Audio Engineering Society

Conference Paper 8

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

On the perception of time-alignment between full-range speakers and subwoofers for sound reinforcement

Thomas Mouterde, Samuel Moulin, Nicolas Epain, and Etienne Corteel

L-Acoustics, 13 rue Levacher Cintrat, 91460 Marcoussis, France

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

ABSTRACT

Most sound reinforcement systems employ a combination of full-range speakers and subwoofers to deliver a consistent sound pressure level over the audience, while maximizing the frequency bandwidth. A time alignment between the main (full-range) and sub (subwoofers) systems is generally required to ensure an efficient summation at low frequencies. This study investigates how time misalignments between the main and sub systems affect the perceived sound quality. We conducted a listening test whereby the listeners were asked to rate the sound quality as a function of the relative delay between main and sub systems. In addition, test participants were requested to qualify the nature of the perceived artifacts using spectral or temporal attributes. Our results suggest that the overall perceived quality does not decrease linearly with increasing delays, and that it reflects the presence of both spectral and temporal degradations. Lastly, temporal degradations are perceived more often when the sub system is delayed with respect to the main system, unlike spectral degradations for which the direction of the delay has very little influence.

1 Introduction

Sound reinforcement systems employ multiple loudspeakers arranged strategically to provide consistent Sound Pressure Level (SPL) and tonal balance over the entire audience area. The *main* loudspeaker system consists of one or more full-range sources and is meant to reproduce most of the audible frequency range across most of the audience area. Often, a *sub* system, consisting of subwoofers, is deployed to extend the frequency bandwidth of the main system towards lower frequencies. The two systems must be configured so that they are aligned in time and combine efficiently with each other in the range of the crossover frequency.

In the case of a large audience area, the distance between the main and sub systems can be such that the relative timing of the emitted waves varies significantly as a function of the listening position. Therefore, once the two systems are calibrated and aligned in time for a given location, there can still be severe misalignments elsewhere in the audience.

The objective of this study is to qualify how such misalignments between the main and sub systems are perceived. More specifically, our aim is to understand how the perceived quality changes as a function of the time delay and to understand the associated perceptual dimensions. The paper first describes the problem of aligning a full-range loudspeaker system with subwoofers, both objectively and perceptually. The design of a perceptual experiment, aiming to qualify the perception of main-sub time alignment, is then presented. Lastly, the test results are presented and discussed.

2 Main-sub alignment

In this section we provide an overview of the issue of main-sub time alignment and define our research questions.

2.1 Main and sub systems

As its name suggests, the main system aims to provide most of the SPL and reproduce most of the frequency response (*e.g.,* 60 Hz to 20 kHz) over most of the audience. The most common layout for the main system is a stereo layout, with stacks of speakers arranged on the left and right sides of the stage. However, immersive sound systems, which typically consist of five or more speaker stacks spanning the performance area, are used more and more often.

The sub system, on the other hand, aims to complement the main system by extending its bandwidth towards the extreme-low frequency range. The crossover between the main and sub frequency responses usually lies within the 63 Hz octave band (typically between 50 and 80 Hz, depending on the system combination) as shown in Fig. 1.

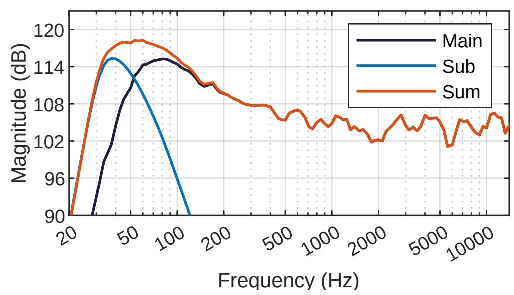


Fig. 1: Measured frequency response of a main system (L-Acoustics K2), sub system (L-Acoustics KS28), and the sum of the two.

Sub systems consist of one or more subwoofers or stacks of subwoofers that can be arranged in various ways. They can be stacked on the ground or flown, centered, or positioned at the left and right sides of the stage. Another type of setup, referred to as *arc sub*, consists of using multiple subwoofers stacked on the ground and distributed along the stage width. Delays are then implemented between elements to control the directivity of the array. Each sub system layout has its pros and cons with regard to system efficiency and SPL distribution [1, 2], effect of the audience [3], noise pollution [4] or impact on audience exposure and auditory health [5]. Ultimately, a sub system layout should be chosen accounting for its capacity to “work” together with the main system.

