Audio Engineering Society

Convention Express Paper 464

Presented at the AES 160th Convention
2026 May 28–30, Copenhagen, Denmark

This Express Paper was selected on the basis of a submitted synopsis that has been peer-reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This Express Paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library (http://www.aes.org/e-lib), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Acoustic and Perceptual Consequences of Time Misalignments in Line Array Speakers

Nicolas Epain and Etienne Corteel

L-Acoustics, 13 rue Levacher-Cintrat, 91462 Marcoussis Cedex, France

Correspondence should be addressed to Nicolas Epain (nicolas.epain@l-acoustics.com)

ABSTRACT

Variable-curvature line arrays achieve their intended directivity and spectral balance through phase-coherent summation across cabinets. Even small timing disparities between elements perturb the interference patterns that shape the array response, with consequences for both spatial coverage and timbre. In this work we quantify these effects end-to-end. Using simulations for a typical 12-element array, we examine how inter-element delays modify the frequency response across an audience area. We then apply an auditory coloration model to predict the perceived impact of those modifications and validate the predictions through controlled listening tests. We observe that delays in the tens of microseconds generate pronounced spectral coloration that listeners consistently judge as degraded quality, whereas coloration becomes detectable at delays on the order of one microsecond. These results translate into synchronization accuracy targets for high-fidelity line-array deployments.

1 Introduction

The acoustic behavior of line arrays is governed by the superposition of wavefronts radiated by adjacent cabinets. In a variable-curvature configuration, relative phase alignment across elements controls the interferences that shape both spectral balance and directivity. When timing alignment is imperfect, *e.g.*, due to inter-amplifier desynchronization, those interference patterns shift. The resulting changes in frequency response can be subtle or severe depending on delay magnitude, frequency, and listener position, and they manifest perceptually as coloration of timbre. While such phenomena are understood qualitatively by practitioners, there remains a need to relate time misalignments in line arrays to human perception quantitatively.

Early discussions in the networking standards' community already anticipated the sensitivity of line-array summation to microsecond-scale timing. Christensen argued within the IEEE 802.1 standard context that, for array applications, PTP clock accuracy should be considerably tighter than 5 µs—and ideally below 1 µs—to preserve intended acoustic behavior [1]. More recently, Kaltheuner and colleagues revisited the problem and presented simulation results illustrating how small time

offsets between amplifiers translate into observable changes in array response [2]. Complementing these analyses, Hortig presented measurement evidence that audio network synchronization schemes differ in their resilience to stress [3]. Some approaches exhibited greater susceptibility to load-induced timing wander, which caused measurable deviations in the radiation pattern of a line array. Together, these contributions suggest that sub-microsecond-class synchronization is desirable in practice. Yet, they also highlight a persisting gap between network timing metrics, the resulting acoustic impact on arrays, and what listeners actually perceive in realistic use cases.

The present paper addresses that gap by focusing on the acoustic and perceptual consequences of inter-element latency. First, we simulate a representative variable-curvature array to characterize how time offsets reshape frequency response across the audience. Second, we map those spectral changes to perceived timbre using an auditory coloration model. Third, we validate the mapping with controlled listening tests designed to assess both overall quality degradation and the onset of audibility. Based on our results, we can derive synchronization accuracy thresholds that can inform sound reinforcement system design.

2 Acoustic simulation of inter-element delays in line arrays

Line sources rely on acoustic interference. This is especially true for line-source systems whose waveguides are designed to produce a coherent, isophase wavefront, making them highly sensitive to inter-element timing errors. As shown in Fig. 1, even small delays affecting one or more elements can modify the array's directivity pattern and frequency response significantly. In this section, we examine how such inter-element delays affect frequency responses in the audience.

2.1 Simulation methods

To investigate the acoustic consequences of timing errors between elements of a line source, numerical simulations were conducted using a realistic sound reinforcement configuration. The goal of these simulations was not to exhaustively characterize array behavior, but rather to establish representative acoustic effects and generate stimuli for subsequent perceptual analyses.

