Audio Engineering Society

Convention Paper 10551

Presented at the 152nd Convention
2022 May, In-person and Online

Spectral and spatial perceptions of comb-filtering for sound reinforcement applications.

Samuel Moulin[1], Etienne Corteel[1]

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

Correspondence should be addressed to Samuel Moulin (samuel.moulin@l-acoustics.com)

ABSTRACT

Most sound reinforcement systems consist of multiple loudspeakers systems arranged strategically to cover the entire audience area. This study investigates the spectral and spatial perceptions of interferences that can be experienced in the shared coverage area between two full-range loudspeakers. A listening test was conducted to determine the effect of lag source delay, relative level, and angular separation, on the perception of spectral coloration and spatial impressions (width, localization shift, image separation). The results show that spectral coloration is considerably reduced when sources are spatially separated, even with a small azimuth angle (10°). It was also found that coloration audibility depends on the interaction between the audio track and the delay introduced. Finally, the type of perceived spatial degradation depends mainly on the spatial separation and on the relative level of the source arriving later in time (lag source).

1 Introduction

1.1 Motivations

When two or more audio signals are presented with a short delay it creates an interference pattern also known as a comb-filtering pattern. This can typically happen when the same audio signal is played over multiple loudspeakers close to each other, or when a reflection combines with the direct sound or a source. Comb-filtering perception can be critical for sound reinforcement systems that often consist of multiple loudspeakers arranged strategically to provide consistent sound pressure level and tonal balance over the entire audience area. The main loudspeaker system is meant to reproduce most of the audible frequency range across most of the audience area. Supplementary systems called fill systems are usually deployed as spatial complements to extend the main system coverage area. Different types of fill systems exist depending on the target audience area relative to the stage. For instance, front-fills are usually installed on the stage to cover the first rows of the audience, whereas out-fill systems are oriented towards lateral parts of the audience (see Figure 1).

Although the main and fill systems are meant to cover separated audience areas, there exists an overlap area where both systems have comparable sound level and arrive slightly delayed in time. In this overlap area, the comb-filtering pattern caused by the systems combination can alter the perceived sound quality.

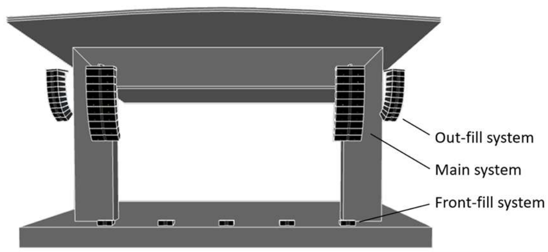

Figure 1. Illustration of typical main and fill loudspeaker systems deployed on a stage.

1.2 Comb-filtering perception

Perception of comb-filtering mainly depends on the delay between the lead source (first-arriving sound) and the lag source (delayed sound) and can alter spectral, spatial, or temporal aspects of the original sound to be reproduced. Most research on the subject has been conducted in the context of room acoustics to study the effect of reflections on sound perception in concert halls or domestic environments.

Overview of perceptual effects

When the delay between two sound sources is small (less than about 1 ms), listeners fuse the location of the two sounds into a single sound (phantom image) that is perceived between the sources. Within this time window, as the delay between the source increases, the perceived localization of the fused image gradually shifts toward the direction of the first-arriving sound [1, 2]. This effect called summing localization is commonly used to place virtual sound sources between two loudspeakers in a stereo configuration.

When the delay between the two sounds is increased, listeners continue hearing a perceptually fused image coming from the direction of the first sound. This phenomenon is the precedence effect, also called Haas effect [3, 4]. The parameters influencing the size of the precedence effect window have been extensively studied and include: the sound level difference between the two sounds, the type and temporality of sound stimuli, the reverberation, and the direction of arrival of sounds [5, 6, 7]. As an example, the size of the precedence effect window for two sounds at the same level is approximately 5-10 ms for clicks and 30 ms for speech signal [1, 2].

If the delay between the two sounds is greater than the upper limit of the precedence effect window, the fusion phenomenon disappears. This leads to the perception of sounds coming from two directions at the same time, or to the perception of a distinct echo (arriving later in time) for even larger delays.

One should note that even when the direction of arrival of the sounds are perceptually fused into one location, the delayed sound impacts the perceptual qualities of the first sound in terms of spatial impressions (increase in spaciousness, width), loudness, or spectral aspects (tonal coloration) [8].

Perception of spectral coloration

Some studies investigated the impact of individual reflections on the perception of the overall timbre in small rooms [9, 10]. These studies indicated that the first-order floor and ceiling reflections are more likely to individually contribute to the overall timbre of the reproduced sound than other reflections. This suggests that the audibility of spectral coloration depends on the direction of arrival of the reflection. Barron also found that coloration is stronger for small differences in direction of arrival between the sound sources [8].

