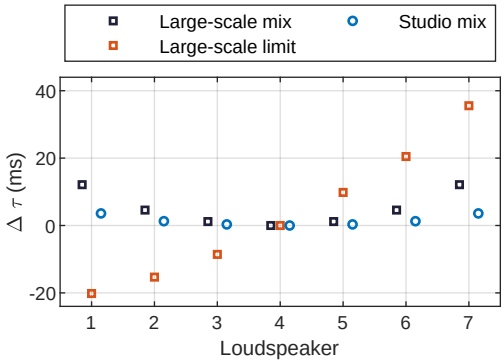

Fig. 3: Propagation time differences observed between the loudspeakers with two different immersive sound systems. We compare the time differences observed in: a) a pre-production studio at the mixing position; and b) a large-scale deployment, at the mixing position and audience limit position.

The differences in sound propagation times between the speakers and the listening positions observed in the studio environment are very different from that measured in the large-scale environment, as will now be shown. Using the central speaker as our temporal reference, Figure 3 presents the difference in sound propagation delays for the different loudspeakers occurring in the studio, as compared to that occurring in the large-scale

venue. A negative delay value means that a sound coming from the considered speaker arrives before a sound in the central speaker. For the studio environment, time offsets are shown for the mix position (blue circles). For the real venue, two listening positions are considered: the mixing position (black squares) and the audience limit position on the side of the audience (red squares). Clearly, the disparity in sound propagation times strongly depends on the environment. In the studio, time differences remain on the order of a few milliseconds regardless of the considered speaker, because the loudspeakers are close to each other. As well, note that for a different listening position in the studio, propagation time differences would remain on the same order of magnitude, as the room is relatively small. On the other hand, with the large-scale deployment, we observe much larger time differences, reaching 20 to 40 ms in the case of the audience limit position. In the following sections, we show that such time differences could lead to audible modifications of the groove, regardless of the employed spatialization method. Indeed this is a consequence of the physical configuration of the sound system, namely the distance separating the loudspeakers.

2.3 Influence of panning on sound source synchronization

So far, we have considered only the physical characteristics of the sound system. However, in an immersive sound event, sound sources are typically modeled as audio objects that are spatialized using the sound system. The spatialization algorithm may modify further the relative timing of the sound sources observed at a given position in the audience. In this section, we briefly review the main spatialization methods employed in large-scale live events and describe how they may impact sound source synchronization.

The first family of spatialization methods is delay-based panning, such as Wave Field Synthesis (WFS) [7]. WFS aims at recreating virtual sources by applying a time offset to the signals played by the loudspeakers. WFS is employed in several live sound solutions: d&b Audiotechnik Soundscape [8], Adamson Fletcher Machine [9], to name a few. In large-scale deployments using 5 loudspeakers above with 3 to 5 m spacing, the wave field is only accurately recreated at very low frequencies, typically below 100 Hz. From a perceptual standpoint, sound localization is mostly driven by the precedence effect.

The second family of spatialization methods is amplitude-based panning, whereby virtual sound sources are recreated solely by controlling the amplitude of the signals sent to the loudspeakers. Vector-Base Amplitude Panning (VBAP) [10] is probably the most widely used amplitude panning technique. For a given object position, the algorithm selects the two or three loudspeakers that surround the object (in 2D and 3D, respectively) and adapts their gains depending on the angular difference between the object and the loudspeakers. L-Acoustics' L-ISA [11] is based on this technique with a few additional improvements. One of the advantages of VBAP is the possibility to place an object exactly in the direction of a loudspeaker (snapped object), which maximizes the proximity effect.

We now investigate the range of time offsets introduced between two audio objects when they are panned over the large-scale loudspeaker deployment described in the previous section. VBAP does not use delays to pan an object. In the simplest case, if the object is localized in the direction of a loudspeaker, the sound is only reproduced by this loudspeaker. In this case, the propagation time, τ_k, between object k and the listener is:

$$\tau_k = \frac{|\vec{x} - \vec{x}_i|}{c}, \tag{1}$$

where $\vec{x}$ is the position in the audience, $\vec{x}_i$ the position of the loudspeaker, and c the speed of sound. With VBAP, the time reference is the loudspeaker itself.

