Plan B! Measuring Speech Transmission Index using low-cost equipment

Thomas Scelo [1], Tony Wehner [2] and Matt Bissell [2]

(1) Acoustic Research & Design Ltd, Auckland, thomas@acousticsrd.com

(2) Barton Sound Systems, Auckland

ABSTRACT

Public address and voice alarm (PAVA) systems communicate spoken information in public buildings and transport environments, from routine announcements and passenger guidance to instructions during incidents, evacuations and other emergencies. Their value depends on audibility and intelligibility. This is typically assessed using metrics such as Speech Transmission Index (STI) and its public-address variant, STIPA. In New Zealand and Australia, measurements are guided by standards including IEC 60268-16, ISO 3382-1, ISO 7240-19 and AS 1670.4, with formal testing generally requiring calibrated Class 1 instrumentation with appropriate STI/STIPA functions. During installation, tuning and fault-finding, however, preliminary sound pressure level (SPL) checks are often undertaken using low-cost or no-cost instrumentation. To complement these preliminary checks with STI/STIPA checks, measurements were conducted using a calibrated Class 1 sound level meter and a range of low-cost equipment on real PAVA installations in indoor and outdoor environments, using continuous recordings and spot checks aligned with standard test methods. The results confirm that low-cost methods cannot replace compliant instrumentation for certification but that, under certain conditions, they can provide useful and credible estimates.

INTRODUCTION

In many public places, the Public address and voice alarm systems (PAVA) is required where the fire design, evacuation strategy, consent documents, or an alternative solution specifically relies on spoken evacuation messages. Most buildings now days would include a PAVA system to alert people to escape to a safe place.

These systems rely of the clear transmission of information to the public, quantified by parameters such as the Speech Transmission Index (STI), its PA alternative (STIPA), Common Intelligibility Scale (CIS), and others. Measurements of STI or STIPA generally need to follow a range of standards for methodology, number of measurement points, methods to feed signals into the system and nature of the system.

In New Zealand and Australia, the commonly cited standards include IEC 60268-16, ISO 3382-1, ISO 7240-19, AS 1670.4, EN 61672-1. These generally require the use of a Class 1 sound level meter equipped with STI or STIPA measurement function. Multiple sound level meters with STI and STIPA measurement function are readily available. These systems require a significant capital investment, regular calibration, and maintenance.

A survey of most suppliers has shown that even the purchase of the additional software to install on commonly owned class 1 sound level meters represent an investment of several thousand dollars. During a 2026 measurement campaign, none were available for hire in the country.

Table 1. Examples of IEC 61672 Class 1 sound level meters with STI/STIPA function on board.

Manufacturer	Instrument	Comment
Norsonic	Nor145 / Nor150	Supports IEC 60268-16 Ed.5. STIPA directly and full STI via impulse response/SweptSine.
NTi Audio	XL2 / XL3	STIPA as option supporting IEC 60268-16:2020 Ed.5, including ambient-noise correction and repeated-measurement averaging.
Svantek	SVAN 979	STIPA as option supporting IEC 60268-16:2020 Ed.5. Also includes RT60, 1/3-octave, FFT, WAV recording and signal generation.
Bedrock Audio	SM90	STIPA and full STI direct-method measurements included supporting IEC 61672. Also includes RT60, FFT and octave analysis.
Brüel & Kjær Rion Acoem/01dB CESVA	None identified	No STI/STIPA function identified.

In real life, a very time-consuming part of the delivery of large PAVA systems is the calibration and setup. This requires preliminary checks, settings and rechecks. (starting with SPL only) and often carried out by the installer with a low-cost (or sometimes no-cost)

instrumentation (Class 2 from a hardware store). The lack of STI information from these simple preliminary surveys often results in surprises when an expert consultant is invited to test the system, this time including the STI. The back-and-forth process is often repeated several times, resulting in delays and high costs to the project. If only these low-cost Class 2 sound level meters could give an estimate of STI alongside the SPL. Once set and equipped with a STIPA test signal, the project could move along more efficiently.

The question explored here is therefore simple: if the standard STIPA signal is recorded along a simple survey, how close can post-processed results come to a reference measurement made with calibrated Class 1 STI/STIPA meter? And is that accuracy (whatever it might be) useful during the delivery of a project?

