ADAPATIVE DELAY SYSTEM (ADS) FOR SOUND REINFORCEMENT

Elliot A. Smith and Jay S. Detsky
Systems Design Engineering, University of Waterloo, Waterloo, Ontario, N2L 3G1

ABSTRACT

At concerts and presentations, the sound system must be carefully calibrated to ensure the entire audience can hear the presenters clearly. For both indoor and outdoor venues, this is done by positioning speakers throughout the audience to reinforce the sound produced on stage. This technique introduces an added complexity, whereby the electrical signal to the speakers in the crowd travels much faster than the sound wave coming from the stage.

The Adaptive Delay System (ADS) for Sound Reinforcement is a new method for synchronizing the sound throughout the audience. Unlike existing methods, it does not require complex calculations when initially configuring the sound system. Furthermore, it is also capable of accounting for time-variant conditions such as wind, which are neglected in the current methods of speaker synchronization. Maximum length sequences, a special type of pseudo-random noise, are injected into a speaker for several seconds. A specially placed microphone picks up this sound which is then cross correlated with the original noise to determine the propagation delay. As this sound is barely audible it can be used during a concert to adaptively correct for changing conditions.

SOMMAIRE

Pendant concerts et présentations, la système auditif doit être précisément calibré pour que tous entendent la présentation. Des haut-parleurs, placés stratégiquement dans l'audience, renforcent le son produit sur la scène. Cette méthode est efficace dans une salle bien qu'en plain air. Cette technique introduit une nouvelle complexité, vue que les signaux électriques destines aux haut-parleurs se propagent plus vite que le son provenant de la scène.

La Système de Délai Adaptif (SDA) pour renforcir le son constitue une nouvelle méthode pour synchroniser le son dans l'audience. Contrairement aux méthodes existantes, celui-ci ne nécessite pas des calculassions complexes lors de la configuration initiale du system du son. En plus, la SDA est capable de prendre en compte les variables comme le vent, qui changent avec le temps. Ces variables sont omises dans les méthodes de synchronisation des haut-parleurs actuelles.

1 BACKGROUND

1.1 Statement of the Problem

There are many situations where multiple sets of speakers (called delay towers) are dispersed amongst a crowd or a room to reinforce music or speech for the entire audience. In these situations, the electrical signal traveling to the delay towers moves much faster than the sound wave coming from the stage speakers, thus the electrical signal must be delayed so that the two sound waves are synchronized. In some cases, such as convocations in gyms or large conferences, the delay is neglected, resulting in poor intelligibility. For large outdoor concert venues, the existing methodology for determining the delay for each speaker is cumbersome. It requires estimates of the weather for concert days and is not adaptable to changing conditions. Thus there is a need for an automated, adaptive system for calculating sound delay.

2.2 Requirement for Adaptive Delay

Large outdoor concerts are often held over two or three days in the summer, when weather changes can be most extreme. Temperature, relative humidity (RH), and changing wind speeds can greatly affect the speed of sound [1]. With a change in the speed of sound comes a corresponding change in the time it takes for sound to travel between delay towers. This means that if the delay is calculated at the beginning of a concert, by the end, the weather conditions might have changed sufficiently to create a noticeable error in the configured delay. The speed of sound can be shown to depend on temperature in the form of the following equation [1]:

$$c = 331.45\,(1+t/273)^{1/2} \tag{1}$$

where c is the speed of sound in m/s, and t is the temperature in degrees Celsius. Relative humidity has a smaller but still

		Temp.	RH	Wind speed	Speed of sound	Delay (ms) for distance:		
		(°C)	(%)	(km/hr)	(m/s)	40 m	60 m	100 m
Typical change	Day	22	80	0	344.4	116.2	174.2	290.4
	Night	10	40	25	330.8	120.9	181.4	302.3
					Error:	**4.7**	**7.2**	**11.9**
Extreme change	Day	40	100	0	358.9	111.5	167.2	278.6
	Night	15	80	55	325.9	122.7	184.1	306.9
					Error:	**11.2**	**16.9**	**28.3**

Table 1. Delay Errors introduced y changing weather conditions.

noticeable effect on the speed of sound.

The delay errors introduced by a typical and extreme change in weather are shown in Table 1. The weather data is extracted from the Environment Canada website [2]. The delay errors are calculated for minimum, typical, and large distances between delay towers, corresponding to 40 m, 60 m, and 100 m. From the table it can be seen that typical changes in weather are sufficient to cause noticeable delay error.