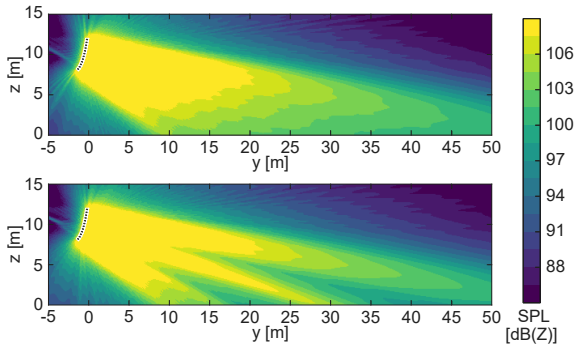


Fig. 1: Impact of inter-element delays on the sound field radiated by a line source of 12 speakers. Cross section of the average sound pressure level over the 1-10 kHz frequency band, when the array elements are perfectly synchronized (top); and random delays (1 ms RMS) are introduced for each group of three adjacent cabinets (bottom).

Line source configuration - The simulated system consisted of a vertical line source made up of 12 L-Acoustics K2 loudspeaker cabinets, arranged with a variable curvature optimized for audience coverage in a typical live sound scenario. Cabinet positions and splay angles were chosen to maximize sound pressure level and frequency-response consistency for a standing audience located between 5 m and 60 m from the array. The array was powered by four amplified controllers, each feeding a group of three adjacent cabinets, reflecting a common deployment practice in large-scale sound reinforcement. Within each group, cabinets were assumed to be perfectly synchronized, while relative delays could be introduced between groups to simulate inter-amplifier timing errors.

Acoustic modeling and observation points - For each cabinet, the acoustic transfer function to observation points in the audience was computed assuming free-field propagation. Audience positions were distributed along the array's main axis, at regular depth intervals of 25 cm, and at a listener height representative of a standing audience (1.7 m). A reference condition was first obtained by summing the contributions of all cabinets with perfect synchronization. Timing errors were then introduced by applying a relative delay to one of the four amplifier-fed groups, while keeping the remaining groups synchronized. Delay values ranged from a few microseconds to several hundred microsec-

onds, covering and exceeding the ranges observed in practical audio-network conditions.

Evaluation of spectral effects - For each delay condition and listener position, the resulting frequency response of the full array was compared to the reference response. The analysis focused on gain differences as a function of frequency and distance, highlighting how inter-element timing errors alter interference patterns within the array. These simulations allowed us to identify the frequency ranges affected by a given delay magnitude and to illustrate how increasing delays progressively impact lower frequencies. The resulting responses were subsequently used as input to perceptual modeling and listening-test stimulus generation, described in the following sections.

2.2 Simulation results

Figure 2 illustrates the gain difference relative to the perfectly synchronized reference condition when one of the four amplified-controller groups is delayed by 10 μs and 100 µs, respectively. For a delay of 10 µs, deviations from the reference response are observed mainly at high frequencies, with clear changes appearing above approximately 5 kHz. Above about 10 kHz, the response oscillates around the reference level across much of the audience area, and gain differences exceed 3 dB at multiple locations.

When the delay increases to 100 µs, the affected frequency range extends substantially toward lower frequencies. In that case, large gain differences are already visible from around 700 Hz, hence the corresponding coloration is no longer restricted to the upper end of the spectrum. At the same time, some frequencies remain comparatively unaffected. In particular, around 10 kHz and 20 kHz, the response is close to that of the synchronized case because a 100 µs delay corresponds approximately to one and two periods, respectively, at these frequencies. Under such conditions, the interference pattern is locally similar to that obtained in the absence of delay.

Overall, the simulations confirm that the acoustic effect of inter-element desynchronization is governed primarily by the delay magnitude. Small delays affect only the highest frequencies, whereas larger delays progressively alter a broader portion of the spectrum. In practical terms, this means that delays on the order