Most studies found in the literature [6-10] focus on determining threshold of detection of spectral coloration rather than identifying conditions leading to subjectively unacceptable coloration. A recent study though investigated the effect of lateralization angle, level, and delay of a single reflection on both audibility and acceptability of tonal coloration [11]. Authors identified the horizontal spatial area for which a -6 dB reflection creates audible or unacceptable tonal coloration when combined with a lead sound placed at 0 °. The unacceptable coloration area was defined as ± 41 ° in the front which was slightly narrower than the region of audible coloration (± 50 °) as illustrated on Figure 2. Although a free elicitation test showed that the delay modified the tonal coloration characteristics, it was unclear whether delay had an influence on the size of the audible and unacceptable coloration horizontal areas.

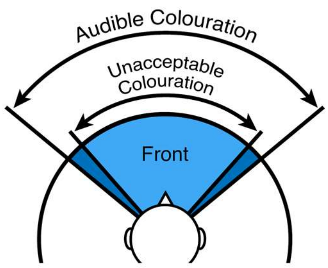

Figure 2. Illustration of audible and unacceptable coloration areas resulting from the combination of a lead source at 0° and a lag source at -6 dB (from [11]).

Limitations of current literature
Literature [1-11] show that comb-filtering perception is influenced by many factors including: the delay, relative level and spatial separation between the lead and lag sources, but also the type of audio material and environment reverberation. The effect of these factors is usually considered individually although combining them in a single experiment would help understanding the relative effect of these factors on interferences perception. In addition, the audibility of comb-filtering has been extensively studied in the context of room acoustics, but the conclusions are not directly applicable to sound reinforcement applications. Indeed, in such application, the spatial separation between the sources is usually limited to 30 ° or so, the level of the sources is comparable (< 6 dB differences), and the sound reproduction is made in a variety of environments including reverberant spaces.
Finally, most available references on spectral coloration due to comb-filtering [6-10] focus on detection thresholds without considering the intensity of the perceived coloration or its acceptability.

1.3 Goals of the study

The present study has the following objectives:
- Investigate perception of comb-filtering created by the combination of two full-range sources using test conditions representative of sound reinforcement applications.
- Identify conditions that are the most critical in terms of spectral coloration and qualify corresponding modification of spatial image.

In this study, two perceptual experiments were conducted. In a preliminary experiment, the threshold of acceptability of spectral coloration is investigated using an adjustment method (see Section 2). The main experiment is then performed focusing on the audibility of spectral coloration associated with perceived spatial modifications (see Section 3).

2 Preliminary experiment

The goal of this preliminary study was to identify the key factors playing a role in the acceptability of spectral coloration, and to identify the relevant range of test conditions to integrate in the main experiment.

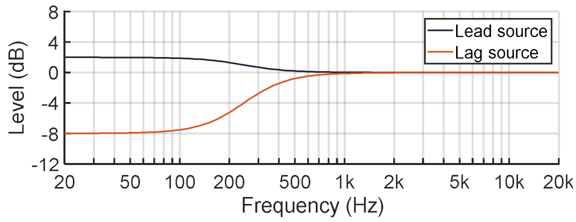

Figure 3. Frequency response of the LF contour modifications applied to the lead and lag sources.

2.1 Test conditions

The test conditions were selected with sound reinforcement applications in mind. An analysis of some sound reinforcement reference designs has been conducted to identify typical range of delay and level differences in the overlap area of two full-range sources, as well as the spatial separation between systems seen from the audience perspective. Focus has been made on the combination of the main loudspeaker system with the out-fill systems or delay systems. Indeed, such configurations represent a large proportion of the audience in comparison to other types of fill systems such as front-fills for instance. Although the overall magnitude response is comparable in the medium and high frequencies between the main and fill systems, the latter usually have limited capabilities in the low frequencies (i.e. LF contour). This difference of frequency content might reduce the audibility of comb-filtering pattern in the LF for the audience located in the overlap area.

Based on the analysis of target applications and informal listening sessions, the following test conditions were selected. The **lead source** was always reproduced on a loudspeaker in front of the participant (0 °). The **lag source** was reproduced with a spatial separation of 10 ° or 30 ° relative to the lead source, and with a delay of 1, 2, 5, or 10 ms. In addition, the LF contour of the lead and lag sources were either similar (same LF capabilities) or modified (up to 10 dB LF contour differential to simulate typical main-fills loudspeaker systems, see Figure 3).

2.2 Audio materials and reproduction system

Two ecologically valid audio materials were tested. A piece of orchestral music mainly involving strings (ORCH.) was taken from the European Broadcasting

Union PEQS CD (track 19) and was chosen because of its rich harmonic content. The second audio material (VOICE) is a dry recording of a male singer (Ben Folds - The Frown song). It was chosen because listeners are sensitive to the timbre of speech which is a familiar stimulus. In addition, this type of signal is delivered in most sound reinforcement applications. Audio files were about 30 seconds long and were presented at a sampling frequency of 48 kHz.

A multichannel artificial reverberation was created and reproduced over surround loudspeakers to provide the sensation of being in a larger and more immersive listening environment. Only the diffuse part of the reverb was created with no intention of adding artificial early reflections. The spectral and temporal characteristics of the reverberation tail were designed based on on-site measurements performed in medium-size venues such as auditoriums and conference rooms (sound reinforcement systems were used as exciting source). As a result, 8 decorrelated reverb signals ($RT_{30} \approx 1.4$ s) were associated with the lead source, and 8 others associated with the lag source.
The dry mono signals were reproduced on frontal loudspeakers placed at 0° for the lead source, and 10° or 30° to the right for the lag source. The reverberation signals were reproduced over 8 surround loudspeakers placed at ± 20°, 55°, 90°, and 145° (see Figure 4). All the loudspeakers were time-aligned and level-aligned at the listening position.