2.3 Perception of Delay

Sound degradation can begin when the primary sound source and the secondary sound source are as little as one millisecond apart [3]. Below this limit, the delay results in stereophonic sound; above this limit, changes to the sound become noticeable, as the tone colour of the sound changes and the 'centre of gravity' (where the listener perceives the sound source to be) begins to shift towards the secondary sound source. Once the delay reaches a certain threshold, known as the echo threshold, what was previously perceived as a single sound event is separated into two distinct events. There is no clear rule for determining the delay at which an echo threshold is reached. In the worst possible case, the echo threshold is reached at two milliseconds; however, in different circumstances the threshold may not be reached until the delay is 30 milliseconds. The threshold depends on multiple factors, such as the angle of the listener to the sound source, the type of sound, and the level of the sound. In the case of sound level, the louder the sound, the shorter the delay before the echo threshold is reached.

A delay that may cause minimal degradation in some circumstances may exceed the echo threshold in other circumstances. Even in cases where the echo threshold is not breached, and two distinct sounds are not recognizable, sound degradation may still be a factor. Therefore, because no standard for delay tolerance can be set, and changing conditions can produce delays varying from 5 ms to 30 ms, it is desirable to minimize the delay as much as possible in all circumstances.

3 SOLUTION

3.1 Solution Framework

It was decided that an open loop analysis would be used to determine the delay time between speakers. This was decided upon because of the ease with which delay can be calculated by using a microphone (behind the rear speaker) to record the sound from the front speaker. Figure 1 shows the setup for this acquisition. By having the same computer processor control the output from the stage speaker *and* record the input sound from the microphone, various mathematical methods (described in section 2.2) can be used to calculate the delay between the two speakers. This setup is considered an open-loop analysis because there is no feedback from the sound downstream of the second speaker to the system used to calculate and adjust the delay.

For the measurement of sound decay, it has been established

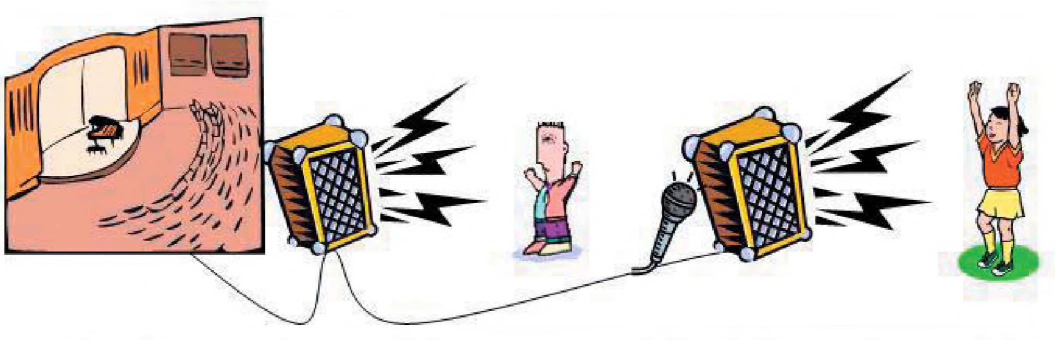
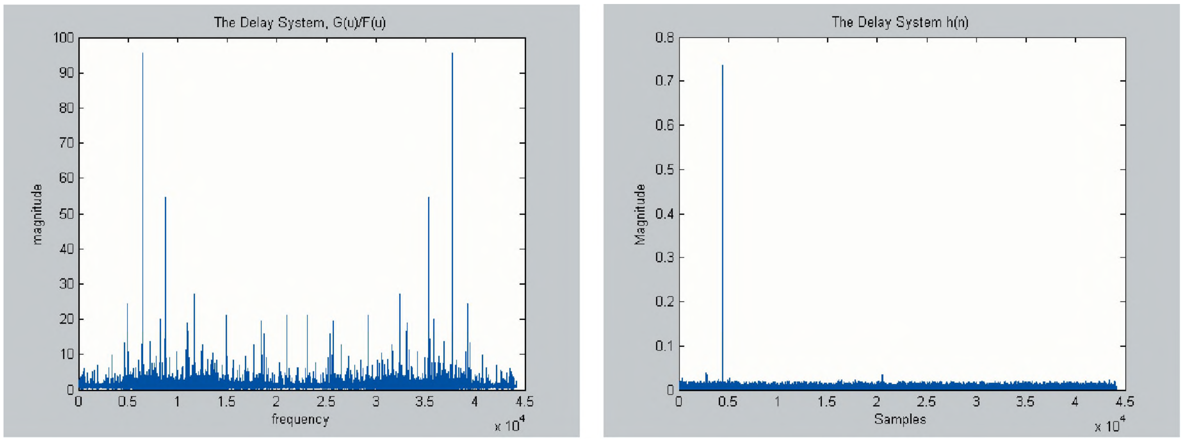


