SELF-MONITORING AND SELF-OPTIMISING PA SYSTEMS

SJ van Wijngaarden Embedded Acoustics BV, Delft, The Netherlands
JA Verhave Embedded Acoustics BV, Delft, The Netherlands

1 INTRODUCTION

Installing and commissioning public address systems can be a complex job. Especially in challenging environments, such as large, reverberant spaces, the level of speech intelligibility obtained with generic hardware, using default settings, is usually far from satisfactory. It takes a skilled engineer, and often a significant amount of time, to determine the optimal set-up and adjustments for a specific venue. Moreover, the conditions at install time are usually not representative for real use (e.g., an empty stadium versus a full stadium).

This paper outlines a recently developed approach to automatically (instead of manually) optimise PA systems. This approach was developed *only* for PA (or VA) systems intended for speech reproduction, and will not produce optimal settings for music. It is assumed that the system needs to be optimized towards speech intelligibility, not sound quality. The typical application would be a voice alarm / voice evacuation system in a public building.

For such systems, minimum performance requirements are usually set in terms of the Speech Transmission Index (STI)[1]. Commissioning of the system then usually involves STI measurements to deliver proof of compliance. The authors of this paper have often used STI measurements not just to verify compliance, but also to optimize the system. By inspecting measurement results on the level of the Modulation Transfer Function (which underlies the STI), causes of intelligibility degradation can usually be identified objectively and, if possible, reduced.

Over the years, we developed specific strategies that work well for manual optimization. We came to the realisation that these strategies can be converted into formal algorithms, making it possible for computer software to do the optimizations autonomously. This led us to develop the concept of the self-optimising PA: a PA system that incorporates its own STI-based optimisation algorithms, and adjusts its parameters as needed. Such a system is also inherently self-monitoring: it keeps a log of all measured STI data, which provides a complete (and traceable) record of the STI over time. Yet more importantly: every change in the acoustic environment, and even aging of the components of the PA system itself, will be detected and compensated for.

A prototype of a self-monitoring PA system has been developed, and is currently undergoing field tests. This paper broadly describes how this prototype operates.

2 DESCRIPTION OF THE OPTIMISATION PROCEDURE

2.1 Objectives for the STI-based optimisation engine

The following objectives were identified:

- The optimisation engine must, first of all, be capable of measuring the STI, at the time of installation and commissioning as well as during normal operation. All data underlying the STI calculation (such as the Modulation Transfer Function) must be available.

- The engine must be capable of quantifying the most probable causes of intelligibility degradation, including reverberation, echoes, noise and non-linear distortion.
- The engine must be capable of controlling parameters of the PA system to counteract causes of intelligibility degradation. The parameters must (at minimum) include the overall gain, EQ and delay, for each channel/loudspeaker.
- The engine must create a traceable log of all measurement and optimisation data
- The engine must be able to compensate (within the applicable physical limits) for aging and failing of components. E.g., if one loudspeaker fails, the optimisation engine must detect this, and try to compensate using other nearby loudspeakers.
- The engine must be able to track changes of background noise levels and reverberation of time, and compensate for these. E.g., in a shopping mall, the optimum gain and EQ settings will depend on the time of day, since a near-empty mall will show far lower background noise levels than a busy mall during peak hours.

These objectives are to be served while respecting the following boundary conditions:

- For all system parameters, settings chosen by the optimisation engine must fall within a pre-determined envelope of allowable settings. E.g., a limit will be imposed on the overall gain to prevent excessive sound levels, and EQ settings will be restricted to prevent highly irregular settings that might be beneficial to the STI, but that will affect perceived sound quality adversely. Also, the settings will be kept within the specified (technical) limits of the system at all times.
- Artificial test signals may not be played back during normal operation of the PA system.

Based on these objectives and restrictions, it is not difficult to envision what such an optimisation system would look like, physically. Although there are several design choices left, it makes the most sense to implement the optimisation system in a digital signal processor of sorts, which is then integrated into the PA system. Contrary to most “normal” PA systems, a self-optimising system will require the (permanent) presence of multiple measuring microphones.