Two related applications were investigated. The first compares point measurements from several recorders with the XL2 reference. The second uses one long recording made while walking along a train platform, from which a spatial STIPA profile is calculated using a sliding analysis window.

METHOD

Recording systems

The following recording systems have been used:

- NTi Audio XL2 (XL2) using its calibrated STIPA function. This is used as the reference system against which the others are compared
- Brüel & Kjær 2250 (BK2250) with BZ7726 signal recording option
- Zoom H3-VR ambisonics microphone (H3VR)
- TASCAM DR-07XP stereo microphone (TASCAM)
- iPhone with Sennheiser XS Lavalier microphone (LAV).

Microphone types

The pieces of equipment listed above do not all measure the same acoustic field. The XL2 and B&K 2250 use conventional omnidirectional measurement microphones. The TASCAM DR-07XP has two directional stereo capsules; converting, selecting or averaging its channels produces a mono signal, but does not make the microphone omnidirectional. The H3-VR was processed using its W channel, the derived omnidirectional component of the Ambisonic recording. The smartphone with the Lavalier microphone could be considered a commercial grade omnidirectional signal.

The differences between the microphones matter because a directional microphone receives a different mixture of direct sound from the PAVA loudspeakers, reflections, diffuse reverberant sound field, even the background noise and capsule/wind induced noise. A simple frequency-response correction cannot reconstruct the sound field that an omnidirectional reference microphone would have received.

The objective of the study was not to turn the more common recorders into calibrated gear, but to see if at an engineering level they could provide useful information to the main contractor and installer/supplier to move forward. For this reason, the core STIPA calculation in the processing of the recordings was kept unchanged and only the selection of the channel, when available, was considered to best replicate the omnidirectional microphone case.

Point measurements and post processing

All results presented here were collected during initial testing phases and do not represent final commissioning results or the final acoustic performance of the projects.

Where corresponding reference and recorder data were available, the recorded WAV file was converted to a single channel and processed using the same direct-STIPA routine. The routine band-filters the received signal into the seven STIPA octave bands, detects the intensity envelope, derives the modulation transfer values and combines them into STI. The reference value for comparison is the NTi XL2 result obtained at the corresponding measurement point.

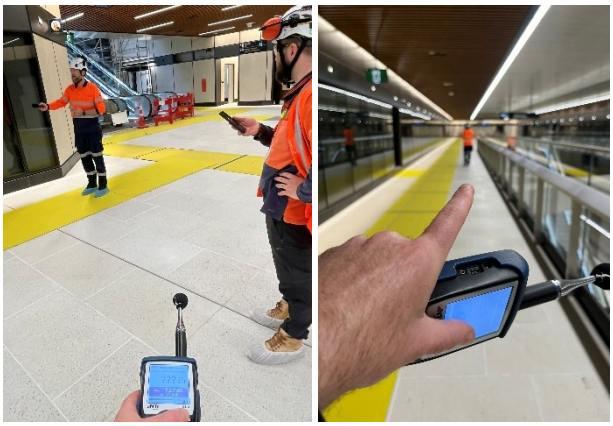


Figure 1. Example of site survey using XL2 and the TASCAM DR-07XP recorder

A total of 10 different spaces has been used for the measurements. These include train platforms (indoor and outdoor), large public lobbies, tall concourse, meeting room and a community hall.

Not all instruments were available for all these measurements. However, a total of 218 measurement points was gathered and is considered sufficient for the investigation, as detailed in Table 2.

Table 2. Number of measurements per instrument.

Instrument	Number of point measurements
XL2	217
BK2250	70
TASCAM	172
H3VR	27
LAV	27

Two statistics have been considered in this study:

- The mean STI over the ensemble of measurement points in each space. This indicates whether a

recorder reproduces the average performance of a surveyed area (often the core element of the project specifications). If correlation exists, then the more detail 2nd statistical parameter can be considered.

- The mean absolute difference against the XL2. This is readily visible in a 1-to-1 graph and indicates the typical point-by-point disagreement. This can be used for finer determination on whether the recorder can be used for more complex specification requirements; for example strict minimum STI value, or percentage of the space with STI above a specified threshold.