Figure 1. Delay tower set up



(a) (b)

Figure 2. The delay system (a) in the frequency domain, H(u), and (b) in the time domain, h(t)

that a room or outdoor environment can be modelled as a linear time-invariant (LTI) system with respect to its acoustics [4]. The general framework for an LTI system can be expressed as [5]:

$$g(t) = f(t) * h(t), \tag{2}$$

where * is the convolution operator, f(t) is the input (in this case the sound at the stage speaker), g(t) is the output (sound behind the rear speaker), and h(t) is the impulse response of the system, which here is a delay process with an impulse at some unknown delay time. The location of the impulse in h(t) is required to accurately determine and implement the required delay for the rear speaker.

The need for an adaptive delay system stems from the fact that the properties of an outdoor concert venue are time-varying. It is, nonetheless, reasonable to consider such a venue an LTI system over the short time (on the order of seconds) required to calculate the delay. Over a time period of several seconds, the factors affecting the speed of sound do not vary wildly, and thus the venue can be considered LTI for the duration of a single test to determine the delay time.

3.2 Alternate Solutions

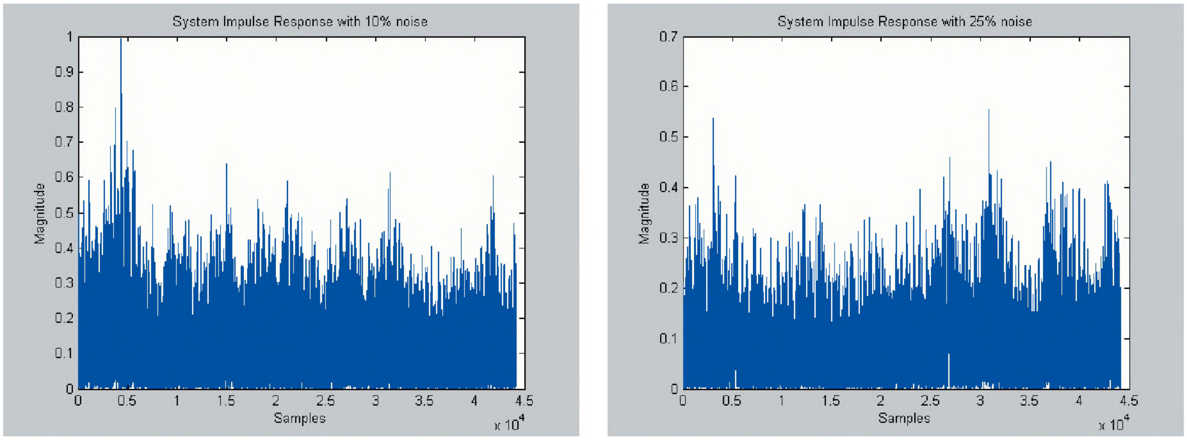
Many possible methods exist to determine the delay. The first two possible solutions examined used the music that would be normally played through speakers at a concert as the signal for determining the required delay. The first method that was investigated uses the inverse system to take advantage of the LTI nature of the sound propagation. The recovery of the delay system requires solving the so-called deconvolution problem in order to determine the delay time encompassed within h(t). This can be most easily solved in the frequency domain where the system equation becomes the following:

$$G(u) = F(u)H(u),$$
or solving for H(u), (3)
$$H(u) = G(u) / F(u).$$

Therefore, taking the Fourier transform of the sample and record systems, dividing, and returning to the time domain should yield an impulse at the time of the delay. Figure 2 shows the delay system H(u) in the frequency domain and the delay system h(t) in the time domain when the above methodology is implemented in Matlab with a 100 ms artificial delay created in software. In Figure 2b, an impulse is present at sample 4410, which at a sampling rate of 44.1 kHz corresponds to a 100 ms delay. This impulse represents the delay in the system and can in theory be used to determine how long to delay the music at the delay tower to ensure that all the music waves are in phase.

