A MODERN SOLUTION FOR AUTOMATED NOISE SEPERATION (AMS-ANS)

C Nicolaides AMS Acoustics Ltd
X Babington AMS Acoustics Ltd

1 INTRODUCTION

Ambient noise sensing is today a very common term in the field of electro-acoustics and is essentially a fundamental part of a compliant PA system in the mass-transit sector. It has proven to be a very useful addition to any PA system since it strives to maintain the output of an announcement at a pre-defined Sound Pressure Level (SPL) above the background noise.

There are currently many available Ambient Noise Sensor (ANS) products available on the market, most of which are developed and supplied by the most common VACIE manufacturers. Although necessary adjustments have been applied along the years to both the hardware and software components of these products, it appears that no in-depth research has, until recently, solely focused on this particular element within the signal chain. This could partly explain why most PA systems with “off the shelve” ANS solutions still have major setbacks and can only work to their full potential in a “favorable” noise environment.

In public spaces where costumer information is paramount, good Intelligibility is essential and therefore so is the signal-to-noise ratio. While a standard ANS system will sample the background noise and consequently adjust the output SPL prior to a PA announcement being broadcast, there is always the potential for the background noise level and/or spectrum to vary during that announcement, hence affecting the STI.

Although some “quick-fix” solutions have previously sprung to mind in order to overcome this problem, this paper will introduce a slightly different and somewhat unconventional concept to ambient noise sensing through a more rigorous and scientific approach while remaining practical in terms of development.

2 GENERAL PRINCIPLE

The main drive behind this research was founded from the idea of active noise monitoring and noise cancellation at the pre-processing stage of the ANS functionality. In simple terms, one would be looking at a microphone monitoring the background noise in real time during a live announcement while in parallel (software / DSP allowing) adjusting the SPL of the broadcast signal at a fixed signal-to-noise (SNR) ratio.

As can be seen from Figure 1, the speech signal is generated by a Digital Voice Announcer (DVA) before being split into two separate paths. Meanwhile the ANS microphone is continuously monitoring the space background noise as well as any simultaneous PA broadcasts which are all gathered by the DSP unit.

For one of the signal paths, the original ‘dry’ speech is convoluted with the acoustic parameters of the space and the resulting output signal is also sent to the DSP. The other signal path introduces a short time delay until the synchronised software processing has completed.

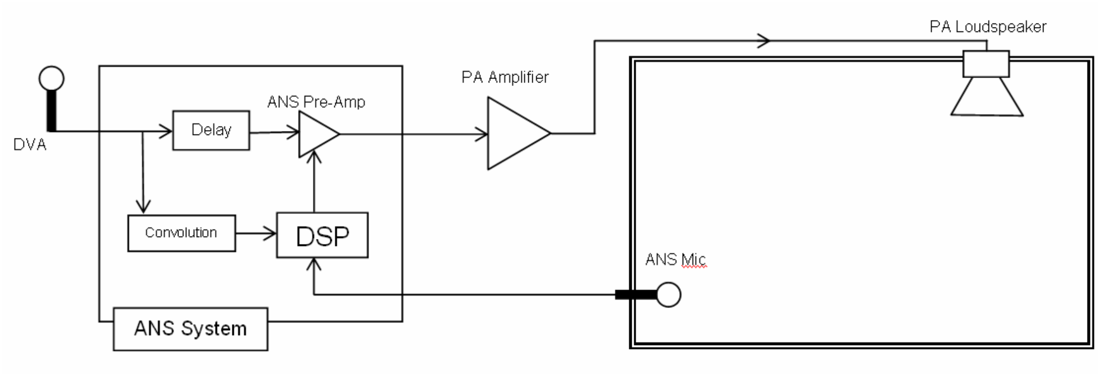

Figure 1 - General concept schematic

The noise separation process is achieved by inverting the convoluted speech signal and carefully time-aligning it with the signal (including noise) picked up by the ANS microphone. Since the processed and measured speech signals are now of equivalent content as well as in anti-phase, they are expected to cancel each other out. This then leaves the background noise information of the space at that particular instant based on which the level of the delayed dry speech signal can be adjusted and sent to the loudspeakers.

In theory, the quicker this complex process is carried out, the smoothest the variations in SPL will be, hence resulting in the best possible intelligibility of the PA broadcast.

This main obstacle to overcome is the effect of reverberation which largely contributes to the spectrum and sound decay of the noise sources within the space. This presents a greater challenge still since all areas will have their bespoke (proprietary) acoustic parameters which will need to be analysed by the ANS device to prove effective.
In theory the acoustic parameters of the space would be pre-stored in the convolution unit but there are obviously limitations due to infinite amount of RT scenarios.