POINT-MEASUREMENT RESULTS

First campaign

The first campaign was conducted over a range of projects, conditions, systems and operators. Not all measurements were conducted by the same engineer and in similar conditions. Outdoor platforms could be very windy, while the large meeting room had an uncharacteristic high background noise.

Table 3 shows the statistical results for mean STI, minimum STI at any location, and percentage of the area with a measured STI over 0.50. These parameters typically feature in project specifications.

Figure 2 shows a 1-on-1 graph including a set of data points for the XL2, BK2250, and TASCAM. These have included 9 different spaces (all but the Community Hall).

Across all combined point data in this initial campaign, the STI averages were almost identical between the XL2, TASCAM and BK2250. There is hope that a simple survey with a common equipment could return useful information to the project.

However, it is noted that agreement of the mean STI value does not mean agreement at individual positions, as shown by the scatter of the TASCAM points in Figure 2. This recorder also exhibits several significant outliers, and a large proportion of points falls within the +/-10dB range (orange zone) from the XL2.

In comparison, the recordings from the BK2250 reproduce the XL2 more consistently with nearly all points falling within the +/-5dB range (green zone) from the XL2 data.

Table 3. Statistical comparison of STIPA processed from STIPA recordings in 5 different spaces.

Space	Parameter	XL2	TASCAM	BK2250
Outdoor train platform, very windy	Mean STI	0.69	0.70	0.72
	Min STI	0.60	0.60	0.67
	% > 0.50	100%	100%	100%
Semi-outdoor train platform, windy	Mean STI	0.54	0.53	0.53
	Min STI	0.44	0.37	0.40
	% > 0.50	72%	63%	62%
Large Meeting Room	Mean STI	0.61	0.62	0.59
	Min STI	0.54	0.53	0.56
	% > 0.50	100%	100%	100%
Tall Indoor Lobby	Mean STI	0.56	0.59	0.59
	Min STI	0.50	0.52	0.57
	% > 0.50	78%	100%	100%
Semi Indoor Long Lobby	Mean STI	0.57	0.59	0.58
	Min STI	0.50	0.56	0.50
	% > 0.50	88%	100%	100%

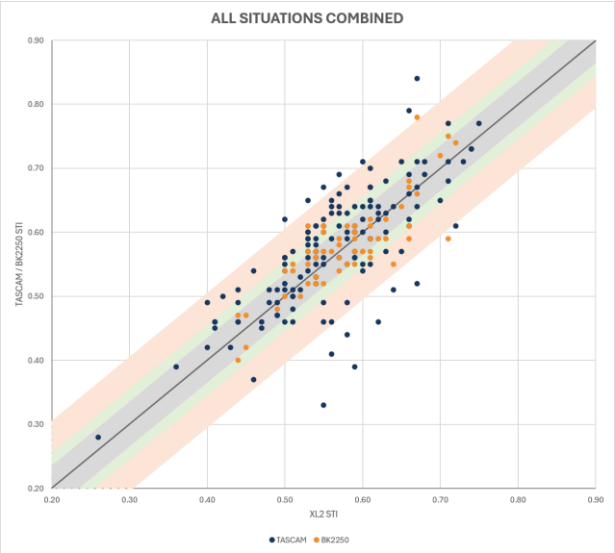


Figure 2. One-to-one graph for the XL2, TASCAM, and BK2250 for 9 out of the 10 spaces tested

Surprisingly, the TASCAM was found to better match the XL2 mean data in noisy and outdoor conditions, and exhibit greater differences in more control spaces such as the Meeting Room and indoor lobbies. This is believe to be due to the fact that the differences between microphone "qualities" becomes less significant in low signal-to-noise situations due to high background noise and wind induced noise.

Second campaign

A second campaign aimed at reducing the number of variables in the measurement conditions (background noise, wind, reverberation and geometry) and introducing more recording instruments. This time, a community hall

in Auckland was hired and used as a control space. All 5 instruments were used simultaneously on a single tripod and subjected to the same background noise conditions 36dBA (no HVAC running) and no wind.



Figure 3. Controlled environment testing with 5 different instruments.

The STIPA signal was fed to a single loudspeaker at a fix location and 27 measurement positions were established in the 23m x 12m x 5.5m space. Curtains were drawn over the windows to control the reverberation time to T_{20} 1.04s.