For the same optimisation system to be effective in a wide variety of scenarios, it needs to have at least three modes of operation:

1. Monitoring. The optimisation engine does not control the system parameters at all, but merely records a traceable log in-situ STI data.
2. Installation and commissioning. The optimisation engine rapidly cycles through multiple optimization iterations, using artificial test signals (STIPA). This is the quickest way to find the optimum settings, but this can only be done if the venue is closed off to the public; otherwise the use of STIPA signals would be highly annoying and disrupt normal operations.
3. Optimizing during normal operation. The STI is assessed based on actual (spoken) announcements, instead of artificial test signals. This is far less efficient, so the optimization data will be gathered at a very low rate (taking days instead of minutes to acquire sufficient “evidence” to base the change of a system parameter on).

We will refer to these three modes as “Mode 1 – 3.”

2.2 Use of speech-based STI

In many ways, Mode 2 (install-time optimisation) is the easiest to implement. This mode is an automated replacement of the optimisation actions currently performed manually by an engineer. For one thing, Mode 2 permits the use of artificial test signals (in particular STIPA) to determine the STI. Usually, there will be a continuous time slot of several hours available for running phase 2 optimisations.

Mode 3 (and also mode 1) rely on real speech to be used as a test signal. The actual announcements themselves can serve as test stimuli. The use of real speech as a test signal is not yet very common, although Steeneken already pioneered speech-based STI in the early 1980s[2].

The major challenge for speech-based STI algorithms is to distinguish modulations in the test stimulus from coincidental modulations present in background noise, or even competing speech. Several researchers proposed correlation-based algorithms to achieve this[3-5]. On alternative approach suggested by Li[6] uses artificial neural networks to extract the STI from running speech.

For our current purposes, correlation-based algorithms are the most suitable approach, since these give physically meaningful results on the level of the m-values. In other words: correlation-based methods of determining a speech-based STI give the same level of details diagnostic information, pointing out *causes* of intelligibility degradation, as a STIPA or even full STI measurement.

The confounding between meaningful and detrimental modulations, which was already mentioned as the prime challenge for speech-based STI methods, can be addressed through the concept of phase-weighting. This concept was implicitly suggested by Drullman et al.[7] and formally developed by Van Gils et al.[8] The concept of phase-weighting is explained further below.

The most general form of computing the speech-based Modulation Transfer Function (and from thereon the STI) is given by Payton[5]:

$$MTF_{payton} \sim \frac{\|crossspectrum\|}{\|autospectrum\|} \tag{1}$$

This uses the cross-spectrum between speech signals at the input and the output of the channel under test, normalised by the modulus of the autospectrum of the intensity envelopes. The MTF is overestimated if uncorrelated speech (or other modulated signals) are present at the channel output.

A way of preventing unrelated modulations from influencing the measurement, is by minding the phase of the cross-spectrum: only if the phase-difference between input and input is small are contributions presumed to be coming from the input signals. This is represented by the following equation

$$MTF_{phaseweighting} \sim \frac{\|crossspectrum\| \cdot f(\angle(crossspectrum))}{\|autospectrum\|} \tag{2}$$

in which $f(\angle(crossspectrum))$ is a function of the phase of the cross-spectrum: the phase-weighting function. Many different forms of this function could be developed that would each have advantages and drawbacks. In practice, we propose simply the following set of functions:

$$w(\phi):\begin{cases}\phi \in (-\pi,\pi] \\ |\phi| \le \pi/2\alpha : w(\phi)=\cos^2(\alpha\phi) \\ |\phi| > \pi/2\alpha : w(\phi)=0\end{cases} \tag{3}$$

Experimentation shows that a value of alpha of 0.5 is a good compromise. If the value of alpha is chosen higher, then the phase-weighting becomes more lenient, and uncorrelated modulations start seeping in. If the value of alpha is chosen lower, then the “penalty” for out-of-phase arrival becomes too high. It should be kept in mind that small “natural” modulation phase shifts may well occur along a transmission path.

Figure 1 shows the phase weighting functions (implicitly) used by the major speech-based STI models.