3 METHODOLOGY

3.1 Computer Simulation

In order to verify the validity of this concept, a theoretical model based on an ideal scenario was simulated on a desktop computer using anechoic speech recordings and the following software tools:

i. CATT Acoustic FIReverb
ii. Adobe Audition 3 (previously Cool Edit Pro)

Using a pre-recorded speech signal along with the RT figures given in table 1, the dry speech signal was convoluted before being contaminated with a constant noise signal. For this example, an arbitrary signal-to-noise ratio of +6dB was chosen.
A multi-track session is then constructed in Adobe Audition which will simulate the processing that takes place in the ANS device. The first track consists of the convoluted speech contaminated with noise and the second track contains the inverted convoluted ‘clean’ speech signal. These two signals are then perfectly aligned in the time domain to achieve maximum cancellation.
After the level of the last signal is adjusted such that the highest level of cancellation is achieved, the master (output) track results in obtaining the noise signal alone.

3.2 Field Measurement

In order to divert from the theoretical model, an ANS system mock-up was installed in a reverberant space to enable the capturing of signals in-situ. The setup was as shown in Figure 2 below:

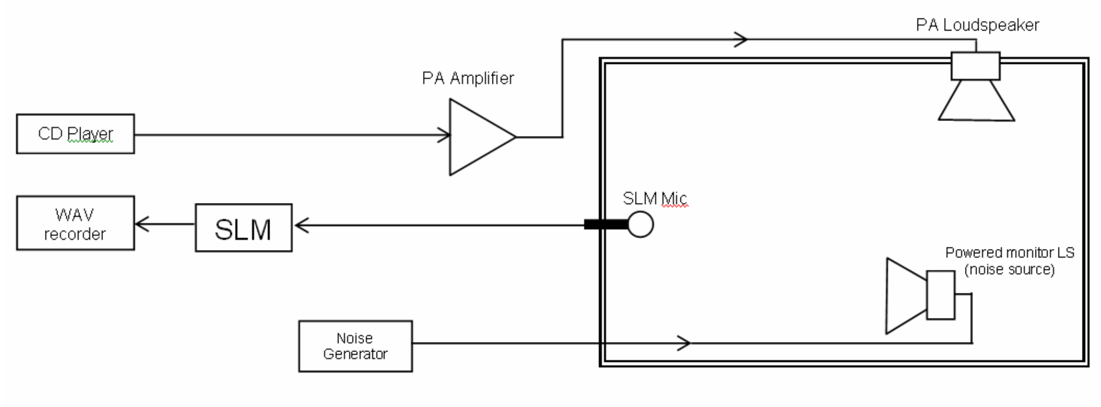

Figure 2 - Field Measurement setup

In the first instance, a WinMLS swept sine wave signal was played through the PA loudspeaker and captured by the ANS microphone (SLM in this instance) to record an impulse response of the room which will later be used in post-processing.

The pre-recorded dry speech signal was then played through the system and recorded at the ANS microphone as a 'wav' file. Background noise was then introduced on top of the speech via a separate loudspeaker source and in turn recorded. This process was carried out at an arbitrary signal-to-noise ratio of 6dB.

The captured impulse response from the PA system in the space was then convoluted with the dry speech signal and the resulting signal was then inverted.

Again using a multi-track session in Adobe Audition, the cancellation process of the ANS unit was simulated by carefully aligning the inverted convoluted signal and the recorded noise contaminated speech.

4 RESULTS

4.1 Computer Simulation

Using the anechoic speech recording along with a preset impulse response created in FIReverb, the convoluted signal was imported into Adobe Audition and subsequently inverted.

Figure 3 - Inverted Convoluted speech

The combination of the convoluted speech signal and background noise introduces a noise floor which masks part of the speech content as shown in Figure 4 below.

Figure 4 - Noise contaminated convoluted speech

Finally, the summation of the two processed signals results in the cancellation of the speech transients and hence the bulk of the speech content. Only the background noise floor remains.

Figure 5 - Isolated noise from signal subtraction

4.2 Field Measurement

Figure 6 shows the inverted result of the reference anechoic speech recording convoluted with the measured impulse response of the room.

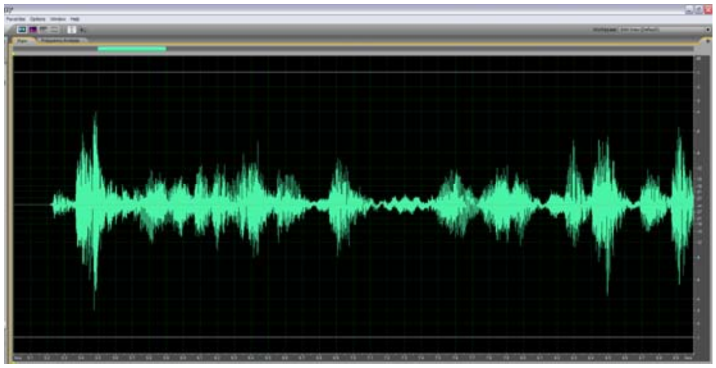

Figure 6 - Inverted Convoluted speech

The pre-recorded speech signal was measured inside the room with the presence of background noise, as shown in Figure 7.