The results of this new campaign are summarised in the Table 4 and Figure 4 below.

In addition to the STI related values, the table include the mean absolute difference (MAD) between the data from each recorder and the data from the XL2 reference instrument.

Table 3. Statistical comparison of STIPA processed from STIPA recordings in 1 space at 27 locations.

Parameter	XL2	TASCAM	BK2250	H3VR	LAV
Mean STI	0.70	0.75	0.77	0.71	0.73
Min STI	0.63	0.66	0.68	0.63	0.68
% > 0.50	100%	100%	100%	100%	100%
MAD	-	0.049	0.033	0.023	0.024
Within +/-0.05 STI	-	52%	91%	96%	93%
Bias	-	+0.05	+0.025	+0.008	+0.021

The data in Table 4 and Figure 4 confirm that in a controlled environment with low reverberation, low background noise and no wind, all instruments have a positive bias (overestimating) relative to the XL2 data.

The worst case being the TASCAM with 52% of the data within 5% of the XL2 measured STI. The best case being the H3VR with 96% of the data within 5% and a negligible bias.

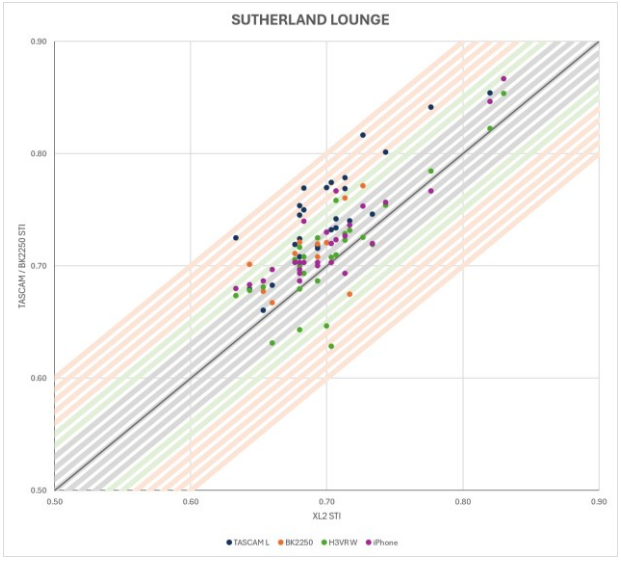


Figure 4. One-to-one graph for the XL2, TASCAM, BK2250, H3VR, and LAV in the Community Hall

Is this good enough for a survey?

One conclusion is that low-cost recordings are not interchangeable with a compliant STI meter. However, they can answer a different question during the pre-commissioning process: has the recent change to the control system moved the system broadly in the right direction? How far are we from the target and what are the risks of calling for a formal commissioning at this stage.

Conversely, recorded measurements should not be relied upon to decide formal compliance at individual points, especially if a coverage criterion, expressed as percentage of the measurement points above a threshold, is used. The number of outliers observed in real-life conditions as well as in a controlled environment is simply too great.

CONTINUOUS RECORDING

The point-measurement comparison led to a second use of recorded audio: replacing a sequence of stationary stops with one continuous recording while walking at approximately constant speed within the space under test.

Each STIPA result at any given time of the recording is then assigned to the position corresponding to the centre of its analysis window, producing a spatial profile, STIPA(x), rather than a set of isolated points.

A 15s analysis window was adopted to reflect the minimum standard STI measurement duration. The calculation step can be equal to the window duration, giving adjacent non-overlapping results, or shorter to create a denser profile. A shorter step does not provide finer fundamental spatial resolution: every value still integrates the acoustic conditions encountered during the same 15s window. At walking speed v, that window represents a finite distance of approximately $15 \times v$ metres.

Figure 5 below shows how each STIPA value is calculated from a finite analysis segment centred at the evaluation time; with approximately constant walking speed, time can then be mapped to distance along the platform.