2.2 Time alignment

The low-frequency (LF) response of the overall (main plus sub) system depends on how well the main and sub systems complement each other, especially around the crossover frequency, where the two equally contribute to the listener’s experience. The crossover between the main and sub systems is thus rigorously designed by loudspeaker manufacturers such that the summation is optimal. However, this design is done for enclosures that are very close to each other, while the main and sub speakers may be separated both horizontally and vertically depending on the selected layout.

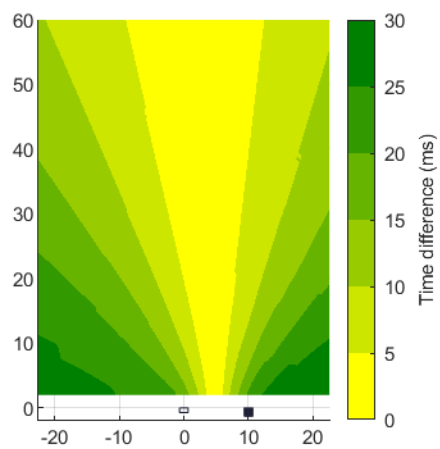


Fig. 2: Typical time differences observed over a flat audience area (45 x 60 m, 1.6 m height) between a house-left main system (12 K2 flown at 10 m from the central axis) and a flown central subwoofer system (8 KS28).

In practice, the main and sub systems must be time-aligned to optimize the frequency response of the overall sound system. This alignment is performed as part of the system calibration, before the show. First, the system's response is measured at one or more locations in the audience area. Then a delay (to be applied to the main or sub system) is determined so as to optimize the system summation. The delay compensates for the difference between travel times from the main speakers to the measurement location, on the one hand, and from the sub speakers to the measurement location, on the other hand. However, this time alignment is "perfect" only at a given location. At other locations in the audience, the two systems are not perfectly aligned and can even cancel each other at the crossover frequency in some cases. As the propagation time differences vary across the audience, time alignment becomes even more critical in large-scale deployments. Such a large-scale scenario is illustrated in Fig. 2, where propagation time differences between a central subwoofer and a House-Left main speaker are represented over a 45 x 60 m audience area. Propagation time differences between the two systems can reach 20 to 30 ms on the side of the audience. Similar values can be observed with other sub layouts, such as arc sub configurations.

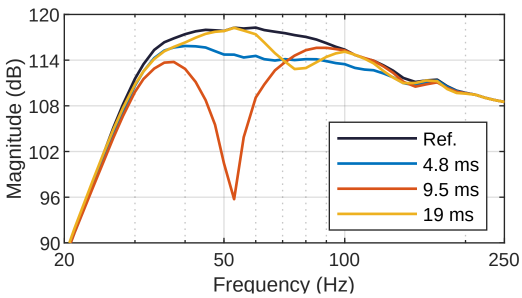


Fig. 3: Magnitude response of the summation of a main-sub system (L-Acoustics K2 and KS28 loudspeakers) as a function of the delay relative to perfect time alignment, measured at the alignment position.

Figure 3 shows the frequency response resulting from the summation of a main and a sub system, as a function of the time delay relative to perfect alignment. When the two systems are aligned (Ref), the summation is maximum at the crossover frequency because the two systems are in phase. When a delay is added between the systems, the gain inevitably decreases around the crossover area and can even collapse if the systems are out of phase, *i.e.* with a delay corresponding to half the period of the crossover frequency. In the case illustrated in Fig. 3, the crossover frequency is 52 Hz, which corresponds to a period of 19 ms. Therefore the systems are out of phase for a delay of 9.5 ms. Still looking at Fig. 3, a large delay may seem better than a small one in terms of magnitude response integrity. Indeed, the magnitude of the frequency response decreases less around the crossover frequency with a delay of 19 ms than with a delay of 9.5 or even 4.8 ms. However, large delays might introduce audible temporal artifacts, as will be shown in the following.

2.3 Perceptual aspects

The perception of main-sub misalignments is a largely unexplored research topic. At the time of writing, there seems to be no data regarding audibility thresholds, nor about the perceived audio quality in the context of main-sub misalignments. However, some research has been conducted on related topics.