This system is straightforward to implement, using well-established signal processing theory. It would cause no concert interference as this system uses only the concert sound and existing speakers to measure the delay. There is, however, a major weakness to this solution. In the presence of noise (such as audience members cheering and clapping, as would be the case at any concert), the signal to noise ratio of the spike from inverse system begins to decrease. This condition was simulated by adding acoustic noise on top of a music file and performing the subsequent inverse system calculation. Figure 3 shows the results of calculating h(t) when adding acoustic noise at 10% and 25% of the total speaker output (with the other 90% and 75% being the music, respectively). With 10% acoustic noise, the impulse is still present, but the 25% case is the marginal situation where the noise floor of h(t) has equal power to the impulse, meaning the delay can no longer be confidently determined. As a result, the inverse system methodology of measuring the delay has poor robustness for external acoustic noise and is not practical for implementation.

A second approach considered using the cross-correlation of the original sound with the delayed sound. Cross correlation is a method for mathematically comparing two signals. At the point where the two signals are perfectly in phase, all the peaks and valleys of the waves will be aligned and the cross correlation yields a detectible spike. When the waves are not in phase, most of the peaks and valleys will cancel one another. The location of the spike from cross correlating the



(a) (b)

Figure 3. (a) Impulse corresponding to 10% noise and (b) 25% noise

speaker and microphone signals will thus provide an impulse at the delay time between the two signals. This method is simpler than the inverse system method to implement because it does not involve Fourier Transforms, and like the inverse system, it does not interfere with the concert. However, like the inverse system, cross correlation of the music signals fails in the presence of external acoustic noise.

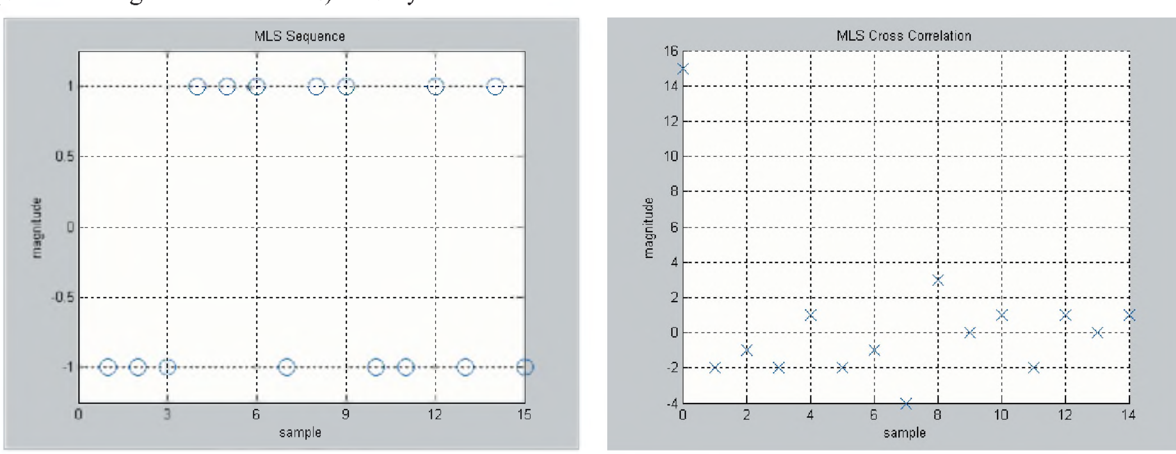
3.3 Maximum length sequences (MLS)

To deal with the problem of acoustic noise, a method to recover the impulse response of a system using maximum length sequences (MLS) [6] and a cross correlation operation was examined. MLS is pseudo-random binary white noise, specifically designed to have no internal periodicity. An MLS sequence can be of any length L, such that