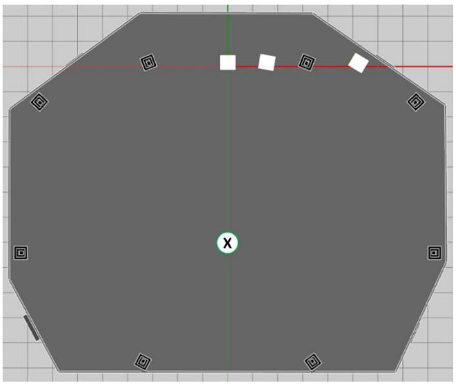

Figure 4. Top view of loudspeaker setup and listening position (white squares represent loudspeakers fed by dry signals, blacks squares are reverb loudspeakers).

The listening environment was a relatively large pre-production studio at L-Acoustics R&D facilities (about 200 m^2, $RT_{30} \approx 0.4$ s), equipped with Syva and Syva Low loudspeakers from L-Acoustics. The loudspeakers were connected to L-Acoustics LA4X amplifiers that were fed by AVB audio signals sent from an RME Digiface AVB audio interface connected to a laptop.

2.3 Experimental method

Adjustment method
An adjustment method was used to determine threshold of acceptability of spectral coloration. Such method is known to be fast and intuitive although slightly less accurate than other psychophysical methods for threshold identification such as method of constant stimuli or staircase methods [12]. Adjustment methods are usually well suited for preliminary experiments.

Participants were asked to define the threshold of acceptability of spectral coloration by adjusting the gain of the lag source. They were not explicitly told that they were adjusting the gain of the lag source. The adjustment was made using a vertical slider corresponding to a continuous scale going from -15 dB (bottom) to 0 dB (top). As shown on Figure 5, no labels or numeric values were presented on the slider to limit potential biases. The output level was normalized in real-time so that no obvious change in sound level could be heard when switching between the test signals, nor when adjusting the slider.

Acceptability judgement
Participants had to make acceptability judgments based on the comparison with a reference signal which corresponded to the lead source alone (i.e. no interferences). Participants were also instructed not to account for potential shift in spatial localization while performing the task. The assumption that subjects can discriminate between a spatial and a spectral change in the reproduced sound is supported by previous work involving adjustment method [13]. Prior to the test, participants were exposed to sound examples to illustrate the type of degradations to focus on (spectral coloration against shift of spatial localization).

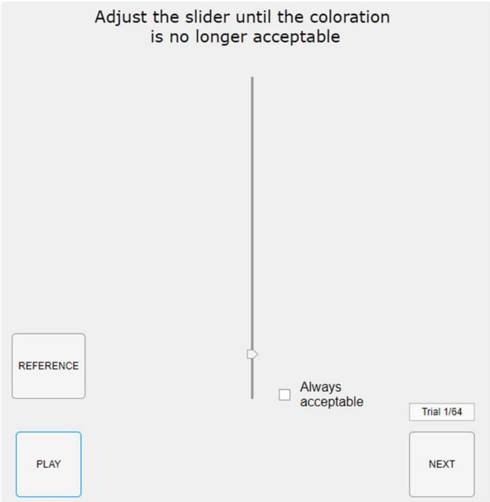

Figure 5. User interface of the preliminary experiment (adjustment method).

Participants could select "Always acceptable" checkbox if the spectral coloration was perceived as acceptable when the slider was at the top position (i.e. gain of 0 dB relative to the lead source).

Procedure

Participants performed the task for each of the 32 test conditions, corresponding to combination of all lag source parameters (4 delays, 2 spatial separations, 2 frequency contour conditions) with the 2 audio materials. A full repetition was included resulting in a total of 64 trials.

2.4 Data collection and analysis

The panel consisted of 8 experienced listeners who all have a background in live audio (former or current system engineers, L-Acoustics R&D staff).

Collected perceptual data were analyzed using Matlab 2020b. One participant was excluded from data analysis as 58 trials out of 64 were judged as "Always acceptable". A univariate analysis of the variance (ANOVA) was performed considering a mixed model, where *Participant* and *Repetition* were treated as random factors, and other factors were treated as fixed factors (*Delay*, *Spatial separation*, *Contour*, *Track*). Main factor effects and first-order interactions were analyzed.

Delay was found to have a significant effect ($F(3, 374) = 18.53$; $p<0.05$), so as the following interactions:

- *Delay* x *Track* ($F(3, 374) = 4.78$; $p<0.01$),
- *Delay* x *Spatial sep.* ($F(3, 374) = 3.58$; $p<0.05$),
- *Track* x *Spatial sep.* ($F(1, 374) = 9.01$; $p<0.01$).