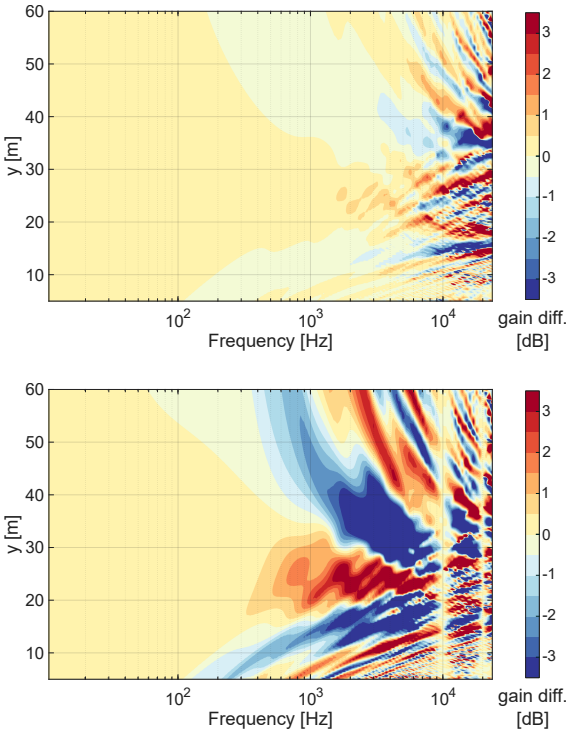


Fig. 2: Effect of inter-element time latency on the frequency response of a line source consisting of 12 L-Acoustics K2 speakers fed by four amplifiers. Gain difference relative to the reference frequency response (zero inter-element latency), as a function of the frequency and on-axis distance when one of the amplifiers is delayed by 10 μs (top) and 100 μs (bottom).

of a few microseconds should result in relatively localized high-frequency changes, while delays in the tens to hundreds of microseconds are expected to produce substantial spectral coloration over a much wider frequency range.

These results also clarify the mechanism by which timing errors degrade perceived sound quality. Because line sources rely on phase-coherent summation between adjacent elements, delaying one group changes the spatial interference pattern and therefore the spectral balance received by listeners. Although these simulations do not provide an exhaustive characterization of all possible array geometries, they illustrate the following trend: as inter-element delay increases, the resulting spectral distortion rapidly becomes large enough to be perceptually relevant.

3 Predicting perceived coloration

The simulations presented in the previous section showed that inter-element delays modify the frequency response radiated by the line source, which should mainly be perceived as a change in timbre. In this section, we estimate the magnitude of that perceptual effect using binaural renderings of the simulated array responses together with an auditory model of perceived spectral difference.

3.1 Binaural rendering methods

We used the same line-source configuration as in Section 2. For each audience position, we first computed the impulse response (IR) between each loudspeaker cabinet and a virtual omnidirectional microphone. These cabinet-to-listener IRs were then converted to binaural signals by convolution with anechoic head-related impulse responses (HRIRs) of a Neumann KU 100 manikin [4]. To do so, each IR was convolved with the HRIR corresponding to the direction of the associated cabinet. Lastly, the binaural contributions of all cabinets were summed to obtain one pair of HRIRs representative of the full line source at the considered listening position.

This procedure was applied both to the perfectly synchronized reference condition and to the degraded conditions in which a relative delay was introduced in one amplifier-fed group. Delay values ranged from 1 to 1000 µs, covering the range over which the simulations of Section 2 indicated potentially relevant acoustic effects. For each condition, the resulting binaural impulse responses were convolved with 0.5 s of Gaussian pink noise, yielding pairs of binaural stimuli with and without inter-element delay.

3.2 Auditory model

To estimate the perceptual effect of the simulated delays, we used the auditory model proposed by McKenzie et al. [5], as implemented in the Auditory Modeling Toolbox [6]. This model predicts the amount of spectral coloration perceived when comparing a test signal to a reference signal, and returns a score ranging from 0 to 100, where 0 denotes no perceived difference and 100 corresponds to the coloration produced by comparing a broadband signal to a version low-pass filtered at 3.5 kHz. In the following, we refer to this metric as the predicted perceived spectral difference (PPSD).

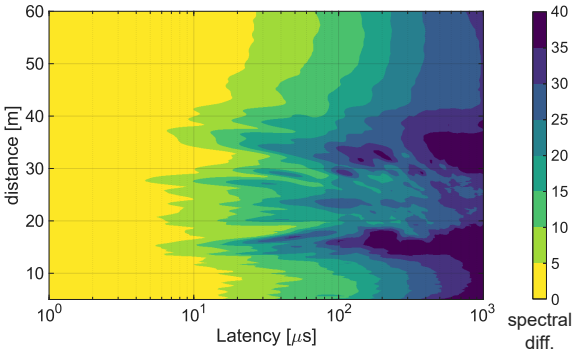


Fig. 3: Impact of inter-element time latency for a line source consisting of 12 L-Acoustics K2 speakers fed by four amplifiers, shown as the predicted perceived spectral difference [5], as a function of the latency of one of the amplifiers and on-axis distance.