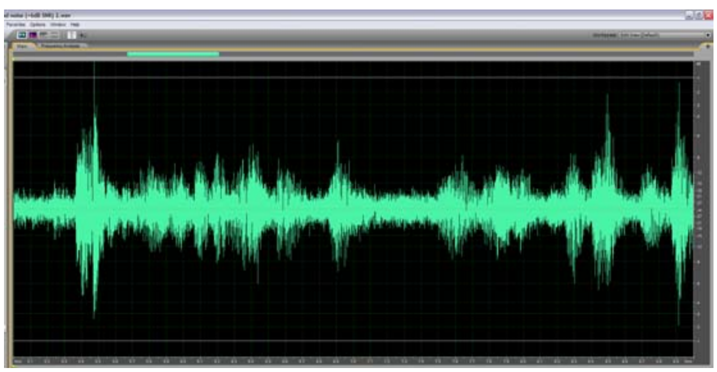

Figure 7 - Measured speech signal in noisy environment

The process of the signal summation results in some degree of reduction of the speech transients with the background noise being the dominant signal. However, elements of the speech content still remain subjectively perceivable above the noise floor.

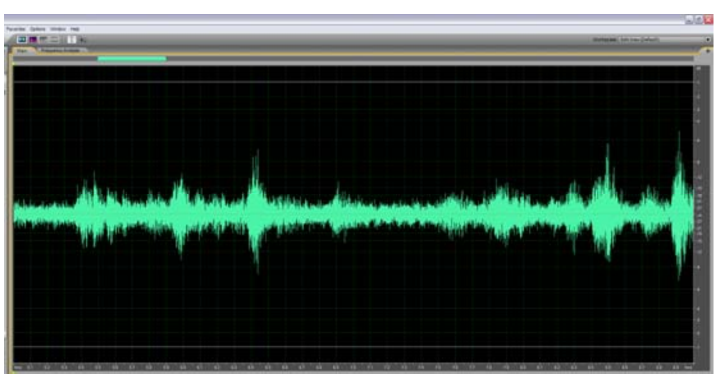

Figure 8 - Resulting output of signal subtraction

4.3 Observations

On the basis of our initial ANS processing simulations implemented using real field measurements, it was noticed that the cancellation achieved was successful towards the beginning of the speech passage but gradually decreased over time. This observation was later verified by re-measuring the signals while introducing a short reference tone at the beginning and end of the speech signals. This process successfully highlighted that distortion in the time domain was taking place throughout the process and that the two opposing signals were eventually out of sync. Although the difference in magnitude was found to be in the order of 10ms over the course of a 20 second audio file, it proved sufficient for the noise separation process to breakdown during playback.
Some of the temporal distortion (2-3ms) was found to take place upon playing the speech signals through the CD player, out of the system, and recording onto the solid state recorder. The remaining and majority of the temporal distortion, was found to be generated during the auralisation process.

In the case of a system that uses pre-recorded messages such as a DVA, the proposed ANS system would not necessarily require the auralisation process to take place. Since the messages are pre-recorded, the signal required to cancel them out could also be pre-established placed onboard the ANS. This scenario would therefore resemble that of the computer simulation process investigated and would be expected to tie up closely with the theoretical model results. An actual recording of each DVA message, made in the space in question, using the ANS microphone is all that would be required for successfully calibrating the ANS system.

5 CONCLUSIONS

Although this paper is limited to simplified noise separation simulations, the preliminary results have successfully demonstrated that the proposed concept is worth the investment of further research and could potentially be viable for real life applications.

For both theoretical and practical models, it was found that the lower frequency content of the speech signals was attenuated most successfully.

In order to maximise the cancellation of the speech signal, this study has proved that the alignment of the anti-phased signals in the time and amplitude domains is crucial and represents the core element of the ANS system.

It has been found that field measurements of pre-recorded signals in-situ can offer great benefits as to the functionality of the ANS system since no auralisation process is involved and hence temporal distortion is significantly reduced.

6 FURTHER RESEARCH

For the system to be able to handle live microphone announcements, the auralisation process would be necessary. This specific element of the ANS functionality would therefore require further investigation in order to gain a better understanding of the actual content of the impulse response, the limitations of the different capture methods, as well as more robust convolution techniques.

Further work would also be focused on the analysis of the spectral content of the noise before and after cancellation. This would allow quantifying and objectively measuring how successfully cancellation can be applied across the frequency range.

Additional measurements are to be carried out using various other RT scenarios in order to observe the degree of contribution from the room's acoustics itself towards the noise separation process. Further measurements would also be repeated using different cases of signal-to-noise ratio.

In order to simulate the effects on a real distributed PA system, the measurement process would be repeated with additional loudspeakers introduced in the space and connected to the signal chain.

7 REFERENCES

1. WinMLS electronic user manual
2. CATT Acoustic user manual
3. Adobe Audition 3 user manual
4. BS EN 5839-8: 2008 Fire Detection and Alarm systems for buildings