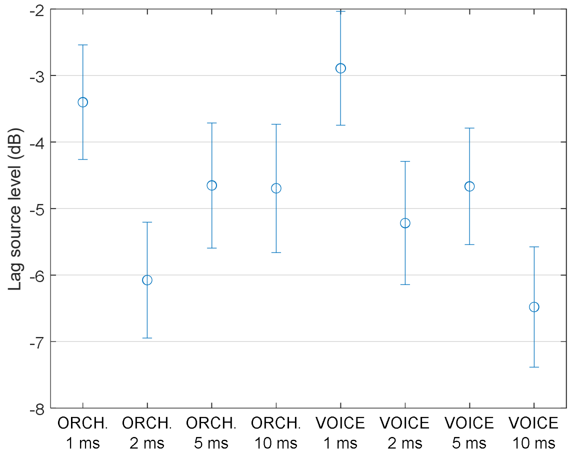

Figure 6. Mean gain and 95% confidence intervals for *Delay* x *Track* interaction.

To illustrate the *Delay* x *Track* interaction, Figure 6 shows that the coloration is perceived as unacceptable at lower gain for VOICE when the lag source is delayed by 10 ms, whereas the most critical condition is found at 2 ms for ORCH.

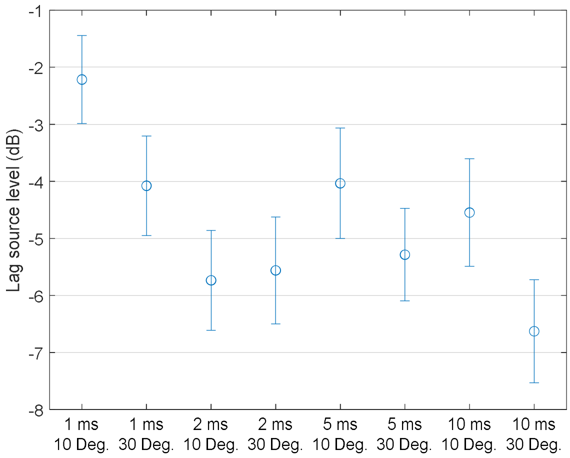

Figure 7. Mean gain and 95% confidence intervals for *Delay* x *Spatial separation* interaction.

The level of lag source corresponding to perception of unacceptable spectral coloration also depends on combinations of delay and spatial separation (see Figure 7). As an example, for the combination 10 ms / 30 °, the coloration is judged as being unacceptable for a lag source level of -6.7 dB (average value) relative to the lead source, whereas it can be as high as -2.2 dB to reach similar impression for the combination 1 ms / 10 °. Surprisingly, differences between conditions at 10 ° and 30 ° are either judged as similarly critical or more critical for 30 ° (e.g. with delays of 1 ms and 10 ms). Visual inspection of *Track* x *Spatial sep.* interaction (not shown here to limit the paper length) also revealed that the increase in spatial separation between the sources lowers the acceptable levels for the lag source, especially for the ORCH. material. Such observation is not consistent with most results in the literature [8, 11], as the binaural decoloration (sometimes referred to as echo suppression [14]) should limit audibility of coloration for larger spatial separation between sources.

2.5 Learning from preliminary experiment

Although the data analysis of this first experiment revealed potentially interesting results (e.g. interactions between *Track*, *Delay* and *Spatial separation*, no effect of LF contour difference), some inconsistencies with the literature were also spotted as discussed in Section 2.4. Feedbacks from participants indicated that the task was difficult to perform, especially because both spectral and spatial modifications were perceived when the slider was adjusted. This could be an indication that, instead of judging acceptability of spectral coloration in comparison to the reference signal, participants rather assessed any perceived differences with the reference no matter the spectral or spatial nature of the degradation. This would potentially explain that conditions at 30 ° were judged as more critical than at 10 ° because the perceived spatial modifications were more noticeable. In addition, the notion of acceptability was also highly subjective and related to personal expectations, which was difficult to handle for some participants. For all these reasons, it was decided to adapt the test method for the main experiment (see Section 3.3).

3 Main experiment

3.1 Experimental setup

The experiment was conducted in the same environment as the preliminary study except that the loudspeakers delivering artificial reverb were not used (see Section 2.2). Indeed, it was decided to run the experiment in an echoic environment with its natural early and late reflections. Informal listening indicated that adding reverb partially masks some of the effects that we are interested in (see Section 5.1 for further discussions on the effect of reverberation).

3.2 Test conditions

As no effect of LF contour conditions was found in the preliminary experiment, this factor was removed from the main experiment. It was also decided to increase the number of tested delays to get a better understanding of the role of this factor and its potential interactions with audio materials, and spatial separation (see Section 2.4). In addition, a spatial separation of 0 ° between the lag and lead source was included in the test. Indeed, this case of electric summation is potentially the most critical in terms of spectral coloration, without introducing additional spatial cues from the lag source. This condition could be seen as a low-quality anchor for spectral coloration and help identifying the most critical conditions (most critical delay for each audio material for instance).