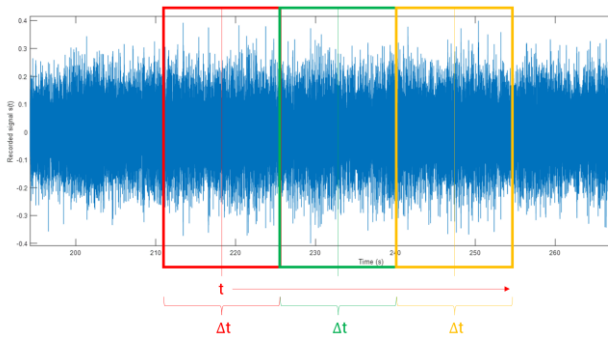


Figure 5. Sliding-window concept for a continuous recording.

Overlap is therefore mainly a presentation and localisation tool. With a 7.5s step, adjacent results share 50% of their audio; with a 1.0s step they share more than 90%. Such results are useful for plotting a smooth spatial function, but adjacent values should not be treated as independent observations. They however may represent better the fluctuations heard in the STIPA signal while walking through the path.

Figure 6 shows how smaller steps increase sampling density and make local fluctuation easier to identify. However, there is no effect on the overall mean STI value.

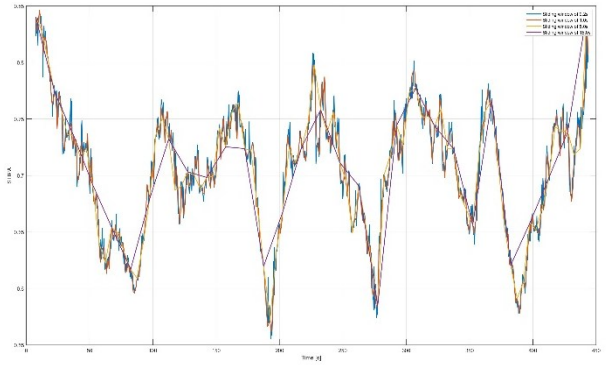


Figure 6. Effect of calculation step for the same 15s analysis window.