For each audience position and delay condition, the model was applied to the two binaural pink-noise stimuli obtained in Section 3.1: the reference condition with perfectly synchronized cabinets, and the degraded condition including a delay on one amplifier-fed group. The resulting PPSD therefore provides a scalar estimate of the timbral difference induced by inter-element desynchronization at a given listening position.

3.3 Results

Figure 3 shows the predicted perceived spectral difference as a function of the delay introduced in one amplifier-fed group and the distance to the line source. As expected, the predicted coloration increased monotonically with delay magnitude. For delays below about 10 µs, the PPSD remained relatively low, indicating that only limited timbral differences should be expected. Above 10 µs, however, the predicted coloration increased sharply.

The results also varied with audience position. Although the whole audience region was affected to some degree, the highest PPSD values occurred at distances of approximately 16 m and 34 m. These distances correspond to positions where the coverage of the delayed amplifier-fed group overlaps most strongly with that of the adjacent groups, so that changes in interference patterns have a larger spectral effect.

These simulations therefore suggest that the timbral consequence of inter-element latency is determined mainly by the delay value, with audience position acting as a secondary factor. In particular, delays on the order of a few microseconds should produce only limited coloration, whereas delays on the order of 100 µs are expected to induce major spectral changes across much of the audience.

4 Listening test A: Perceived quality as a function of the predicted coloration

In this section, we investigate how the spectral coloration caused by inter-element delays in a line source maps to perceived audio quality. More specifically, the aim of this first listening test was to assess how audio quality ratings vary as a function of the PPSD.

4.1 Stimulus generation

The test stimuli were generated using the same line-source configuration as in Section 2 and 3. Four IRs were first computed, corresponding to sound propagation from the four groups of three K2 cabinets to one virtual omnidirectional microphone located in the audience, at a distance of 31 m from the array and a height of 1.7 m. Because the stimuli were intended for headphone reproduction, these impulse responses were equalized to compensate for the low-frequency contour of the loudspeaker system.

Three source signals were used: four bars of *Lola Rastaquouère* by Serge Gainsbourg, four bars of *Around the World* by Daft Punk, and 10 s of Gaussian pink noise. Each of these signals were filtered using the four IRs to obtain four group signals. Reference stimuli were obtained by directly summing these four signals. To generate degraded stimuli simulating synchronization errors, delays were applied to the four group signals prior to summation. Two delay modes were considered: a static mode, in which delays remained constant throughout the excerpt, and a dynamic mode, in which delays oscillated at 1/20 Hz to simulate clock wandering.

For each track and delay mode, five degraded stimuli were generated with increasing target coloration. This was achieved by selecting delays at random for each group of cabinets and repeating the process until the resulting stimulus matched a target PPSD value within a tolerance of 0.5. The target values were 10, 20, 30, 40, and 50. In the dynamic condition, PPSD was computed over successive 20 ms frames and then averaged over time.

4.2 Protocol and participants

The experiment consisted of six MUSHRA evaluations [7], corresponding to the three source signals combined with the two delay modes. For each evaluation, listeners were asked to rate the quality of the stimuli relative to a reference, based on how similar they sounded to that reference. In addition to the hidden reference and the five PPSD conditions, each page included a mid anchor obtained by low-pass filtering the reference at 7 kHz, and a low anchor obtained by low-pass filtering it at 3.5 kHz. The test was implemented using webMUSHRA [8]. The six evaluations were presented in random order and were preceded by a short familiarization phase.

The test took place in a soundproof booth and stimuli were presented over Beyerdynamics DT770 headphones. Participants were free to adjust the listening level and to take breaks if needed. No specific instruction was given regarding response time.

A total of 22 listeners took part in the experiment (8 female, 14 male), aged 24 to 54 years (median: 39). None reported hearing difficulties. About half of the panel could be considered expert listeners with experience in critical listening, although no specific expertise was required for participation. The median test duration was 11 min 31 s. Screening was based on ratings assigned to the hidden reference: listeners who rated the reference below 80 at least once, or below 90 at least twice, were excluded. According to this criterion, data from two participants were discarded.