To sum up, the **lag source** was presented with the following parameters (values are given relative to the **lead source**, placed at 0 °):

- a spatial separation of 0, 10 or 30 °,
- a delay of 0.5, 1, 2, 3, 5, or 10 ms,
- a level of -6, -3, 0, or +3 dB.

3.3 Adaptation of experimental method

To prevent participants from making a combined assessment of their spectral and spatial impressions, both perceptual aspects had to be evaluated in parallel (see Figure 8). The audibility of **spectral differences** was rated on a continuous scale with “Not audible” and “Very audible” as endpoints. This approach was preferred to the estimation of acceptability threshold as it is less related to participant’s expectation and therefore less hedonic by nature.

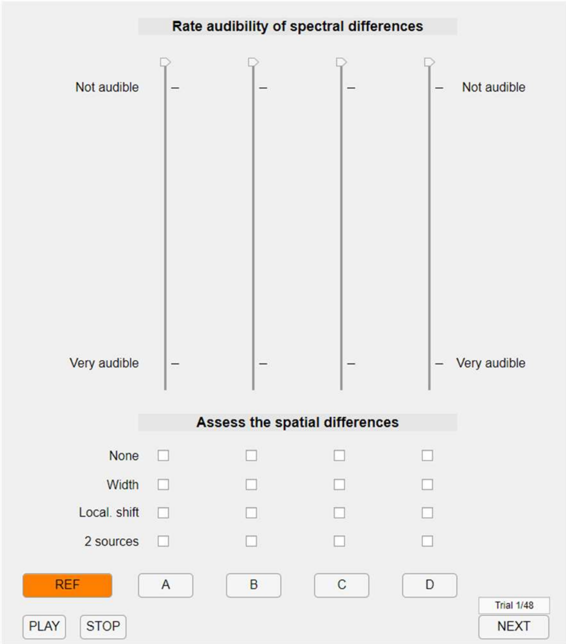

Figure 8. User interface of the main experiment.

In addition to spectral aspects, the assessment of **spatial differences** was made using checkboxes. Participants could give indications about the nature of the perceived spatial difference if any. Participants could choose one or more perceptual attribute based on the following descriptions:

- *Width*: a difference is heard in terms of image width, source width, localization blur, etc.,
- *Localization shift*: the spatial origin of the sound is shifted,
- *2 sources* (spatial separation): the sound is perceived as coming from two sources in space.

Participants were instructed to base their assessments on the comparison to the reference signal, not on preference.

3.4 Experimental design and procedure

The test consists of 48 pages like the one shown on Figure 8. The 48 pages correspond to 24 combinations of delay (N=6) and level (N=4), for each audio material (N=2), and presented in a randomized order.

On each page, the participant had to assess spectral and spatial differences between the reference signal (REF) and the 4 items (A, B, C, and D on Figure 8) that included, in a random order:

- a hidden reference (i.e. single source at 0 °)
- lead source at 0 ° + lag at 0 °
- lead source at 0 ° + lag at 10 °
- lead source at 0 ° + lag at 30 °

A familiarization step was performed prior to the test where some sound examples were presented to the participants as a complement to the written instructions. The idea was to illustrate key perceptual attributes (*Spectral coloration*, *Width*, *Localization shift*, *Spatial separation*) and provide clarifications in case of misunderstanding of test instructions. Then, participants could perform the task for 8 pages to get familiar with the task and the UI before starting the experiment.

4 Results

A panel of 9 expert listeners participated in this test. Most of them were sound engineers at Radio France while the others were members of L-Acoustics R&D staff. Out of the 9 participants, only two of them took part in the preliminary experiment which was performed a few months before. All the assessors self-reported normal hearing conditions and participated in the experiment on a voluntary basis.
It took 110 minutes on average to perform the entire test including the familiarization phase, and short breaks every 20 minutes.

4.1 Audibility of spectral coloration

The scores for audibility of spectral differences between test signals and the reference signal were rated between 0 and 100 (0 = very audible, 100 = not audible). The term "Spectral fidelity" is mentioned on the figures to help understanding the direction of the scale. Scores were analyzed using a mixed ANOVA model, with *Participant* (N=9) treated as random factors, while other factors were treated as fixed factors: *Level* (N=4), *Delay* (N=6), *Item* (N=4), *Track* (N=2). Main factor effects and first-order interactions were analyzed. Due to the number of parameters under test, only the most relevant effects are presented in this paper.

Effect of *Item*

The ANOVA model revealed a significant effect of the factor *Item* on the audibility of spectral coloration (F(3, 1724) = 84.57; p<0.001).

Figure 9 shows that the hidden reference (Ref.) was well detected by the participants, and that the electric summation (0 °) was assessed as the most different in in terms of spectral coloration in comparison to the reference signal. Results shows that increasing spatial separation between sources reduces the audibility of spectral coloration although the difference is subtle between 10 ° and 30 °. A complementary analysis revealed that the difference observed between 10 ° and 30 ° is mainly driven by the VOICE audio track.