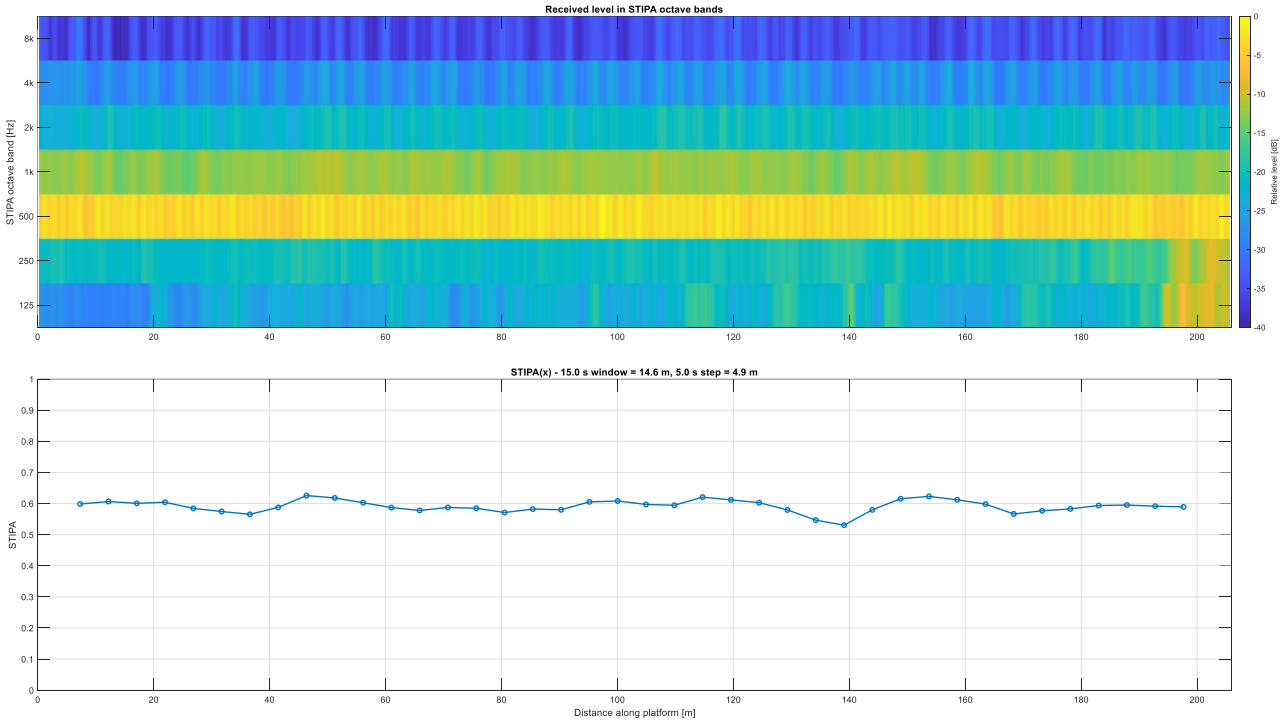


Figure 7. Top: spectrogram of the entire walk. Bottom: continuous STIPA(x) measurement during a steady-speed walk along the platform as a function of the location.

Is a continuous measurement helpful

Continuous recording has shown to provide useful information for long and linear spaces (platforms, concourses). It can cover the full length of the space rapidly while retaining the complete audio record for later analysis. It can reveal the shape and location of broad STI troughs that may be missed by the standard sparce measurement locations. This is particularly attractive during PAVA tuning, when several iterations may otherwise require repeated stop-and-measure surveys, lasting much longer.

However, a moving measurement is not equivalent to a stationary direct-STIPA measurement at the centre point as dictated by standards and project specifications. Each 15s value integrates a changing sound field over a finite distance; it is a broad brush.

In addition, wind and background noise as well as microphones directivity and sensitivity affect low-cost recorders more strongly than measurement microphones. For this reason, STIPA(x) is best interpreted as a screening profile that directs attention to areas requiring

adjustment or subsequent compliant measurement, rather than as a replacement for the formal survey.

In terms of usefulness, the distinction between mean STI survey accuracy and point accuracy is central to how the low-cost method should be used.

If the survey is to judge the general outcome of a gain or equalisation change, a recorder that reproduces the survey mean STI within a known and understood percentage may be entirely adequate. If the survey is to determine whether a particular location passes a criterion such as STI 0.55, the same error may be unacceptable. The method is more robust at the stage of the project when decisions are based on trends, averages, and identifications of poor areas.

A contractor-oriented workflow follows naturally. The installer can make rapid SPL and recorded-STIPA checks during adjustment, repeat the process after gain, EQ or fault corrections, and use the resulting point data or STIPA(x) profile to identify areas that remain suspect. The expert survey can then concentrate on formal measurement, verification of coverage and investigation of borderline locations. This does not reduce the need for compliant commissioning; it can reduce the amount of expert time spent discovering issues that could have been identified earlier.

For recorded-audio surveys to be repeatable, the recording chain should also be treated consistently. The selected channel and microphone configuration should be documented, automatic gain control or limiting should be avoided where possible, wind protection should be used for walking measurements, and the original WAV files should be retained. Keeping the raw recording is a particular advantage of the method because alternative channel choices, window lengths or quality checks can be evaluated later without repeating the site survey.

CONCLUSIONS

The following conclusions have been drawn from the study:

- Low-cost recording of STIPA signals and post-processing was shown to provide reasonable estimates of the mean STI value.
- Point-by-point agreement is much weaker than mean agreement and will highly depend on the recorder and the microphones it uses. Audio equipment is surprisingly better than low-cost consumer recorders; it is a balance between cost, user expertise, and required accuracy.
- Recorded-audio STIPA is therefore credible as a contractor tool for tuning, screening, identifying spatial trends and deciding where and when expert testing is required. This method is not suitable for formal commissioning or borderline pass/fail situations.
- Continuous walking recordings extend the method from discrete points to a spatial STIPA(x) profile that can be used to identify areas of concerns, understand how local noise sources and reverberation/architecture interfere with STI.

REFERENCES

[1] IEC 60268-16:2020, Sound system equipment - Objective rating of speech intelligibility by speech transmission index.

[2] ISO 7240-19, Fire detection and alarm systems - Design, installation, commissioning and service of sound systems for emergency purposes.

[3] AS 1670.4, Fire detection, warning, control and intercom systems - System design, installation and commissioning - Emergency warning and intercom systems.

[4] P. Zaviska, Brno University of Technology, MATLAB direct-STIPA implementation used for post-processing (2023-2024).