Several publications have addressed the issue of group delay perception. Group delay can be seen as a measurement of a system's latency as a function of frequency. In [6], loudspeakers with different group delay characteristics were compared in a listening test. Results show that, in the range 300 Hz – 1kHz, group delays below 1 ms are rarely detected but above 2 ms they are almost always detected. This study also suggests that group delay values can exceed 10 ms below 200 Hz without being detected. These results are in line with the common belief that group delay below 1 or 1.5 periods cannot be detected.

Studies addressing the perception of group delay often rely on stimuli that may not be representative of musical signals, such as clicks or impulses (*e.g.*, [7]). In [6], the authors suggest that the frequency content of the excitation signal may impact the perception. This aspect is to be considered when studying main-sub alignment since the frequency response of sound reinforcement systems is usually not flat, as shown in Fig. 1.

Other studies investigated subjects closely related to ours. In [8], the authors conducted a listening test whereby an arc sub system was compared to a reference left-right subwoofer system. However, this study addressed the summation between subwoofers and not between the main and sub speakers. In [9], the perception of comb filtering in the shared coverage area of two full-range systems was investigated. The application case and frequency range of the present study are, nonetheless, different.

Despite the lack of references available in the literature, we can anticipate two types of perceivable degradations when the main and sub systems are misaligned:

- spectral degradations, related to a decrease in sound level around the crossover frequency, and
- temporal degradations, associated with a loss of temporal precision, or even the perception of two different sources separated in time in the case of sufficiently large delays.

This study addresses the following research questions:

- What is the impact of main-sub misalignments on the perceived audio quality for typical sound reinforcement applications?
- How does the perceived quality change as a function of the delay between the two systems?

3 Experimental design

This section describes the perceptual experiment designed to evaluate the perception of the time misalignment between main and sub systems.

3.1 Experimental setup

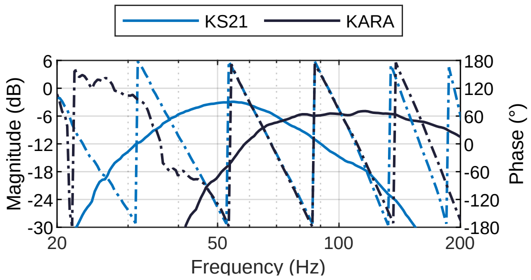
The experimental setup is illustrated in Fig. 4. The listening test took place outdoors, with the speakers stacked on the ground. Speakers are stacked on the ground to avoid the notch that would be created by the reflection on the floor at ear height with flown subwoofers. Two combinations of main/sub systems from L-Acoustics were selected to represent the variety of equipment used in medium or large-scale sound reinforcement applications. The smaller system consisted of four KARA II full-range loudspeakers (main) combined with two KS21 subwoofers (KS21_100 preset). This main-sub combination has a crossover frequency of 73 Hz, as shown in Fig. 5a. The larger system consisted of four K2 full-range loudspeakers (main) combined with three KS28 subwoofers (KS28_60 preset). For this system, the crossover frequency is 52 Hz, as shown in Fig. 5b. The loudspeakers (full-range and subwoofers) were connected to L-Acoustics LA12X amplifiers that were fed by AVB audio signals sent from an RME Digiface AVB audio interface connected to a laptop.

The main and subwoofer systems were time-aligned based on acoustic measurements performed at the listening position, which was located 15 m from the speakers. The measurements were done using the L-Acoustics M1 measurement software. The alignment parameters were selected so as to maximize the cross-correlation between the main and sub system impulse responses. As shown in Fig. 5, it also corresponds to a very good match of the main and sub phase traces. Therefore the frequency response magnitude resulting from the summation of the two systems is maximum. We refer to these alignments as reference alignments in the following sections.

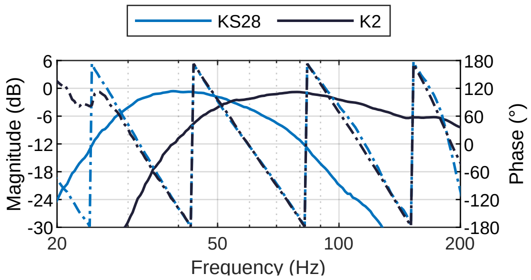
Fig. 4: Overview of the experimental setup and loudspeaker systems, from left to right: KS21 (sub) with KARA II (main), and K2 (main) with KS28 (sub).