With WFS, every loudspeaker contributes to synthesizing the wavefront for each object. Loudspeaker gains and delays are computed according to the object's position, including the distance at which it is located. For instance, objects 1 and 2 are located 3 m behind the speakers in the example shown in Figure 2. The WFS algorithm involves delays corresponding to the distance between the object (the virtual source) and the loudspeakers. In addition, because every speaker contributes to recreating the virtual sound source, the perceived object location will be driven by the first wavefront reaching the listener. The arrival time of the first wavefront reaching the listener for object k is given by:

$$\tau_k = \min_i \left(\frac{|\vec{x}_i - \vec{x}_k|}{c} + \frac{|\vec{x} - \vec{x}_i|}{c} \right), \tag{2}$$

where $\vec{x}_k$ is the position of the object. Hence, with WFS, the time reference is almost the object itself.

Let us consider the propagation time difference, $\Delta\tau$, between objects 1 and 2:

$$\Delta\tau = \tau_2 - \tau_1. \quad (3)$$

Figure 4 presents the value of $\Delta\tau$, observed at the mixing and audience limit positions, for different object configurations. In every configuration, Object 1 is located on the center axis, while Object 2 is panned to the directions of Loudspeakers 1 to 7. Note that, with VBAP, the object's distance is not taken into account, therefore a single $\Delta\tau$ value is obtained for a given panning direction. On the contrary, with WFS, the delay difference depends on the objects' respective distances. Therefore, with WFS, object distances ranging from 1 to 8 m behind the speakers were considered and a distribution of $\Delta\tau$ values was obtained for each panning direction. Also, for WFS, the delay corresponding to the shortest distance between the object and the loudspeakers was subtracted from the delay values. This normalization is often employed in live events to reduce latency.

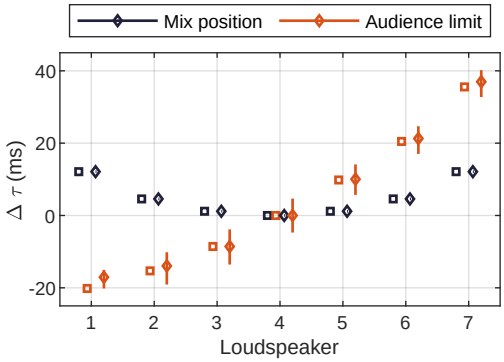

Fig. 4: Propagation delay difference between a sound object panned to the center and an object panned successively to the direction of the other loudspeakers, at the mix position and at the audience limit position. Comparison between VBAP (single value: squares) and WFS (statistics: diamond for the median, vertical line for the 5th to 95th percentile interval).

It can be observed that the propagation time differences obtained when using WFS are very similar to that obtained when using VBAP, regardless of the objects' distances. In both cases, the time offset between Objects 1 and 2 is maximum when Object 2 is located in the direction of Loudspeaker 7. Considering only the loudspeakers that span the performance area (Speakers 2 to 6), the propagation time differences reach around 5 ms at the mixing position, but 20 ms at the audience limit position. Propagation time differences up to 40 ms are observed when panning objects to the extension speakers (speaker 7).

In summary, using a large-scale sound system to spatialize sound objects may result in significant time offsets between the objects for certain listening positions in the audience. These offsets originate from the physical distance between speakers and occur regardless of the employed spatialization method. Note that we derived these results for the case of a frontal sound system, but even larger time offsets could be expected in the event where surround and overhead speakers were used. In the case of off-center positions in the audience, time offsets can be larger than 20 ms, which has been shown to induce changes in the perceived quality or groove of music [3, 4]. In the following, we investigate how time shifts in this range of duration affect the perception of musical groove.

3 Perceptual experiment

This section describes a perceptual experiment to evaluate how large-scale immersive sound systems impact the perceived groove of a given piece of music.

3.1 Conditions and stimuli

Three different audio tracks were used for the perceptual test. They were selected so as to provide examples of various music genres. The tracks were extracted from pieces of music for which we have multi-track recordings:

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

Each track combines a rhythmic part with another instrument. Tracks 1 and 2 associate drums with a guitar (a funk guitar and an arpeggio guitar, respectively). Track 3 associates an electronic beat and a synthetic

bass. Note that these tracks and instrument combinations were selected, through listening sessions, as being particularly critical in terms of instrument timing.