4.3 Test Results

A multi-factor analysis of variance (ANOVA) was first conducted to assess the effects of track, delay mode (static versus dynamic), and stimulus condition, as well as their interactions. The results showed that delay mode was not a significant factor. In other words, stimuli generated with dynamic delays were not rated differently from those generated with static delays when they had the same PPSD. Consequently, delay mode was not considered further in the analysis.

A second ANOVA run for parameters stimulus condition, track and their interaction (see Table 1) showed

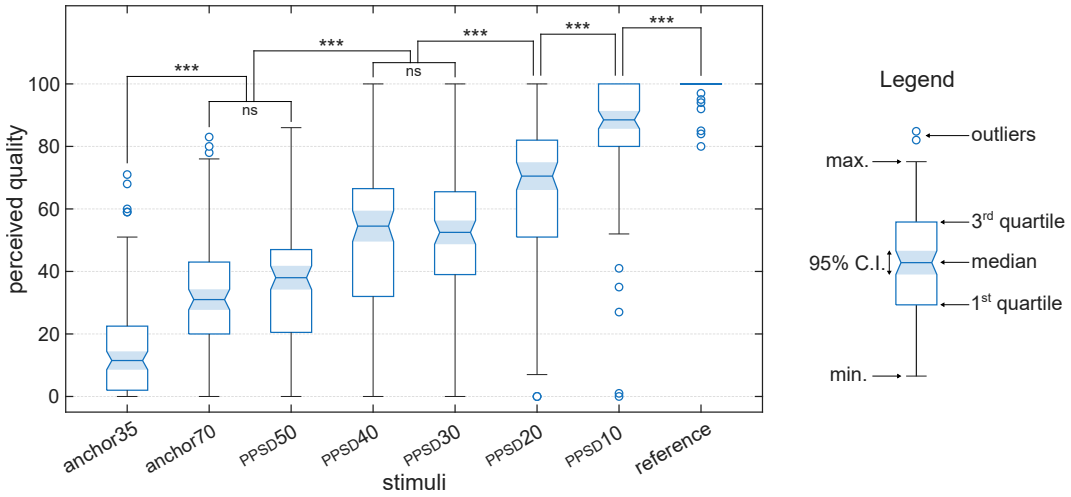


Fig. 4: Results of listening test A: audio quality perceived for the different stimulus conditions (anchor / PPSD / reference). All groups were significantly different from each other, as assessed by a Bonferroni test ($p < .001$), with the exception of the (anchor70, PPSD50) and (PPSD30, PPSD40) pairs.

Var.	Sum sq.	dof	Mean sq.	F	Prob.
cond.	651298.5	7	93042.6	278.95	$<.001$
track	21151.8	2	10575.9	31.71	$<.001$
co.*tr.	32547.5	14	2324.8	6.97	$<.001$
Error	312198.1	936	333.5		
Total	1017195.9	959			

Table 1: ANOVA table: Variation in audio quality perceived during Test A, explained by factors "stimulus condition" (low-pass filtering, PPSD*xx*, or reference), "track", and their interaction. Latency mode (dynamic vs. static) was not significant.

that the three had a significant impact on perceived quality. However, the effect of stimulus condition was by far the dominant one: it accounted for about 64% of the variance in the ratings, whereas track and the track-by-condition interaction accounted for only about 2% and 3%, respectively. Thus, although the source content had a measurable effect, perceived quality was driven primarily by the amount of predicted coloration.

Fig. 4 shows the perceived quality as a function of the stimulus condition. As expected, quality ratings decreased monotonically with increasing PPSD. The low anchor received the lowest ratings, followed by the mid anchor, while the five degraded conditions were mostly ordered from PPSD50 to PPSD10 in increasing quality. A Bonferroni post-hoc comparison indicated that all adjacent conditions were significantly different from one another, except for the pairs (anchor70, PPSD50) and (PPSD30, PPSD40). In perceptual terms, stimuli with a PPSD of 50 were rated similarly to the mid anchor and can therefore be considered of poor quality, whereas stimuli with PPSD values of 20 and 10 were typically rated as good and excellent, respectively.