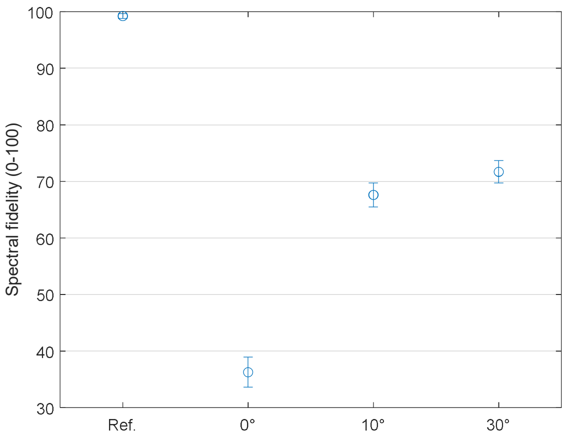

Figure 9. Mean score of audibility of spectral difference and 95% confidence intervals for *Item* (mean score averaged over *Participant*, *Level*, *Delay*, and *Track* factors).

Effect of *Level*

As expected, a significant effect of *Level* was found (F(3, 1724) = 28.68; p<0.001) and the model also revealed a significant effect of *Level* x *Item* interaction (F(9, 1718) = 9.62; p<0.001).

Results presented on Figure 10 suggest that spectral coloration is more audible when sources have the same level (0 dB) although the difference with -3 dB and + 3 dB conditions is tight, even at 0 °. A level of -3 dB or +3 dB relative to the lead source seems to create comparable perception of spectral coloration (full overlap of confidence intervals).

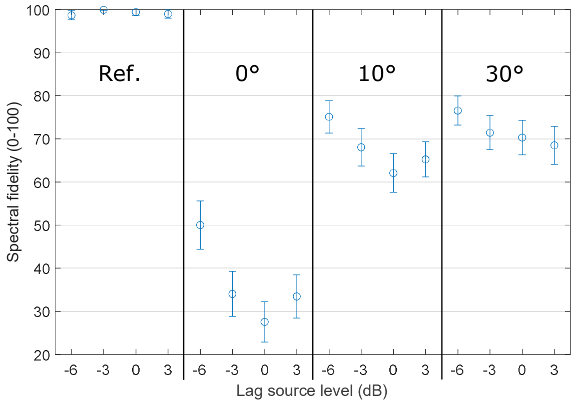

Figure 10. Mean score of audibility of spectral difference and 95% confidence intervals for *Level* x *Item* interaction.

Effect of *Delay* x *Track* interaction

A significant effect of *Delay* was found (F(5, 1722) = 4.31; p<0.001), but a stronger effect was found for the *Delay* x *Track* interaction (F(5, 1722) = 42.44; p<0.001) as illustrated on Figure 11. The smallest delays are more critical for ORCH. whereas for VOICE, the most critical delay is 10 ms. This suggests that there is not a single delay for which the perceived coloration is more audible as it depends on the audio tracks. Although not shown here, this interaction is very strong at 0 °: mean scores for ORCH. are between 17 and 74 for delays of 0.5 and 10 ms respectively (VOICE mean score between 38-30 for delays of 0.5-10 ms resp.).

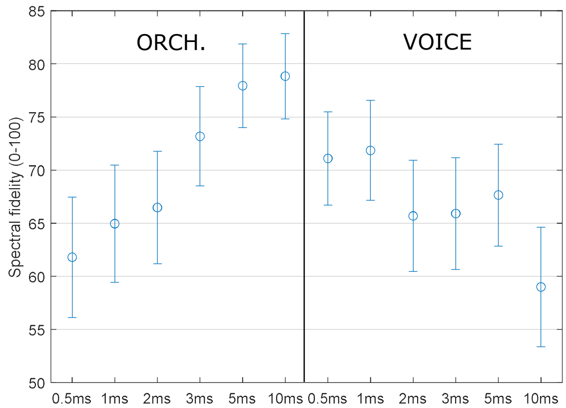

Figure 11. Mean score of audibility of spectral difference and 95% confidence intervals for *Delay* x *Track* interaction.

A more detailed analysis of the audio materials spectrum could potentially reveal such dependance by looking at the frequency of the notches of the comb-filtering pattern created by the tested delays.

4.2 Detection of spatial modifications

In addition to assessing the spectral coloration audibility, participants evaluated the spatial modifications caused by the combination of the sound sources. Participants had to indicate if they detected a spatial difference or not ("No spatial diff.") and to specify the nature of the perceived changes by selecting one or more descriptors: "Width", "Localization shift", "Spatial separation" (see Section 3.3 for definitions). The collected data on spatial perception were analyzed based on the number of selections and are here presented as percentages.

Effect of *Item*

Results obtained for each tested items are illustrated on Figure 12. This figure shows that the hidden reference (i.e. single source at 0 °) was well detected because no spatial difference was perceived in almost 100 % of the cases. When adding a lag source at 0 °, this score remains high (92%) confirming that this condition can be seen as a low-quality anchor for spectral coloration without introducing additional spatial cues from the lag source. When the spatial separation between the lead source (0 °) and the lag source increased to 10 ° or 30 °, spatial modifications were more noticeable. Most modifications were perceived in terms of Width at 10 ° and were more balanced between Width, Localization shift and Spatial separation at 30 °.