The audio stimuli used in the test were created by shifting the harmonic instrument (guitar or bass) ahead of the rhythmic instrument. Offsets of this kind have been shown to be perceptually more critical [4, 5]. Four levels of time offset were used: 1/48, 2/48, 2/32, and 3/32 beats. This corresponds to about 10, 20, 30, and 45 ms for tracks 1 and 3, and 15, 30, 45, and 67.5 ms for Track 2. Note that the relative timing of the different elements that form the rhythmic part was not modified.

In addition, the impact of sound spatialization was investigated. In the stimuli, the rhythmic part was always located at the front (0° azimuth), while the associated instrument was either located at the front or 30° to the right. As the experiment was conducted using headphones, the stimuli were synthesized using Head-Related Impulse Responses (HRIRs). To reflect the listening conditions of a live event as closely as possible, HRIRs were measured outdoors at the L-Acoustics headquarters, using a Neumann KU100 artificial head and an L-Acoustics Kara II loudspeaker.

To summarize, the following independent variables were tested: the music track (different musical genres and instrument associations), the location of the harmonic instrument (0° or 30° azimuth), and the time offset duration. The test was divided into 12 successive trials, corresponding to two occurrences for every combination of track and spatialization conditions. The order in which trials were presented was randomized.

3.2 Test methodology

The interface of the test is shown in Figure 5. The test method was inspired by the MUltiple Stimuli with Hidden Reference and Anchor (MUSHRA) method. Participants were presented with an explicit reference, which consisted in the track as originally 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 labeling of the different stimuli under test (from 1 to 5) was picked at random for every trial. The stimuli with a time offset of 3/32 beats were used as a low anchor: the duration of these offsets was larger than that simulated for large-scale sound systems but was expected to provoke an invariably detectable degradation to the track's musical quality.

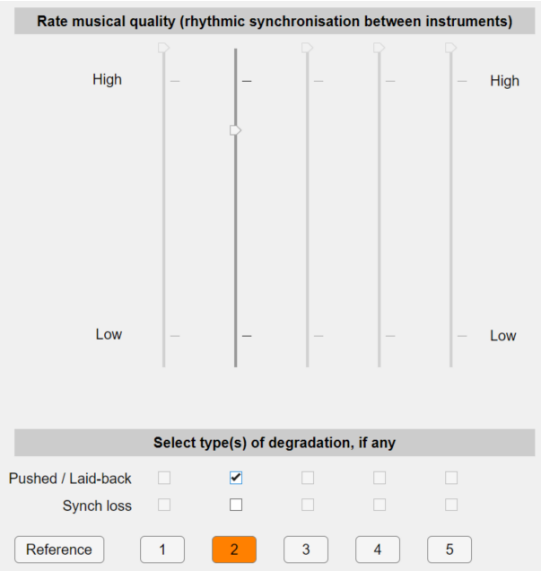

Fig. 5: Graphical user interface of the application used for the perceptual test.

Participants were asked to rate the rhythmic synchronization between instruments for each of the 5 conditions (HRef, 1/48, 2/48, 2/32, 3/32 beats offset) compared to the explicit reference. The quality rating was done on a continuous scale ranging from low to high quality, the highest quality meaning that there was no perceived degradation. For each trial, participants could freely listen to the different stimuli, going back and forth between conditions as many times as needed.

In addition to the quality rating, participants were asked to classify the perceived degradation using 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 stated.

3.3 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 so as 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, a two-step familiarization protocol was proposed, in order for the participant to understand the test interface and task:

1. participants were presented with examples of timing degradation;
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, with 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 "synchro loss". Then, participants took two test trials using the same music excerpt, with the following time offset values: 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.

A panel of 15 people (2 females, 13 males, aged between 23 and 52) participated in the test: 8 sound engineers from Radio France, and 7 engineers from L-Acoustics' R&D department. All the testers reported normal auditory acuity and could be considered expert listeners. The test lasted between 25 and 60 min, with an average duration of 40 minutes. As no influence of musical education was observed in [4] and [3], this parameter was not taken into consideration here.

3.4 Results

A Kolmogorov-Smirnoff test (*kstest* function in Matlab) indicated that quality ratings were normally distributed for every stimulus. Therefore, parametric methods could be used. An analysis of variance (ANOVA) was performed on the ratings with the following factors: test participant (N = 15), track (N = 3), source location (N = 2), repetition (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.