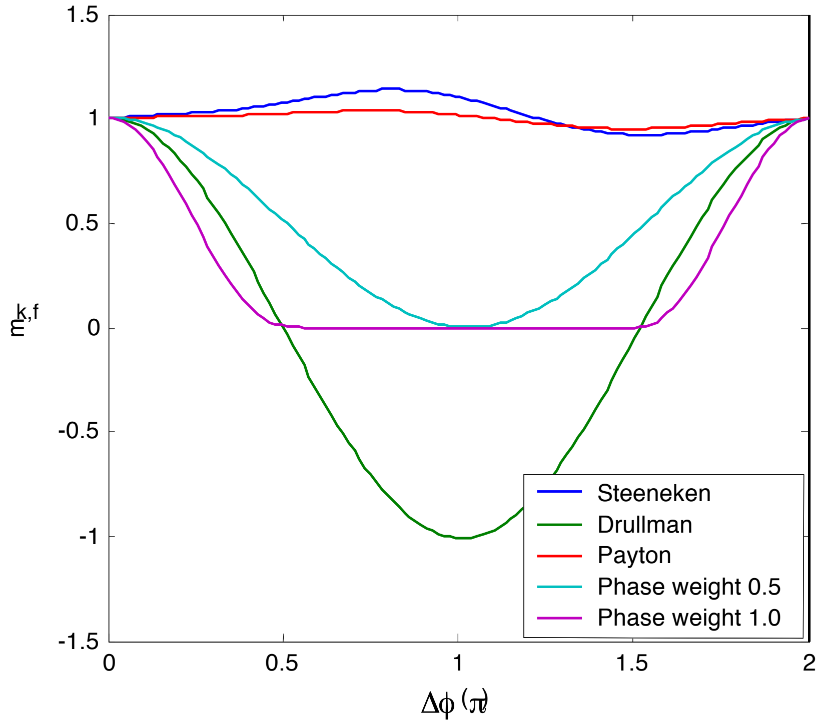

Figure 1. The influence of phase weighting functions of various speech-based STI methods on the MTF, giving modulation index m as of function of the difference between input phase and output phase.

Figure 1 shows that the main difference between the Drullman approach and the phase weighting process as given in equations 2 and 3 is that Drullman effectively penalises out-of-phase modulations, whereas the true phase-weighting approach simply ignores these.

The process as described above, combined with all of the "standard" algorithms for calculating the STI as described in IEC-60168-16[1], are used to determine the STI using any spoken message as a test stimulus.

An additional complication is that a minimum of approx.. 60 seconds of speech is needed to arrive at a sufficiently accurate STI estimate. PA announcements are rarely that long. This is solved by pooling speech from multiple announcements into a single calculation. Data from each announcement is kept; multiple announcements are then combined into a single STI-measurement by fusing the data on the level of the MTF. Similar techniques (which are beyond the scope of this paper) are also used to calculate the STI with fluctuating noise[9].

2.3 Optimisation algorithms

Assuming that sufficiently reliable data has been gathered, the next challenge for the optimisation engine is to determine the optimal system settings from this data. A series of optimisation strategies were (as already used in manual optimisation sessions) were formalised into optimisation algorithms.

In its most general form, the optimisation engine can be described by the block diagram given in figure 2.