Overall, these results indicate that the perceived audio quality of the degraded stimuli was determined mainly by the amount of spectral coloration predicted by the auditory model. This supports the use of PPSD as a meaningful descriptor of the perceptual impact of inter-element delays, and provides a first mapping between coloration magnitude and quality judgment.

4.4 Mapping Inter-Element Delay to PPSD

In order to determine which ranges of timing error correspond to a given degree of perceived coloration, we now relate the PPSD value to the underlying inter-element delays. To do so, we generated a large set of simulated delay configurations using the same line-source model as in the previous subsections. For each simulated stimulus, delay values were assigned at random to the different amplifier-fed groups, uniformly

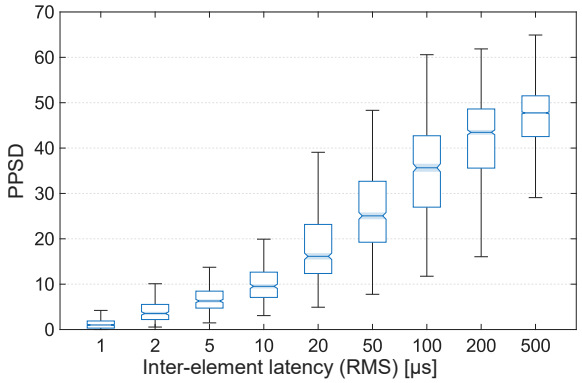


Fig. 5: Predicted perceived spectral difference (PPSD) as a function of the RMS time delay between amplifiers. Simulation run for 12 L-Acoustics K2 speakers fed by four amplifiers. Note that the box plotted for a given RMS delay value x corresponds to the PPSDs obtained for delays which are closer to x. For instance, the box plotted for 10 µs corresponds to delay values ranging from $\sqrt{5 \times 10}$ and $\sqrt{10 \times 20}$.

over a logarithmic scale, so as to span a wide range of timing-error magnitudes. The root-mean-square (RMS) value of the inter-element delays ranged from approximately 0.5 µs to 1 ms. For each configuration, the corresponding binaural response was rendered and its PPSD was computed relative to the perfectly synchronized reference. In total, 5000 such stimuli were generated.

The results are shown in Fig.5, which plots the distribution of PPSD values as a function of RMS inter-element delay. As expected, predicted coloration increased with delay magnitude. Some dispersion can nevertheless be observed for a given RMS value range, because the amount of coloration depends not only on the overall delay magnitude but also on the exact distribution of delays across amplifier groups, and therefore on the resulting interference pattern. Despite this variability, the overall trend remains clear: very small delays produce little coloration, whereas delays in the tens of µs and above lead to substantially larger PPSDs.

More specifically, PPSD values remained generally below 10 for RMS delays on the order of 10 µs and below. By contrast, for RMS delays on the order of 50 µs and above, PPSD values were mostly greater than 20. In listening test A, stimuli with PPSD = 10 and 20 were typically rated as having excellent and good quality, respectively, whereas larger PPSD values were associated with a clearer degradation in perceived audio quality. This implies that RMS inter-element delays should remain on the order of 10 µs and below for the reproduced sound to be perceived as consistently good to excellent, while delays reaching the tens of microseconds are expected to produce a more substantial quality degradation.

5 Listening test B: Audibility Threshold

In this second test, we focused on smaller impairments to determine under which conditions the effects of inter-element delay become audible. Specifically, the aim was to estimate the minimum amount of predicted coloration for which a stimulus could be distinguished from the perfectly synchronized reference.

5.1 Stimulus generation

The stimuli were generated using the same general method as in Listening Test A, but with two modifications. First, only two source signals were used: the reggae excerpt (*Lola Rastaquouère*, by Serge Gainsbourg) and Gaussian pink noise. Second, the target coloration values were chosen in a lower range, with target PPSD values of 1, 2, 5, 10, and 20. The selected PPSD range was intended to span the transition from imperceptible to clearly audible coloration. As in the previous test, degraded stimuli were generated by applying delays to the group signals until the resulting binaural rendering matched the target PPSD within a small tolerance.