Two main factors were found to have a significant effect on the quality rating: the time offset ($F(4,895) = 192.34$; $p < 0.001$), the track ($F(2,897) = 12.63$; $p < 0.001$), as well as the interaction between these factors ($F(8,891) = 5.98$; $p < 0.001$). However, note that the effect of the track and offset-track interaction are much smaller than that of the offset itself. This suggests that the relative timing between instruments plays a major role in the perception of musical groove, but that listeners' sensitivity to instrument timing may vary as a function of the piece or musical genre.

On the other hand, the ANOVA determined that instrument location had no significant effect on the quality rating, neither as the main factor ($F(1,898) = 0.88$; $p = 0.3635$) nor in interaction with the track ($F(2,897) = 1.85$; $p = 0.1579$) or time offset ($F(4,895) = 0.88$; $p = 0.4765$). This result suggests that the perception of the musical groove is mostly dictated by timing and mostly independent of space. Also, no effect of repetition was found ($F(1,898) = 0.75$; $p = 0.3996$), which indicates that memory had no influence on the perceived quality.

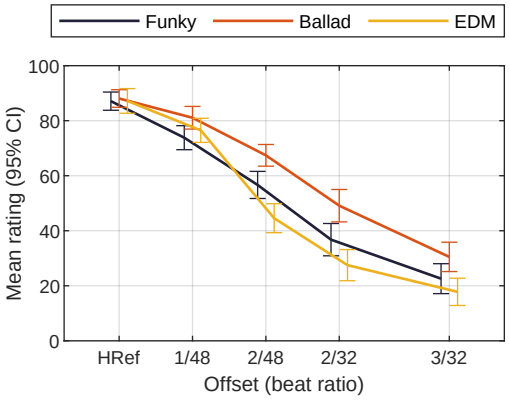

Fig. 6: Mean rating and 95% confidence interval as a function of the time offset for the 3 tracks.

Figure 6 shows the mean rating and 95% confidence intervals (CIs) as a function of the applied time offset for the three tracks. Ratings are the highest for the hidden reference, at around 90/100. Then, for every track, a continuous degradation of the perceived quality is observed as the time offset between objects increases. The low anchor obtained a poor rating (less than 40/100)

for all the tested tracks. Even though time-offsets are relative to the music tempo, the ballad appears to be less degraded than the other tracks for a given offset. However, the CIs corresponding to the different tracks overlap for the majority of the time offset values.

For every track, the ratings obtained with a time offset of 1/48 beat are significantly different from that obtained with the hidden reference. A t-test between the two gives the following results: $p < 0.001$ for Track 1 (*Funky*), $p = 0.0059$ for Track 2 (*Ballad*), and $p < 0.001$ for Track 3 (*EDM*). Thus, an offset of 1/48 beat is above the audibility threshold for tracks 1 and 3, while it can be considered close to the threshold for Track 2 (*Ballad*).

For time offsets larger than 1/48 beat, we observe that the perceived quality degrades faster for the *EDM* track than for the two others: the mean rating is below 50/100 with a time offset as small as 2/48 beats. This difference between the EDM track and the two others can also be observed by looking at the "Sync. Loss" and "Pushed/Laid-back" degradations reported by the participants, as illustrated in Figure 7. With a time offset of 2/48 beats, more than 50% of the participants perceived the instruments of the *EDM* track as being no longer synchronized. On the other hand, the *Ballad* track seems much more robust to time offsets than the two others: with an offset of 2/48 beats, almost no participant reported a loss of synchronization, and with an offset of 2/32 beats, a majority of the participant qualified the track as "Pushed/Laid-back".

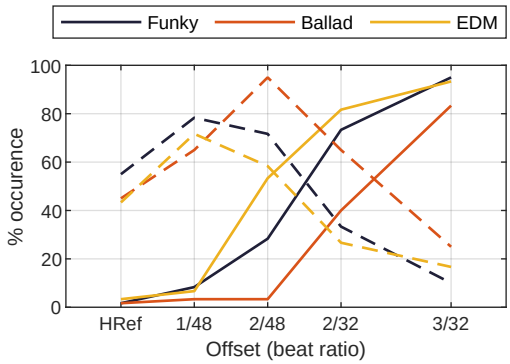

Fig. 7: Percentage of "Sync. Loss" (continuous line) and "Pushed / Laid-back" (dashed line) boxes checked as a function of time offset, for the three tracks.