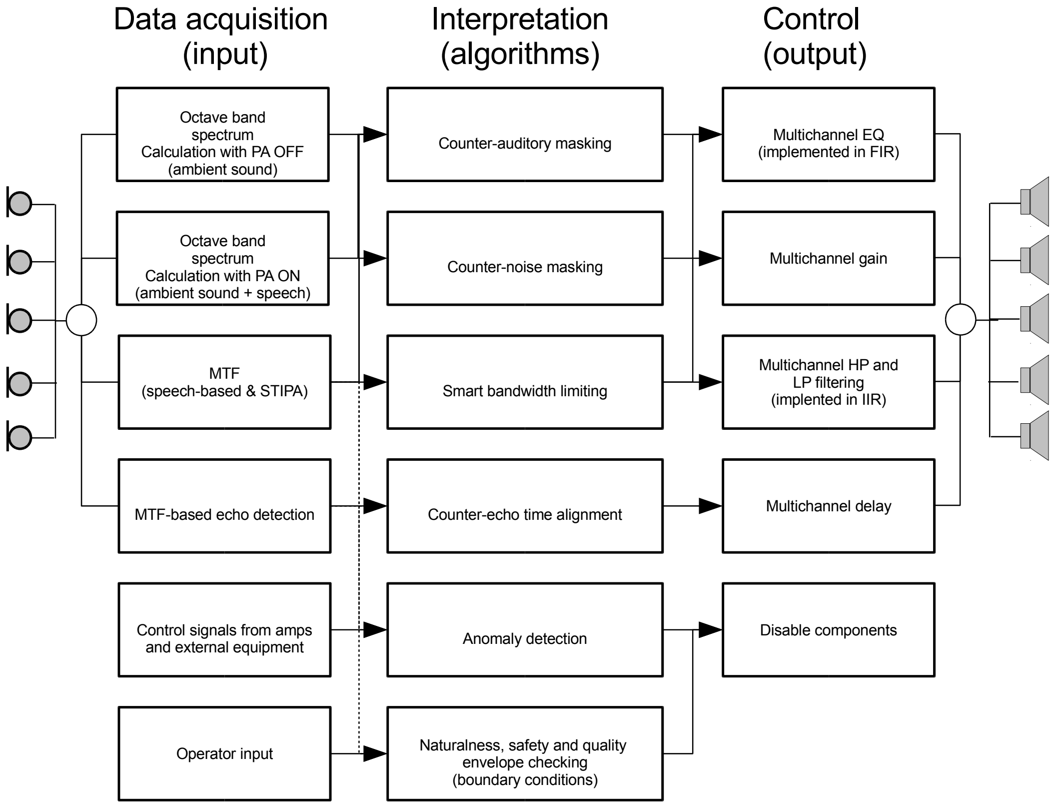

Figure 2. Block diagram of the optimisation engine in its most general form.

Data gathered by the optimisation system as shown in Fig. 2 is not just restricted to typical STI-related data (such as the MTF and octave band spectra). The sense microphones are also used to measure ambient noise in the absence of spoken messages. A dedicated pre-processing step is also taken to detect echoes from the MTF; quite often, “echoes” are induced by poorly time-aligned loudspeakers rather than the acoustic environments. This can be fixed by adjusting delay lines within the PA. Input is also taken from external digital sources, such as amplifier control lines (e.g. used for signalling that clipping occurs).

On all of this (processed) input data, the system runs multiple optimisation algorithms in parallel:

- Counter-auditory masking: the effects of upward spread of masking can be countered by lifting octave bands, iteratively going through all bands from low to high. EQ and gain settings are adjusted.
- Counter-noise masking: the speech signal is shaped for optimal SNR relative to the background noise. EQ and gain settings are adjusted.
- Smart bandwidth limiting: this limits energy spent outside the effective bandwidth of the system (i.e., bands in which the STI contribution is absent or marginal), increasing the headroom for other bands. LP and HP filters are set; possibly gain and EQ as well.
- Counter-echo time-alignment. The system attempts to eliminate echoes by iteratively adjusting delays between channels. If this is successful, than the delay lines are adjusted permanently.
- Anomaly detection: Error signals from equipment are interpreted. Also, results that deviate from the expected (such as earlier measurements obtained around the same time of day) are flagged. Suspect components (usually loudspeakers or amps) are disabled; gains on nearby components are optimised again, this time without the failing equipment being switched on.
- Continuously, all settings are checked against a series of preset boundaries, partly controlled by the operator, partly resulting from the technical limits of the PA system.

3 IMPLEMENTATION OF PA OPTIMISATION

3.1 Hardware

In order to keep our solution as generic as possible, the choice was made to implement the entire optimisation engine in a single 19” DSP. This forms the core of an otherwise conventional PA system. An example of a (generic) configuration is shown in Fig. 3.