5.2 Protocol and participants

Because the perceptual differences under study were much smaller than in Listening Test A, the experiment followed the ITU-R BS.1116 methodology [9], which is specifically intended for the assessment of small impairments. The test interface was implemented using webMUSHRA and began with a familiarization phase during which participants could listen to all stimuli, explicitly labeled and sorted by condition. This phase was followed by 10 evaluations, corresponding to the combination of 2 tracks and 5 coloration conditions.

In each evaluation, listeners were presented with a reference stimulus, labeled A, and two comparison stimuli, labeled B and C. One of these two stimuli was the

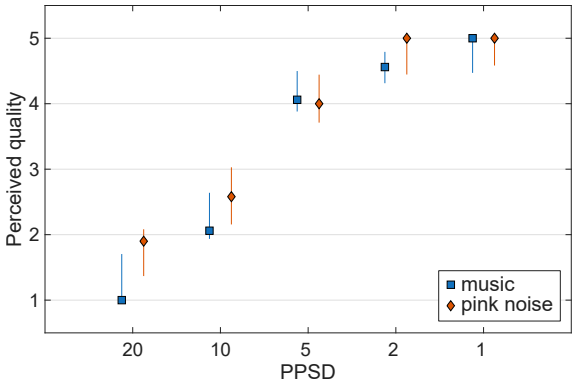


Fig. 6: Results of listening test B: perceived quality (median values and interquartile ranges) as a function of the stimulus PPSD for the music and pink noise tracks.

hidden reference and the other was the degraded stimulus under test. Participants were asked to rate the perceived difference relative to the reference on a scale from 1 ("very annoying") to 5 ("imperceptible"), in accordance with the ITU recommendation.

A total of 16 expert listeners took part in the test (2 female, 14 male), aged 25 to 55 years (median: 42). All reported normal hearing and were accustomed to critically comparing sound systems. Screening was performed using the ratings assigned to the hidden reference, following the same principle as in Listening Test A. None of the participants were excluded.

5.3 Test results

Figure 6 summarizes the ratings obtained for the two program materials as a function of PPSD. As expected, the smaller the predicted coloration, the closer the degraded stimuli were rated to the reference. For PPSD values of 5 and above, the differences were clearly perceptible for the large majority of listeners. For PPSD values of 2 and below, however, the distinction from the reference became less consistent.

For the music excerpt, a t-test indicated that the ratings obtained for PPSD = 1 were not significantly different from those assigned to the reference ($p = 0.155$). For pink noise, ratings were not significantly different from the reference for both PPSD = 1 and PPSD = 2 ($p = 0.298$ and $p = 0.171$, respectively). These results indicate that the audibility threshold lies around a PPSD of 2, below which the spectral changes caused by inter-element delay are generally difficult to detect. This suggests that the perceptual threshold for inter-element delay is much lower than the threshold at which quality becomes substantially degraded.

Referring back to the simulation results of Section 4, a PPSD on the order of 2 corresponds to inter-element delays on the order of 1 µs. This suggests that coloration caused by inter-element delays becomes audible at extremely small timing offsets. Unexpectedly, the effect appeared to be slightly easier to detect with the music excerpt than with pink noise, although the overall trend was consistent across both program materials.

6 Discussion

The present results show that inter-element delays in a line source are perceived mainly as spectral coloration. More specifically, Listening Test A showed that increasing PPSD consistently leads to lower quality ratings, while Listening Test B indicated that much smaller PPSD values are already audible. The delay-to-PPSD mapping helps interpret these perceptual results in terms of synchronization error. In the simulations, RMS inter-element delays on the order of 1 µs corresponded to coloration close to the audibility threshold, whereas delays in the tens of microseconds were associated with much larger PPSD values and clearly lower quality ratings. For line-source systems, this suggests that synchronization should remain on the order of 1 µs or below if perceptual transparency is desired. More generally, the results show that timing accuracy in networked sound systems should be considered as a perceptual requirement, not only as a technical one.