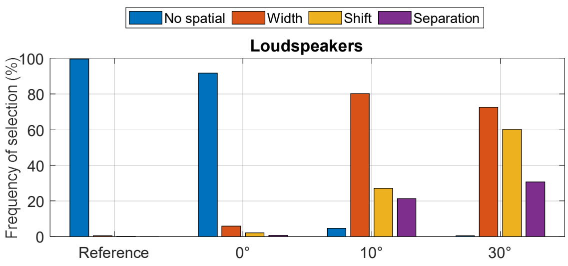

Figure 12. Spatial modifications detected for each item depending on the reproduction method.

Effect of *Level* and *Delay*

In addition to the spatial separation between the sources, other factors seemed to influence the spatial perception of interferences including the level and the delay of the lag source. As an illustration, the results presented on Figure 13 correspond to the case of the lag source at 10 ° and 30 ° (Reference and 0° conditions were filtered out from the data).

It can be seen on the top part of Figure 13 that the level of the lag source relative to the lead source have a strong influence on spatial perception. Cochran's Q tests were performed for each spatial impression individually. These tests revealed the existence of strongly significant differences between the tested levels for No spatial diff., Width, Shift, and Separation ($p<0.001$).

At -6 dB, the presence of lag source primarily modifies the width of the perceived image. As the level of the lag source increases, the modifications of the spatial image are more and more perceived as a shift in localization or even as a separation in space. A post-hoc Dunn test was performed with Bonferroni adjustments to identify which pairs of conditions were significantly different. The post-hoc test confirmed that all pairs were different in terms of Shift and Separation except the pair "0 dB vs. +3 dB" for Separation (see Table 1).

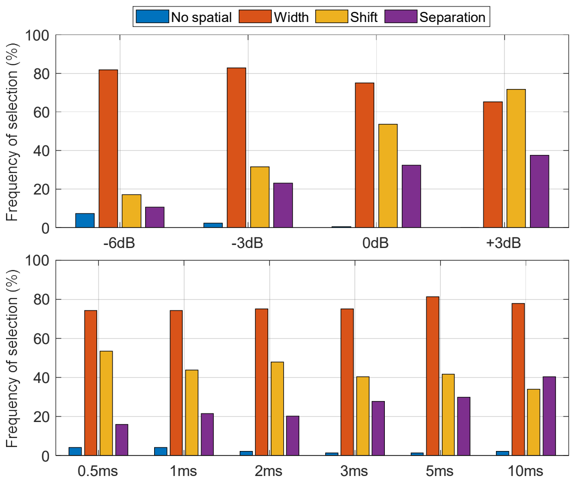

Figure 13. Spatial modifications over loudspeakers for each level (top) and each delay tested (bottom).

	Pair	No spatial diff.	Width	Shift	Separation
Level	-6dB / -3dB	***	-	***	***
	-6dB / 0dB	***	-	***	***
	-6dB / +3dB	***	***	***	***
	-3dB / 0dB	-	-	***	**
	-3dB / +3dB	-	***	***	***
	0dB / +3dB	-	**	***	-
Delay	0.5ms / 1ms	-	-	-	-
	0.5ms / 2ms	-	-	-	-
	0.5ms / 3ms	-	-	-	-
	0.5ms / 5ms	-	-	-	*
	0.5ms / 10ms	-	-	**	***
	1ms / 2ms	-	-	-	-
	1ms / 3ms	-	-	-	-
	1ms / 5ms	-	-	-	-
	1ms / 10ms	-	-	-	***
	2ms / 3ms	-	-	-	-
	2ms / 5ms	-	-	-	-
	2ms / 10ms	-	-	-	***
	3ms / 5ms	-	-	-	-
	3ms / 10ms	-	-	-	-
	5ms / 10ms	-	-	-	-

Sig. levels: $p < 0.001$ *** $p < 0.01$ ** $p < 0.05$ * $p > 0.05$ not sig.

Table 1: Post-hoc test and significance levels for each pair of the tested levels and delays.

The effect of delay of the lag source on spatial perception is more subtle (see Figure 13, bottom). It seems that increasing the delay slightly reduces the frequency of selection of localization shift (the only significant difference was found for the pair "0.5 ms vs. 10 ms"), while perception of separated sources in space occurs more and more frequently. Cochran's Q tests did not reveal differences between the frequency of selection of No spatial diff. or Width depending on the delay ($p > 0.05$) as shown in Table 1.

5 Discussions

5.1 Effect of reverberation

The potential effect of reverberation on the audibility of spectral coloration has not been investigated here. Indeed, each experiment was conducted under a single reverberation condition: an echoic environment with an artificial reverb of about 1.4 s in the preliminary experiment, and a natural reverberation time of about 0.4 s in the main experiment. These reverberation times are representative of typical environments for sound reproduction. Indeed, a RT of 0.4 s can be met in a pre-production space with acoustically treated walls, whereas a RT of 1.4 s would be more representative a medium-size concert venue [15]. However, modifications of reporting methods between these experiments make the comparison of the results obtained from the two protocols irrelevant. The primary goal of the first experiment was indeed to identify the most relevant factors and test conditions to integrate into the main experiment.