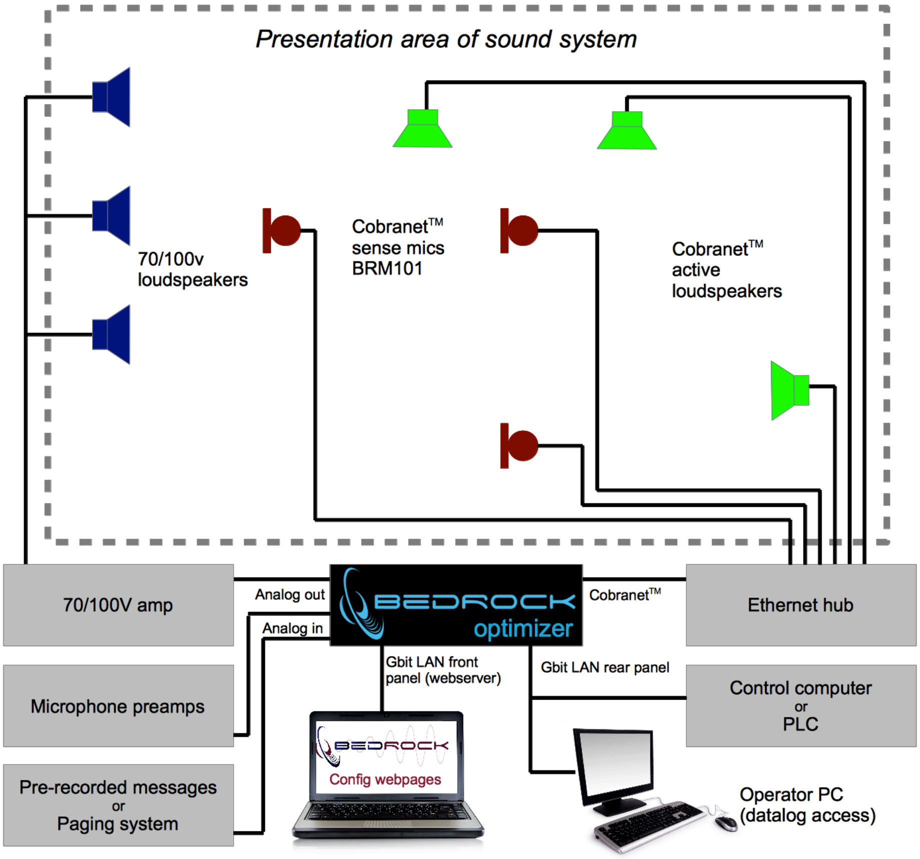

Figure 3. Example of a PA configuration with “Bedrock Optimizer” as the central optimisation hardware, implemented in a 19” enclosure

The hardware used for the prototype optimisation engine is the Bedrock DSP500 platform. The DSP500 combines two sources of processing power:

- The Bedrock Signal Processor 64 (BSP64) board, developed and manufactured by Embedded Acoustics. The BSP64 allows up to 64x64 channels to be processed in parallel on an FPGA-based architecture running at relatively low clock speeds (50 MHz). This is a highly stable audio processing board, designed towards the specifications of EN54, on which all audio DSP functionality runs. The processing capabilities for each channel comprise FIR filtering, IIR filtering, delay and compression.
- An Intel Atom-based industrial (fanless) PC running Windows Embedded Standard 7. This is a very flexible and versatile platform, on which complex algorithms are easily implemented. However, this type of processing platform is not stable enough to be relied on for online processing of audio.

The design is such, that (software) failure of the PC-platform will result in the PC-platform being rebooted automatically. The BSP64 will remain running without interruption, so the functionality of the PA system remains unaffected.

The BSP64 interfaces with the PA system through its analogue interfaces (8x8) and/or its Cobranet interface. The Cobranet interface is also used for the remote sense microphones, which form an essential part of the overall system.

Figure 4. Bedrock Microphone with Cobranet interface (left) and the BSP64 processing board (right), which is the core of the Bedrock DSP500

The sense microphones form a critical link in the entire optimisation chain. During mode 1 measurements, it is feasible to place (temporary) microphones at plausible listener positions. The permanent microphones for mode 2 and mode 3 measurements need to be placed inobtrusively, usually fixed to a wall or ceiling.