One limitation of this study is that the listening tests relied on instantaneous comparison between a degraded stimulus and a perfectly synchronized reference. In practice, however, timing errors in deployed systems may develop progressively. The aim of the present study was not to reproduce the exact temporal dynamics of such impairments, but rather to determine under controlled conditions whether the resulting spectral changes are perceptually detectable and how they affect judged quality. In that respect, direct comparison provides a sensitive estimate of both audibility and quality thresholds. It is possible that slowly varying desynchronization would be perceived differently, or would be less salient than in the present tests. Nevertheless, this would not alter the conclusion that the underlying

spectral distortion becomes perceptually relevant once inter-element delays reach sufficiently large values.

Another limitation is that our simulations and tests were based on one representative line-source configuration. However, the exact relationship between inter-element delay and spectral coloration depends on the array geometry. The thresholds reported here should therefore be interpreted as order-of-magnitude results rather than universal values. That said, the configuration considered is typical of large-scale sound reinforcement, and the effects could in fact be more severe in systems where each loudspeaker is driven by a separate amplifier, for instance.

Finally, the effect of coloration seem to depend somewhat on program material, although this effect was small compared to that of the stimulus condition itself. Despite these limitations, our results converge toward the same conclusion: microsecond-scale timing errors are perceivable, and delays in the tens of microseconds can substantially degrade perceived quality.

6.1 Conclusions

This study examined the perceptual consequences of inter-element timing errors in line-source systems. The results showed that such delays modify the array interference pattern and are perceived primarily as spectral coloration. Listening tests further indicated that this coloration affects both perceived audio quality and audibility, with detection occurring for delays on the order of 1 µs and more substantial degradation appearing once delays reach the tens of microseconds range.

These findings suggest that synchronization accuracy in networked sound reinforcement should be treated as a perceptual requirement, not only as a technical one. More generally, they provide practical targets for the design and operation of line-source systems in contexts where preserving consistent timbre and high audio quality across the audience is essential. The perceptual results presented in this paper are complemented by a companion study in which timing inaccuracies are measured and characterized in Dante, AES67, and Milan-AVB Audio over Ethernet networks [10].

References

[1] Christensen, C., "AVB Application Requiring Better than 5uS Accuracy of the PTP Clock," technical report contributed to the ieee 802.1 standard, The Institute of Electrical and Electronics Engineers (IEEE), 2007.

[2] Kaltheuner, H., Kronauer, G., Lave, M., and Corteel, E., "Requirements and Solutions for Audio Networking in Sound Reinforcement Systems," in *AES 2024 International Conference on Acoustics & Sound Reinforcement*, Le Mans, France, 2024.

[3] Hortig, P., "Time and Media Clock Synchronization Resilience in Converged Audio Networks," in *33rd Tonmeistertagung (TMT33)*, Düsseldorf, Germany, 2025, ISBN 9783982468020.

[4] Bernschütz, B., "A Spherical Far Field HRIR / HRTF Compilation of the Neumann KU 100," in *39th DAGA*, pp. 592—-595, Merano, Italy, 2013.

[5] McKenzie, T. and Brinkmann, F., "Toward an Improved Auditory Model for Predicting Binaural Coloration," *Journal of the Audio Engineering Society*, 73(3), pp. 115–126, 2025, doi:10.17743/jaes.2022.0192.

[6] Majdak, P., Hollomey, C., and Baumgartner, R., "AMT 1.x: A toolbox for reproducible research in auditory modeling," *Acta Acustica*, 6(19), 2022, doi:10.1051/aacus/2022011.

[7] "ITU-R Recommendation BS.1534-3. Method for the subjective assessment of intermediate quality level of audio systems," standard, The ITU Radio communication Assembly, 2015.

[8] Schoeffler, M., Bartoschek, S., Stöter, F., Roess, M., Westphal, S., Edler, B., and Herre, J., "webMUSHRA — A Comprehensive Framework for Web-based Listening Tests," *Journal of Open Research Software*, 6(1), p. 8, 2018, doi:10.5334/jors.187.

[9] "ITU-R Recommendation BS.1116-3. Methods for the subjective assessment of small impairments in audio systems," standard, The ITU Radio communication Assembly, 2015.

[10] Duval, B., Doukopoulos, X., Hannotin, D., Epain, N., Iortoman, A., and Kronauer, G., "Qualifying Timing Errors in Audio-over-Ethernet Networks for Live Sound," in *160th AES Convention*, Copenhagen, Denmark, 2026.