$$L = 2^n - 1, \text{ where n is a positive integer} \qquad (4)$$

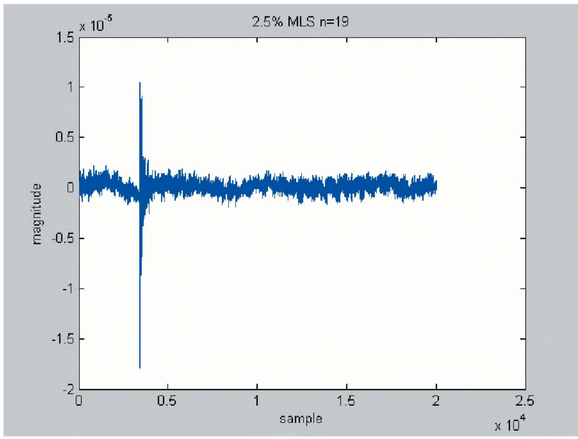
and its elements are either +1 or –1. When an MLS sequence is cross correlated with itself the result for all elements except at element zero (corresponding to zero phase between the two sequences being cross correlated) is very small. This occurs because the cross correlation yields a series of plus-ones and minus-ones, and when these are summed the plus-ones approximately cancel with the minus-ones. In the case where the sequences are in phase, the cross correlation yields a series completely of plus-ones and therefore the sum adds to L, the length of the sequence. Figure 4a shows a sample MLS sequence of length 15. Figure 4b shows the result of that sequence cross correlated with itself.

The MLS illustrated in Figure 4a is 15 elements long and thus the cross correlation results in a spike of value 15; the value of the next largest element is three. This sequence, however, is atypically short, for illustrative purposes. If a more typical MLS of length $2^{18}–1$ (six seconds at a 44.1 kHz sampling rate) is generated, the spike has a value of 262×10^3 and the next largest element has value 600. The delay of an acoustic propagation system can therefore be characterized by playing MLS through the stage speaker, recording it at the delay tower, cross correlating both the original and recorded signals, and locating the spike. The pseudo-random nature of MLS makes it sound like static, similar to noise heard when tuning a radio. As a result of this property, low-level MLS can be injected into a sound system while a concert is

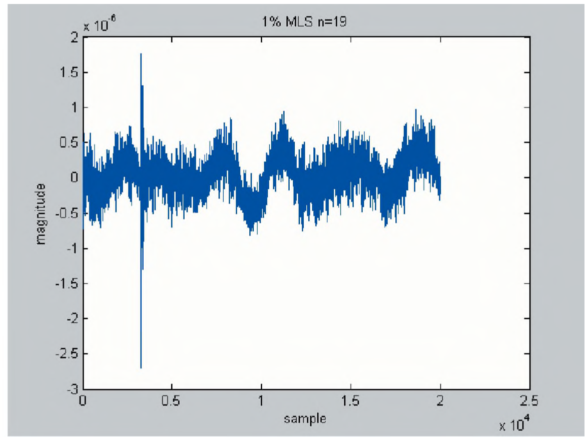


(a) (b)

Figure 4. (a) MLS of length 15 and (b) MLS cross correlated with itself



(a)



(b)

Figure 5. (a) Impulse generated by cross correlating an MLS sequence lasting 12 seconds at an intensity of (a) 2.5% and (b) 1.0%

in progress without an overly adverse affect on sound quality. Because the spike from cross correlating MLS is significantly larger than the background noise, MLS can be injected at a fraction of the intensity of the music [5]; for example, MLS noise can be controlled to be 10% of the total sound output of the speaker with the other 90% being the concert music.

Implementing MLS only differs from implementing the cross correlation solution in two ways. Firstly, MLS must be mixed in with the music for the period during which the system is being characterized, and secondly, the recording taken at the delay tower is cross correlated with the original MLS signal rather than the music. The use of MLS to initially characterize a system and establish the delay time has no negative impact on the concert because the calibration can be done before the audience has arrived. In order to *adaptively* correct delay to account for changing conditions, MLS must be played during the concert, which will have some impact on sound quality. In low-noise applications, barely audible MLS can be used to calculate delay; in noisier environments the MLS power must be boosted. Therefore, the interference caused by MLS is proportional to the background noise; however, the noisier the environment, the less sensitive the audience will be to the added noise generated by MLS. Furthermore, the interference caused by MLS can be optimized through a balance of intensity and duration; the longer the sequence used, the lower the required intensity to generate a measurable spike. Therefore, as crowd noise increases, the duration of the MLS can be increased instead of its intensity. Conversely, a shorter MLS at a higher intensity can also be used measure the delay.