3.2 Software environment

The software on the industrial PC-platform consists of a series of WES7 applications, which serve various tasks:

- User interface: the user controls the Bedrock DSP500, including the optimisation engine, through an embedded Webserver. A laptop is simply connected to an RJ45 network connector on the front panel. System settings of the Bedrock system may then be altered by means of a web browser.
- External control interface: a second network connector on the rear panel is used for connecting the DSP500 to other controlling devices. This is useful with PA systems that are controlled through an external controlling system, such as a PLC. A software process monitors communications through this port.
- The optimisation engine itself.

4 PRELIMINARY RESULTS

The first series of trials were run in the Maastunnel, a monumental tunnel in Rotterdam, the Netherlands. Although Voice Evacuation systems are required to be in tunnels by law in the Netherlands, the Maastunnel does not have such a system (yet). Since the tunnel is a monument, and because the tunnel ceiling is relatively low, the usual ceiling-mounted loudspeaker solutions are rejected. Based on conventional horns, the minimum STI-requirement of STI>0.45 are impossible to meet without taking additional optimisation measures.

Using the optimisation approach as described in Fig. 2, the STI was found to exceed 0.45 even during the harshest of noise conditions (rush hour traffic). The end user was especially appreciative of the fact that the system adapts to the noise environment, matching the intensity of the traffic inside the tunnel. An additional benefit from a legislative standpoint is the fact that the system is *optimised* fulfils the criterion that performance is “as good as reasonable achievable.” This is a basis for obtaining a waiver in case the usual minimum STI cannot always be met.

5 CONCLUSIONS AND DISCUSSION

We conclude that the concept of self-optimising PA systems is entirely feasible. An effective and stable optimisation system can be constructed based with a limited amount of (affordable) hardware.

It should be noted that different applications and different types of venue call for a different set of optimisation algorithms. In particular, the boundary conditions will have to be set differently. For instance, optimising the STI (hence optimising intelligibility) to the absolute maximum sometimes leads to settings that are unpleasant to the listener. This decrease in perceptual quality is usually not a concern for evacuation systems that are rarely used, but it is highly undesirable in systems that are also used for (for instance) commercial messages.

This implies that, whereas the required hardware is affordable, there will be a significant expense in tailoring a self-optimising system to the needs of each specific venue. The true benefit to the end user is the assurance that the system will remain performing at its peak, not only at install-time, but throughout the lifecycle of the system, with minimal maintenance efforts.

6 REFERENCES

1. International Electrotechnical Commission. "IEC 60268-16. Sound system equipment-Part 16: Objective rating of speech intelligibility by speech transmission index." Rev. 4, (2011).
2. H. J. M. Steeneken and T. Houtgast, ‘‘The temporal envelope spectrum and its significance in room acoustics,’’ Proceedings of the 11th ICA, 7, Paris, pp. 85–88 (1983).
3. Payton, Karen L., and Louis D. Braida. "A method to determine the speech transmission index from speech waveforms." The Journal of the Acoustical Society of America 106 (1999): 3637.
4. Goldsworthy, Ray L., and Julie E. Greenberg. "Analysis of speech-based speech transmission index methods with implications for nonlinear operations."*The Journal of the Acoustical Society of America* 116 (2004): 3679.
5. Payton, K. L., et al. "Computing the STI using speech as a probe stimulus." In: Past, present and future of the speech transmission index (ed. S.J. van Wijngaarden) (2002): 125-138.
6. Li, F. F., and T. J. Cox. "Speech transmission index from running speech: A neural network approach." The Journal of the Acoustical Society of America 113 (2003): 1999.
7. Drullman, Rob, Joost M. Festen, and Reinier Plomp. "Effect of reducing slow temporal modulations on speech reception." *The Journal of the Acoustical Society of America* 95 (1994): 2670.
8. Van Gils, Bastiaan J., and Sander J. van Wijngaarden. Objective Measurement of the Speech Transmission Quality of Vocoders by Means of the Speech Transmission Index. (2005) TNO Human Factors,
9. Rhebergen, Koenraad S., and Niek J. Versfeld. "A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners." *The Journal of the Acoustical Society of America* 117 (2005): 2181.