3.2 Test conditions

For each main-sub combination, different misalignment conditions were tested around the reference alignment by adding delays, either to the main or sub system. Since the two main-sub combinations have different crossover frequencies, applying the same delay in milliseconds results in different summation patterns in the crossover frequency range, due to phase relations. Consequently, instead of introducing delays in ms, we introduced delays expressed as a ratio of the period (T) at the crossover frequency. This allows us to compare the two main-sub system combinations for a given delay value. For instance, a misalignment of $T/2$ should induce a full cancellation at the crossover frequency for both main-sub combinations although it corresponds to a delay of 6.9 ms for the KARA/KS21 combination and 9.6 ms for the K2/KS28 combination. Table 1 shows the tested delay values, expressed both in periods and



(a) System KARA/KS21



(b) System K2/KS28

Fig. 5: Measurements of the crossover between main and sub systems: magnitude (continuous line, left axis), and phase (dashed-dotted line, right axis).

ms. As shown in Fig. 2, delay values up to $3T/2$ correspond to values that may occur in real application cases.

In this experiment, we also tested the influence of the order of arrival on the perception of LF summation quality. Due to masking effects [10], we can expect the perception to differ, depending on whether the sub system is early or late with respect to the main system. In the following, the condition labeled "delay on main" refers to the case where the sub is early, and vice versa.

The test thus consisted of four pages like the one shown in Fig. 6. The four pages corresponded to the number of main-sub combinations ($N = 2$) multiplied by the number of orders or arrivals ($N = 2$). For each page, stimuli corresponding to different delay values were compared with a stimulus corresponding to the reference alignment. The four pages were presented in a random order for each test participant.

3.3 Audio materials

Audio materials were selected based on the presence of low frequencies with sharp transients (*e.g.*, precise kick drum, bass guitar, etc.) that could reveal both spectral and temporal degradation induced by system misalignments. In addition, the material was chosen so that degradations could be perceived equally well using both main-sub combinations. For example, tracks with a sustained bass note matching the crossover frequency of one of the main-sub systems were disregarded. Ultimately, a single track was selected: a 20 s excerpt from *Leviticus*, by Me'Shell Ndegeocello. This track was judged as being very critical, as it made it easier to perceive temporal and spectral degradations caused by misalignments. The intention here was not to offer a large variety in terms of music genre, spectral characteristics, etc., but rather to identify the perceptual dimensions and key factors involved in the perceived audio quality.

System		KARA/KS21	K2/KS28
Crossover ($1/T$)		72.5 Hz	52.2 Hz
Delay	$T/6$	2.3 ms	3.2 ms
	$T/4$	3.5 ms	4.8 ms
	$T/2$	6.9 ms	9.6 ms
	$3T/4$	10.4 ms	14.3 ms
	T	13.8 ms	19.1 ms
	$3T/2$	20.7 ms	28.7 ms
	$2T$	27.6 ms	38.2 ms

Table 1: Crossover frequencies and tested delays (in ratio of the period T at the crossover frequency and milliseconds) for the two main-sub combinations.

3.4 Test methodology

The test protocol was inspired by the MUltiple Stimuli with Hidden Reference and Anchor (MUSHRA) method. Following the ITU recommendation, an explicit reference was presented to the participant. As explained in Section 3.1, this reference corresponded to the main and sub systems being time-aligned at the listening position, with no additional delay. The participants were asked to compare eight stimuli to this reference: seven stimuli corresponding to the delay values presented in Table 1, as well as a hidden reference stimulus referred to as HRef.

The test conditions included two perceptual anchors that matched the two main perceptual dimensions under investigation. The $T/2$ condition served as a spectral anchor, as it causes maximum cancellation at the

Fig. 6: Graphical user interface of the application used for the experiment.

crossover frequency (*i.e.*, maximum decrease in SPL). In addition, a temporal anchor (T.Anc) was introduced, which corresponded to a delay of $2T$ applied to the main system (*i.e.*, the sub system arrives earlier than the main system). A preliminary listening session showed that a similar delay applied to the sub system was not perceived as badly.

For each of the eight time-alignment conditions (HRef, $T/6$, $T/4$, $T/2$, $3T/4$, T, $3T/2$, and T.Anc), participants were asked to rate the LF summation quality compared to the reference alignment. They could freely listen to any of the nine stimuli, in the order they wished. Quality ratings were done using the quality scale shown in Fig. 6. Participants were instructed to rate quality in comparison to the reference stimulus, and not on their preference. In addition to the quality score, participants were requested to select one or more types of degradations among the following:

- Level loss: loss of SPL in the low frequencies,
- Precision loss: loss of temporal precision (*e.g.*, tightness or attack of kick drum, bass),
- 2 sources: perception of two distinct sources in time,
- Other: any perceived degradation not described by the other categories.