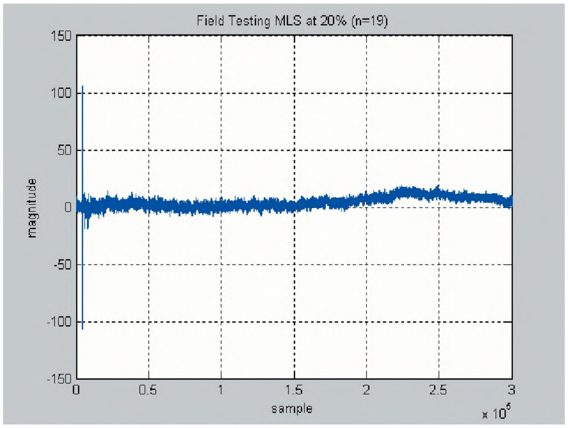
The most significant benefit to using MLS is its ability to reject noise. Because the cross correlation is done against the original maximum length sequence and not the music, noise is treated differently here than in the other solution concepts. Here, MLS is the signal of interest and both the music and crowd noise is regarded as noise in the cross correlation. Figure 5 shows the impulse generated using an MLS sequence lasting twelve seconds ($L = 2^{19}-1$). This sequence length was chose because it is practical for use. In Figure 5a, the MLS power is 2.5% of the speaker output and in Figure 5b it is 1.0%, yielding signal-to-noise ratios of 39 and 99 respectively. The spikes in both cases are clearly discernable.

4 IMPLEMENTATION

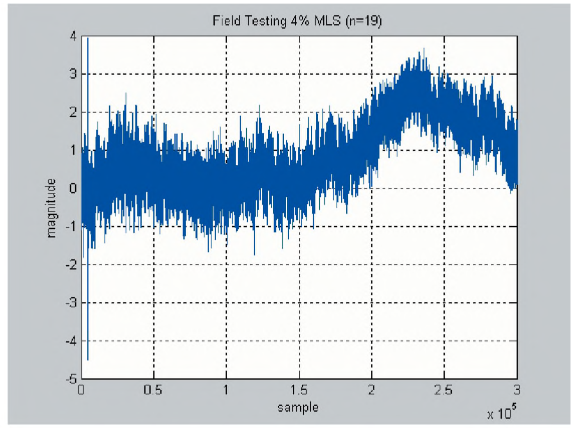
The Adaptive Delay System (ADS) is comprised of Matlab software written to calculate the delay of a system using MLS. This software is run on a personal computer that controls the output of both the stage and delay speakers and receives input from the microphone. An MLS sequence is generated using software obtained from the Matlab website [7]. This MLS sequence is then added on top of the output music of the front speaker at a fraction of total speaker output that is controllable. The delayed signal at the position of the rear speaker is recorded with a microphone, and this recorded signal is cross correlated with the original MLS sequence, yielding a spike at the delay time between the two speakers. The rear speaker is then delayed by the time corresponding to this spike in order to synchronize the sound from the two speakers.

One problem encountered with the software was computer-specific calibration. The delay calculation depends on playing MLS and immediately starting to record the sound being captured by the microphone. However in Matlab it takes a certain amount of processing time after calling *wavplay* before the computer can process the record command *wavread*. This causes the measured delay to be offset by the amount of processing time required by the specific computer being used. Therefore, calibration software was written to determine the computer-specific processing delay error. Note that this problem is relevant only when using Matlab on a personal computer, and that the idealized final product would use a microchip that would not have this calibration issue.

In theory, the MLS methodology is sensitive to a single sample, which at 44.1 kHz is approximately 0.02 ms. The calibration calculations, however, typically have a standard deviation of 1.0 ms so in practice this limits the sensitivity



(a)



(b)

Figure 6. Field testing result using an MLS sequence of 12 seconds at (a) 20% and (b) 4% of total speaker output

of the calculation to one millisecond. To test this limitation, a series of five measurements was repeated using the same parameters, including a 100 ms delay. The resulting five calculated delays had an average of 100.1 ms with a standard deviation of 0.1 ms which confirmed that the sensitivity is within one millisecond.

5 TESTING

Two large speakers, a microphone, and all necessary equipment to connect this audio setup to a personal computer in order to test the ADS was installed in the main gymnasium at the University of Waterloo to perform field measurements and validate the ADS prototype. The two speakers were set up 100 metres apart, with a microphone placed behind the second speaker. The speakers and microphone were hooked up to a laptop computer running the ADS software as described in section 3.