It would be interesting though to assess the evolution of spectral coloration perception under different reverberation conditions, in terms of reverberation length or direct-to-reverberant ratio for instance. We believe these aspects should be addressed in a dedicated experiment. In addition, it is anticipated that reverberation characteristics play a role on other perceptual dimensions altered by comb-filtering such as temporal aspects for large delay values [8].

5.2 Consequences on sound reinforcement applications

These experimental results can have multiple implications on sound reinforcement applications such as the alignment strategy of main and fills loudspeaker systems or on panning choices between adjacent sources of an immersive sound system.

For the main-fills loudspeakers alignment strategy, it seems that there is no universal "coloration-free" delay range as the audibility of spectral coloration depends both on the delay and the audio material itself. However, considering the spatial separation in azimuth usually observed between typical main and fills loudspeaker systems, audibility of spectral coloration should naturally be limited in the shared coverage area. In addition, the strong dependance to the lag source level supports the fact that directivity of full-range loudspeaker systems should allow for a sharp transition at the limit of coverage. Indeed, this would increase level differences in the overlap area, reducing audibility of spectral coloration.

Regarding immersive sound reinforcement systems, and especially amplitude-based panning approaches, object-oriented mixing techniques can be adapted to prevent the perception of spectral coloration. For instance, although the spatial separation between systems naturally limits coloration audibility, it could be further reduced by panning audio objects closer from one source rather than in dead-center between adjacent loudspeakers. Again, this would favor level difference over the audience area and diminish risk of detecting spectral degradations.

6 Conclusions

This paper presented a series of listening tests on perception of interferences created by the combination of two full-range sources.
In a preliminary experiment, the relevant factors and conditions inspired from sound reinforcement applications were identified. It also helped refining the method for data collection. Indeed, some results suggested that the initially proposed adjustment task was ambiguous, potentially because both spectral and spatial modifications were perceived when the slider was adjusted. Based on this, the main experiment was designed investigating the effect of lag source delay, relative level, angular separation, and audio materials on perception of interferences. The influence of these factors on audibility of spectral coloration were assessed, together with spatial impressions in terms of width, localization shift, and image separation.
The results showed that audibility of spectral coloration due to interferences depends on the spatial separation between sources. Indeed, the most critical case was observed when the lead and lag sources were reproduced at 0 °, whereas the audibility of coloration was considerably reduced with a separation as small as 10 °. As expected, increasing the level difference between the sources reduced the audibility of spectral artifacts. In addition, the delay at which the coloration was more audible strongly depended on the audio material. When interfering sources were separated by 10 ° or 30 °, the perceived spectral coloration was associated with modifications of the spatial impressions. The type of perceived spatial modification was found to depend mainly on the relative level of the lag source, with limited influence of the delay between the two full-range sources.

Acknowledgments

Authors would like to thank Radio France and especially Hervé Desjardin and Frédéric Changenet for their help in facilitating this experiment recruiting expert listeners.

References

[1] Blauert, J. (1997). Spatial hearing: the psychophysics of human sound localization. MIT press.

[2] Moore, B. C. (1995). Hearing. Academic Press.

[3] Wallach, H., Newman, E. B., & Rosenzweig, M. R. (1949). A precedence effect in sound localization. The Journal of the Acoustical Society of America, 21(4), 468-468.

[4] Haas, H. (1972). The influence of a single echo on the audibility of speech. Journal of the Audio Engineering Society, 20(2), 146-159.

[5] Brown, A. D., Stecker, G. C., & Tollin, D. J. (2015). The precedence effect in sound localization. Journal of the Association for Research in Otolaryngology, 16(1), 1-28.

[6] Toole, F. E. (2008). Sound Reproduction: Loudspeakers and Rooms. Taylor & Francis.

[7] Olive, S. E., & Toole, F. E. (1989). The detection of reflections in typical rooms. Journal of the Audio Engineering Society, 37(7/8), 539-553.

[8] Barron, M. (1971). The subjective effects of first reflections in concert halls—the need for lateral reflections. Journal of sound and vibration, 15(4), 475-494.

[9] Bech, S. (1995). Timbral aspects of reproduced sound in small rooms. I. The Journal of the Acoustical Society of America, 97(3), 1717-1726.

[10] Bech, S. (1996). Timbral aspects of reproduced sound in small rooms. II. The Journal of the Acoustical Society of America, 99(6), 3539-3549.

[11] Johnson, D. (2018). Towards the perceptual optimisation of virtual room acoustics (Doctoral dissertation, University of Huddersfield).

[12] Ehrenstein, W. H., & Ehrenstein, A. (1999). Psychophysical methods. In Modern techniques in neuroscience research (pp. 1211-1241). Springer, Berlin, Heidelberg.

[13] Bech, S. R. (1998). Spatial aspects of reproduced sound in small rooms. The Journal of the Acoustical Society of America, 103(1), 434-445.

[14] Zurek, P. M. (1979). Measurements of binaural echo suppression. The Journal of the Acoustical Society of America, 66(6), 1750-1757.

[15] Adelman-Larsen, N. W. (2014). Rock and Pop Venues: Acoustic and Architectural Design.