The labeling of the different stimuli under test (from 1 to 8) was picked at random for every trial.

3.5 Test procedure

The experiment was conducted with only one participant at a time. The sound pressure level was set by the organizer so as to be loud enough to hear changes induced by misalignments at low frequencies but remain comfortable after a one-hour test session (≈ 81 dBA). The experiment started with the reading of the test instructions. A familiarization step was then performed prior to the test so that the participant understood the test interface and the task before starting the experiment. The familiarization consisted of two steps. First, examples illustrating the different perceivable degradations were presented: level loss at $T/2$, temporal precision loss at T, and source separation at $2T$. Then, the testers could experiment with the test interface during a pre-test. This pre-test was a single page where four stimuli could be compared to the reference (time offsets of $T/4$, $T/2$, T, and $2T$). After this familiarization, the tester was invited to confirm that everything was clear and that he/she consented to participate before starting the test.

4 Perceptual test results

A panel of 23 expert listeners aged between 25 and 55 participated in the test. Most of them were members of L-Acoustics R&D staff with or without past professional experience as mixing or system engineers in a live sound context. All the assessors self-reported normal hearing conditions and participated in the experiment voluntarily. On average, it took the participants 45 minutes to go through the test, including the familiarization phase.

4.1 Analysis of variance

A Kolmogorov-Smirnoff test (*kstest* in Matlab) was done on quality ratings. The test results indicated that quality ratings were normally distributed for all the test conditions. Therefore, parametric data analysis methods could be used. An analysis of variance (ANOVA) was performed on the ratings with the following factors: test participant ($N = 23$), main-sub combination ($N = 2$), delay direction ($N = 2$), and delay ($N = 8$). 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 results indicate that all the examined independent factors have a significant effect on quality. However, the effect size for the delay ($F(7,728) = 133.93$, $p < 0.001$) is much greater than that of the main-sub combination ($F(1,734) = 11.02$, $p = 0.003$), delay direction ($F(1,734) = 12.82$, $p = 0.002$), and interactions. Therefore, the delay added to the reference alignment plays the main role in the perception of the alignment quality. Consequently, in the following, the other influence parameters are inspected separately as a function of delay. The influence of the test participant is not studied here.

4.2 Test results for the KARA/KS21 combination

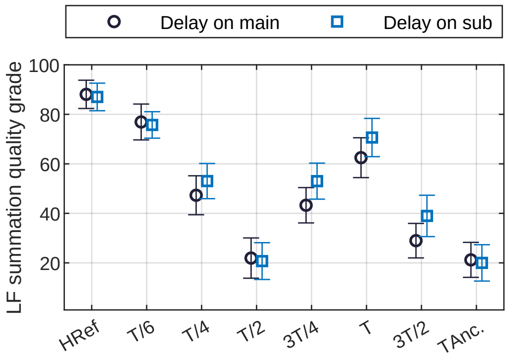


Fig. 7: Mean quality scores and associated 95% confidence intervals for the KARA/KS21 combination.

Fig. 7 shows the quality ratings obtained for the KARA/KS21 combination. We observe that the LF summation quality strongly depends on the delay but not in a linear way. In other words, larger delays do not necessarily mean lower LF summation quality. The hidden reference (HRef) was well detected by participants and obtained the best quality ratings, with 95% confidence intervals comprised between 80 and 100 (which corresponds to excellent quality, as shown in Figure 6). As the delay increases, the perceived quality decreases and reaches a minimum for a delay of $T/2$, which corresponds to the spectral anchor (maximum cancellation at the crossover frequency). This was expected because, between 0 and $T/2$, increasing the delay simply increases the SPL loss in the crossover area. However, with a delay of $T/6$, the quality is still rated as good and the difference with the rating of the hidden reference is only slightly significant. As delays increase above $T/2$, quality scores increase until a local maximum is reached for a delay of T. For this delay, mean ratings are comprised between 60 and 80, which corresponds to a "good" quality. Then, for delays larger than T, the quality decreases again.

There is no clear indication that the order of arrival has an influence on the perceived LF summation quality based on quality scores only. Indeed, 95% confidence intervals overlap almost entirely except for $3T/4$, T, and $3T/2$ conditions, for which quality ratings are slightly higher when the delay is applied to the sub system.