Figure 6a shows the result of injecting MLS for 12 seconds at a level of 20% of the total speaker output. The spike in this case is prominent over the background noise. At this level, however, the MLS was deemed to be intrusive. Tests were repeated until the minimum intensity of MLS that still produced a discernable spike was found. This occurred when the MLS was at a level of 4% and is shown in Figure 6b. In order to evaluate the potential to use longer sequences at a lower intensity, the test was repeated with an MLS lasting 48 seconds, which is the longest sequence that can be cross correlated using Matlab and one gigabyte of memory. The lowest intensity for a sequence of this duration was found to be at 2% of total speaker output (results for this case shown in Figure 7). This confirmed the hypothesis that longer sequences can be used as an alternative to playing the sequences at a higher volume. The improvement in audio

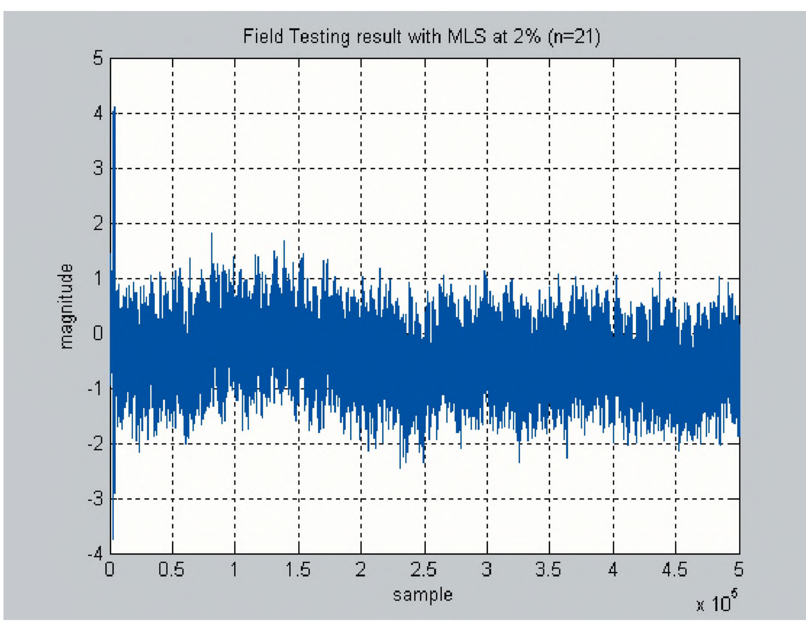


Figure 7. Field testing result with a longer MLS sequence (48 seconds) at 2% total speaker output

quality after implementing the ADS system was significant.

6 CONCLUSIONS

Issues with audio delay at large venues can cause poor sound quality and low speech intelligibility. The existing tools on the market to deal with delay are not adaptive to changing conditions, and more significantly, are time-consuming to configure and set up. The Adaptive Delay System is a new adaptive and automated method for synchronizing the sound between speakers at these large venues, in order to compensate for the electrical impulses that travel faster to the delay towers than the sound that travels from the stage.

In order to deal with the noise present at concerts, the ADS uses low level MLS added on top of the audio coming from the stage speaker in order to accurately determine the required delay time to the delay towers. This prototype was validated through field testing done in the University of Waterloo's main gymnasium. Positive results from prototype testing clearly demonstrated the success of the Adaptive Delay System for Sound Reinforcement.

The ADS was entered into the Ontario Engineering Competition and Canadian Engineering Competition in the Corporate Design category, where it placed first and second respectively.

7 REFERENCES

[1] Bohn, Dennis A. *Environmental effects on the speed of sound.* J. Audio Eng. Soc., Vol. 36, No. 4, April 1988.

[2] Environment Canada. "Canadian climate normals." Online at http://www.climat.meteo.ec.gc.ca/climate_normals/index_e.html.

[3] Blauert, Jens. *Spacial Hearing: The Psychophysics of Human Sound Localization.* MIT Press, Massachusetts: 1983.

[4] Schroeder, M. R. *Integrated-impulse method measuring sound decay without using impulses.* J. Acoust. Soc. Am. Vol 66, No 2. August 1979.

[5] Oppenheim, Alan V. and Alan S. Willsky. *Signals and Systems, Second Edition.* Prentice Hall, New Jersey: 1997.

[6] Rife, Douglas D. and John Vanderkooy. *Transfer-Function Measurement with Maximum-Length Sequences* J. Audio Eng. Soc., Vol 37, No 6. June 1989.

[7] Matlab website. "Matlab central file exchange – sound". Online at http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=1246&objectType=file.