Still referring to Figure 7, the percentage of perceived loss of synchronization appears to be correlated to the quality ratings. On the other hand, the percentage of "Pushed/Laid-back" occurrences follows a slightly different trend. For time offsets larger than or equal to 2/48 beats, the number of occurrences decreases symmetrically to the increase in the occurrences of "Sync. loss": in other words, as the time offset increases, participants tend to perceive time offsets as a problem and no longer as an acceptable change to the groove. Interestingly, about 50% of the participants checked the "Pushed/laid-back" box for the hidden reference, which could be a sign that they had difficulties identifying the reference from the stimuli corresponding to the smallest offset. In other words, the difference between the two was subtle, even for expert listeners.

3.5 Discussion

For all the tested audio stimuli, the introduction of microtiming between instruments induces a drop in the perceived quality of the groove. However, the quality is only perceived as a little degraded for time offsets shorter than 1/48 beats (10 to 15 ms depending on the track). Such small offsets were not perceived as causing a loss of synchronization but increased the feeling of a pushed or laid-back groove compared to the hidden reference. However, the notion of "pushed" or "laid-back" is not necessarily associated with a large drop in the perceived quality. These results are in line with the literature, which states that small microtimings (below 20 ms) are not perceived as a timing error. In [5], the groove quality was perceived as preserved with a bass shifted 2/32 beats earlier relative to the drums.

Concerning time offsets of 2/48 beats, results strongly depend on the track. For Track 2 (*Ballad*, offset duration 30 ms), this offset is overwhelmingly perceived as a change in the musical groove. For Track 1 (*Funky*, offset duration 20 ms) a significant amount of the participants perceived a loss in synchronization, but the majority mentioned a "Pushed/Laid-back" groove. Lastly, for Track 3 (*EDM*, offset duration 20 ms), the majority of the listeners identified the offset as a synchronization loss and rated the quality as low (less than 50/100). This large drop in perceived quality observed for the *EDM* track between 1/48 and 2/48 beats may be explained by the very "snapped on the grid" groove of the original track. As no microtiming is present in the original track, any perceived change in timing is seen as

an artifact and detrimental to the quality of the groove. Lastly, time offsets larger than or equal to 2/32 beats (30 to 67 ms) were for the most part perceived as causing a loss of synchronization between the instruments and the associated quality ratings were low. The test results show, otherwise, the absence of impact of the spatialization of sound on the perceived quality.

Note that this experiment left out several potentially important parameters, such as the complexity of audio tracks, the presence of reverberation, etc. Additionally, the tracks employed in this test were selected because they were expected to be relatively sensitive to instrument timing. Nevertheless, these results confirm that time offsets in the range of 30-40 ms can cause significant drops in the perceived music quality for certain styles or instrument combinations. In the following section, we discuss how these results can be taken into account when planning live music events in large venues.

4 Anticipating potential timing issues in the spatial mix

State-of-the-art immersive sound systems offer an important shared speaker coverage and provide great freedom and precision in audio object positioning. However, in the previous sections, we have established that spatializing objects using such systems could induce significant time offsets between sound objects, which in turn can modify the perceived musical groove. In other words, there may be a trade-off between the width spanned by the objects and the timing of these objects at certain listening positions. When preparing the mix for a live event, care must therefore be taken in order to optimize sound quality for any position in the audience.

Recalling the sound system modeled in Section 2, propagation time differences between objects that are panned among the three central speakers (Speakers 3 to 5 in Figure 2) are shorter than 10 ms, even for listening positions located far off-center in the audience. According to the results of the listening test presented in Section 3.4, offsets below 10 ms do not result in significant drops in perceived musical quality, therefore objects panned using the central speakers should be perceived with the intended groove. Hence it is advisable to keep the core rhythmic elements of the music, such as the drums and bass, at the center of the stage. Note that this is in line with common practices: in a rock music band the bass is often physically close to the drums on stage, for instance.

For instruments or sections that are less rhythmically critical, such as the arpeggio guitar from the *Ballad* song in the perceptual experiment, objects can be placed more freely. According to our simulation results, time differences over 15 ms are only observed for sources panned to the edge of the stage (Speakers 2 and 6). Moreover, these values are reached for off-center positions, which account for around 15% of the shared coverage area. Therefore, if a minor change in timing is deemed acceptable for a small part of the audience, some instruments can be freely panned along the entire stage width (Speakers 2 to 6).