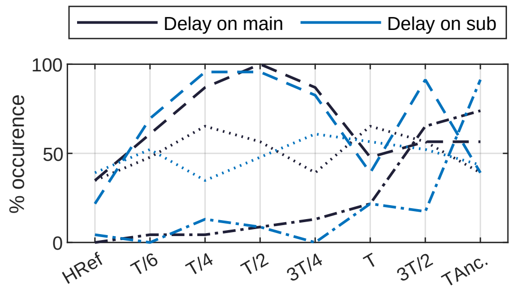


Fig. 8: Perceived degradations for the KARA/KS21 combination: level loss (dashed line), precision loss (dotted line), and source separation (dash-dotted line).

Fig. 8 shows the percentages of level loss, precision loss, and source separation reported by the listeners for the KARA/KS21 combination. The level loss is perceived by 100% of participants when a delay of $T/2$ is applied to the main system, and 96% when it is applied to the sub system. Results regarding a loss in temporal precision are difficult to interpret because percentages vary around 50% for every stimulus, with no clear trend. This corroborates the participants' feedback on the fact that this perceptual dimension was the most difficult to identify during the familiarization phase.

Source separation was perceived by most participants for the temporal anchor (T.Anc). In addition, the order of arrival seems to have an influence on the perception of source separation for the $3T/2$ delay: 65% of the participants perceived it when the delay was applied to the main system, against 17% when it was applied to the sub. In the latter case, the high percentage of perceived level loss (91%) might explain the low-quality ratings obtained with this condition. Note that the delay of $3T/2$ is the largest delay tested with the two different orders of arrival because, as described in Section 3.4, T.Anc corresponds to a delay of $2T$ applied to the main system.

4.3 Test results for the K2/KS28 combination

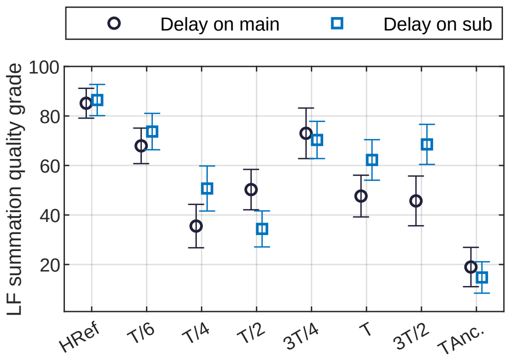


Fig. 9: Mean quality scores and associated 95% confidence intervals for the K2/KS28 combination.

Fig. 9 shows the quality ratings obtained for the K2/KS28 combination. The results are overall comparable to the ones obtained for the KARA/KS21 combination. The general trend is the same, with low-quality ratings around $T/2$ and T.Anc and better-quality ratings in between. However, a few differences can be observed.

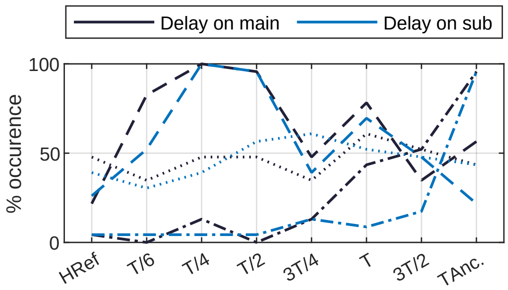


Fig. 10: Perceived degradations for the K2/KS28 combination: level loss (dashed line), precision loss (dotted line), and source separation (dash-dotted line).

First, the condition rated with the lowest quality (T.Anc excluded), depends on the order of arrival. Indeed, a delay of $T/4$ applied to the main is perceived as equally bad as a delay of $T/2$ applied to the sub. Fig. 10 shows that these conditions were associated with an equally high percentage of perceived level loss. This result is unexpected. Then, Fig. 9 shows that LF summation quality was good at $3T/4$ regardless of the order of arrival. As shown in Fig. 10, this condition is associated with relatively low percentages of perceived level loss in comparison to $T/2$ and T delays for instance, and low percentages of perceived source separation (13%). This might explain why quality ratings are good for this condition. Note that similar quality ratings were obtained with the KARA/KS21 combination but with a delay of T. Lastly, there seems to be an influence of the order of arrival on quality ratings obtained for delays of T and $3T/2$. This could be linked to the higher percentage of perceived source separation observed when delays are applied to the main system.

4.4 Source separation detection threshold

Participants perceived a source separation for large delay values, especially when the delay was applied to the main system (*i.e.*, when sounds emitted by the sub system arrive first). Source separation was systematically associated with medium to low-quality scores. Psychometric functions were fitted to the data presented in Figs. 8 and 10 to estimate a threshold for the perception of source separation. The fitting was done using the Psignifit Matlab toolbox, which implements the maximum-likelihood method described in [11]. Fig. 11 shows the lognormal psychometric functions corresponding to the two main-sub combinations.

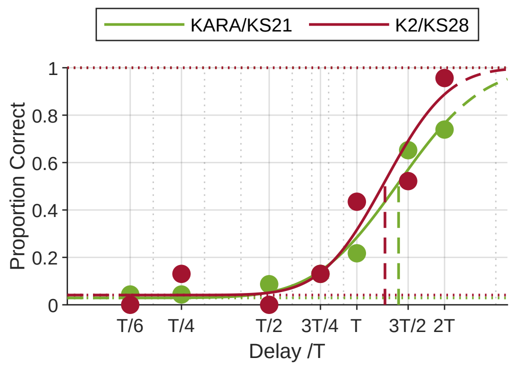


Fig. 11: Proportions of source separation perceived when delays are applied to the main system (K2 or KARA), and fitted psychometric functions.

The estimated source separation thresholds correspond to $1.25T$ (or 23.9 ms) for the K2/KS28 combination, and $1.39T$ (or 19.2 ms) for the KARA/KS21 combination, respectively. These thresholds correspond to 50% of the listeners perceiving a source separation.

4.5 Summary of results

Overall, adding a delay to the reference main-sub alignment lowers the perceived quality of the summation. However, the perceived quality is not a linear function of the delay but oscillates depending on the main-sub phase relationship. With a delay of $T/6$ the quality is rated as good. For longer delays, ratings decrease, reach a local minimum around $T/2$, then increase again, reach a local maximum around T, and then decrease again.

Two different types of degradations impact the quality perceived by listeners: spectral degradations and temporal degradations. Spectral degradations consist of a loss of energy around the crossover frequency. Degradations of this type can be perceived for relatively small delays, but their intensity depends on the main-sub phase relationship: it is maximal for delays of $T/2$ and $3T/2$ and minimal for a delay of T. Regarding temporal degradations, two effects were expected: a loss of temporal precision, and the perception of two distinct sources. However, our results indicate that the loss of precision is perceptually irrelevant. On the contrary, listeners seem to perceive two distinct sources for delays longer than T. This result is consistent with that observed in studies focusing on the perception of the group delay [6] but for lower frequencies here. Further, our results indicate that temporal and spectral degradations can result in an equally bad perceived quality. Regarding temporal degradations, an asymmetrical behavior can be observed: listeners perceive two sources mostly when the sub system is early relative to the main system. On the contrary, the intensity of spectral degradations does not seem to depend on the delay direction.

4.6 Discussion

Local maxima in perceived quality seem to occur for different delay values, depending on the main-sub combination (*i.e.*, T for Kara/KS21 versus $3T/4$ for K2/KS28). This may be explained by the fact that the two systems have a different crossover frequency, while the audio material used to generate the stimuli was the same. The impact of a loss of energy around the crossover frequency could thus depend on the spectral content of the signals. This could also explain why, for a delay of $3T/2$, a level loss was perceived more often with the KARA/KS21 combination than with the K2/KS28 combination. Confirming this hypothesis would require complementary perceptual tests with a greater variety of stimuli.

In the light of our results, we can predict how main-sub alignment may affect the perceived audio quality in the case of large-scale sound reinforcement systems. In the example proposed in Section 2, the main system is flown at 10 m of the central axis, while the sub system is flown on the central axis. With systems located 10 m from each other, the area where the perceived quality is rated as good would be small. In particular, the audio quality would be perceived as poor in front of the main system. In order to improve the main-sub time alignment over the largest possible portion of the audience, a simple system design recommendation is to place the two systems as close to each other as possible. Ideally, most of the audience should experience main-sub time offsets in the $-T/6$ to $T/6$ range relative to perfect alignment. This recommendation is consistent with that proposed in [4] regarding noise pollution: flying the subwoofers close to the main system allows for optimizing the alignment within the audience and achieving good rejection in the crossover frequency range outside of the audience.