Further, note that this study focuses on musical styles and instruments that make changes in timing relatively obvious. However, during the informal listening sessions that took place during this research work, we noticed that sounds with little transients, effects, and ambient elements were much more robust to time offsets. Hence, instruments and musical elements of this kind could safely be panned anywhere along the frontal system (Speakers 1 to 7).

Additional listening tests would be required to establish a complete set of mixing guidelines, which cover any musical style and possible sound source. In any case, the spatial mix in an immersive sound live event should always be assessed from different positions in the audience. Mixes are often prepared in studios with much smaller sound systems, which induce very short time offsets between sound objects, as illustrated in Figure 3. Nevertheless, note that it is possible to anticipate for possible timing issues even in the studio, with the help of simulation. The idea is to emulate the propagation delays and level differences expected at a given position in the audience. Such scale simulation tool is available in L-Acoustics' L-ISA controller [11] (see Figure 8).

5 Conclusion

The perception of the musical groove and, more generally, of the musical quality, is strongly linked to the rhythmic synchronization between instruments. As we have shown in this paper, spatializing objects using a sound system such as that deployed for large venues leads to propagation time differences between objects.

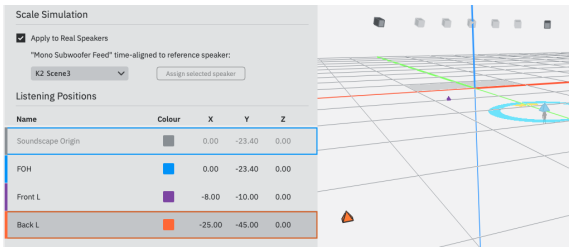

Fig. 8: Scale simulation panel in L-ISA Controller, outlining simulated monitoring positions.

These time offsets depend on the physical distance between loudspeakers and listening position, and occur regardless of the employed panning technique.

We conducted a perceptual experiment to determine to what extent the time offsets induced by immersive sound systems affected the perceived quality of music. The results of this test show that the perceived musical quality does not depend on spatialization but decreases as the length of time offsets increase. Additionally, the drop in quality for a given offset seems to depend on the musical style and/or instrument combination. With short offsets (10 to 15 ms), the perceived degradation is small for the three different musical excerpts under consideration. With longer offsets (20 to 30 ms), however, the perceived quality strongly depends on the audio stimulus.

Simple mixing rules can be deduced from the results of the listening test. First, it is generally safe to pan instruments to the three loudspeakers located at the center of the stage. Therefore, it is advisable to keep the most rhythmically critical instruments at these positions. Second, in the case of less critical instruments or musical styles, objects can be safely panned further away from the stage center, or even outside the stage area when the sound sources have few transients. Such considerations can be made early in preparation for a live event, as scale simulation makes it possible to synthesize the expected mix at any position in a venue and listen to it in a studio.

Note that the music excerpts used in our study were picked because the change in musical groove caused by time offsets between instruments is particularly obvious. Further investigations would be required to devise more exhaustive mixing guidelines that take into account the musical genre, tempo, and type of instrument.

Acknowledgment

The authors would like to thank Radio France and Hervé Déjardin for inviting some of their sound engineers to take part in the perceptual experiments.

References

[1] Moulin, S. and Corteel, E., "Spectral and spatial perceptions of comb-filtering for sound reinforcement applications." in *Audio Engineering Society Convention 152*, Audio Engineering Society, 2022.

[2] Keil, C., "Participatory discrepancies and the power of music," *Cultural Anthropology*, 2(3), pp. 275–283, 1987.

[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] Berkhout, A. J., de Vries, D., and Vogel, P., "Acoustic control by wave field synthesis," *The Journal of the Acoustical Society of America*, 93(5), pp. 2764–2778, 1993.

[8] "d&b Audiotechnik Soundscape website," https://www.dbsoundscape.com/global/en/, 2023.

[9] "Adamson Fletcher Machine website," https://adamson-fletcher-machine.com, 2023.

[10] Pulkki, V., "Virtual sound source positioning using vector base amplitude panning," *journal of the audio engineering society*, 45(6), pp. 456–466, 1997.

[11] "L-Acoustics L-ISA website," https://l-isa.l-acoustics.com/, 2023.