However, several limitations may nuance the conclusions of this study. First, we used phase-matched loudspeaker combinations for the test. The two main-sub combinations also have group delays that match fairly well. Combinations involving other subwoofer configurations (cardioid configurations, for instance), or combinations involving other loudspeaker models must be studied because a discrepancy between group delays could affect the perception of temporal degradations. Second, the case where the main and sub systems have very different levels has not been considered in this study. In some cases, for instance, when ground-stack subwoofers are used in combination with flown main speakers, a very different SPL would be experienced at certain locations in the audience. A difference in levels would result in a shift of the crossover frequency and is thus expected to impact the perceived quality. Lastly, in the present study, only one sub and one main system were used. In practice, stereo main systems are very common, complemented by subwoofer deployments consisting of one or more sources (mono sub, left-right layout, arc sub, etc.).

5 Conclusion

Sound reinforcement loudspeaker systems usually consist of a full-range main system, which covers most of the audible frequency range, and a sub system, which extends the system's capacity at low frequencies. The main and sub systems are time aligned for a specific location and the alignment can be poor elsewhere in a wide audience area. In this paper, we have presented a perceptual experiment designed to investigate how time misalignments between the main and sub systems affect the perceived sound quality.

The results of the perceptual test demonstrate that the perceived quality is mainly driven by the perception of: 1) a loss of energy around the crossover frequency (spectral degradation); and 2) a temporal separation of sources (temporal degradation). Spectral degradations are perceived even for relatively small delays ($T/6$) and their perceived intensity oscillates as a function of the delay. Temporal degradations are perceived for longer delays (between T and $3T/2$) and are typically experienced when the sub system is early relative to the main. Both degradation types seem to impact equally the overall perceived quality and result in a non-linear relationship between delay and quality. In the light of these results, the following system design recommendation can be made: place the subwoofers as close as possible to the main speakers so that the two systems are well aligned over the largest possible audience area.

In future work, in addition to studying additional factors as outlined in Section 4.6, research should investigate the influence of the environment on alignment quality. In the present study, the perceptual experiment was performed in outdoor conditions, with ground-stacked systems. Test conditions were therefore close to free-field conditions, but the acoustics of a room could impact the main-sub alignment in a different manner. The presence of an audience could also affect the accuracy of the alignment if calibration measurements were done when the room was empty, as suggested in [3].

References

[1] Hill, A. J., "Live Sound Subwoofer System Performance Quantification," in *Audio Engineering Society Convention 144*, 2018.

[2] Corteel, E., Coste Dombre, H., Combet, C., Horyn, Y., and Montignies, F., "On the Efficiency of Flown vs. Ground Stacked Subwoofer Configurations," in *Audio Engineering Society Convention 145*, 2018.

[3] Mouterde, T., Corteel, E., and Melon, M., "Audience effect on the response of a loudspeaker system in the low frequency range, part 1: magnitude," in *Audio Engineering Society Convention 149*, 2020.

[4] Mouterde, T. and Corteel, E., "On the comparison of flown and ground-stacked subwoofer configurations regarding noise pollution," in *Audio Engineering Society Convention 151*, Audio Engineering Society, 2021.

[5] Hill, A. J., "Understanding And Managing Sound Exposure And Noise Pollution At Outdoor Events," in *AESTD1007.1.20-05.*, 2020.

[6] Liski, J., Mäkivirta, A., and Välimäki, V., "Audibility of Loudspeaker Group-Delay Characteristics," in *Audio Engineering Society Convention 144*, 2018.

[7] Flanagan, S., Moore, B. C., and Stone, M. A., "Discrimination of group delay in clicklike signals presented via headphones and loudspeakers," *Journal of the Audio Engineering Society*, 53(7/8), pp. 593–611, 2005.

[8] Shabalina, E., Ramuscak, J., and Vorländer, M., "A Study of Human Perception of Temporal and Spectral Distortion Caused by Subwoofer Arrays," in *Audio Engineering Society Convention 130*, 2011.

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

[10] Moore, B. C. J., "Masking in the Human Auditory System," in *Audio Engineering Society Conference: Collected Papers on Digital Audio Bit-Rate Reduction*, 1996.

[11] Wichmann, F. A. and Hill, N. J., "The psychometric function: I. Fitting, sampling, and goodness of fit," *Perception & psychophysics*, 63(8), pp. 1293–1313, 2001.