
ON
SPEECH INTELLIGIBILITY & P.A. SYSTEMS

Rolins Thomas Roy
B.Arch, M.Sc Acous(LDN). MIOA

Introduction

The objective of the experiment.

Apparatus used.

-Speech intelligibility & its importance in P.A systems.

- What affects intelligibility?

- 'Machine measure' methods of intelligibility

(emphasizing more on the method used in this experiment: STI-PA)

Procedure :

How the experiment was conducted.

Step by Step explanation of the method of 'Machine Measure' of Speech Intelligibility conducted.

Results, Analysis & Discussion

Conclusion

An overall interpretation of the results obtained.

Assessment of the speakers & the environments.

General inferences

References

There's an important difference between music and speech. The brain is capable of “filling in” a fair amount of missing information in music, because there's a high degree of predictability (generally while hearing music and if you didn't get the bass line or some part of the song which you are keen on listening in the first four measures, you'll pick it up when it repeats in the next four beats) **But speech is rich in constantly-changing information.**

At large distances between a talker and listener, intelligible communication is difficult. If in an enclosed reverberant space, the reverberant sound would mask the speech syllables since the direct sound would be weak and the reverberant sound dominant. As the talker and listener move closer together, then the direct sound increases and speech communication improves. If even a modest percentage of the information is jumbled or missing, the brain can't decipher the message.

So this experiment was conducted to obtain a practical knowledge of speech intelligibility & also to gain experience of setting up a public address sound system.

Various speaker systems were assessed and their respective intelligibility (STI-PA scores) were noted to compare with each other; the environments or rooms in which these tests were performed were also assessed. Hence the capability of the space to accommodate a good sense of speech intelligibility could also be judged with the measured data and conclusions.

'Speech intelligibility' & its importance in P.A (Public Address) systems.

Intelligibility could be defined as the degree to which speech can be understood. With specific reference to speech communication system specification and testing, intelligibility denotes the extent to which trained listeners can identify words or phrases that are spoken by trained talkers and transmitted to the listeners via the communication system.

Public address systems in building complexes have to inform persons about escape directions in case of emergency. Such public buildings include airports, railway stations, shopping centres or concert halls. However if such announcements are misunderstood due to poor system quality, tragic consequences may result. Therefore, it is essential to design, install and verify sound reinforcement systems properly for intelligibility. In addition, a variety of other applications such as legal and medical applications may require intelligibility verification. Speech communication systems (Public Address Systems) therefore are subject to **more stringent requirements** than music systems.

“Human speech is a continuous waveform with a fundamental frequency in the range of 100-400 Hz. (The average is about 100 Hz for men and 200 Hz for women.) At integer multiples of the fundamental are a series of changing harmonics called “formants” which are determined by the resonant characteristics of the vocal tract. Formants create the various vowel sounds and transitions among them. Consonant sounds, which are impulsive and/or noisy, occur in the range of 2 kHz to about 9 kHz.

The sound power in speech is carried by the vowels, which average from 30 to 300 milliseconds in duration. Intelligibility is imparted chiefly by the consonants, which average from 10 to 100 milliseconds in duration and may be as much as 27 dB lower in amplitude than the vowels. The strength of the speech signal varies as a whole, and the strength of individual frequency ranges varies with respect to the others as the formants change.”¹

(In **Fig.1** is a vocal spectrum graph for male and female speakers with an “idealized” human vocal spectrum superimposed)

The listener's challenge is to analyze speech sounds into meaningful units of language - a complicated task. Gaps in the sound don't necessarily correspond to word or syllable breaks.

¹ Section-1 : 'Speech Intelligibility Papers'- Written by Ralph Jones. Edited by Rachel Murray P.E
<http://www.meyersound.com/support/papers/speech/>

Speech sounds also are not discrete events: rather, they merge and overlap in time, and the articulation of a given phoneme differs in different contexts and with different speakers.

In fact, the precise ways in which the ear-brain mechanism decodes speech remain something of a mystery. Such **factors as loudness, duration and spectral content** certainly affect speech perception, but how they may interact is not fully understood.

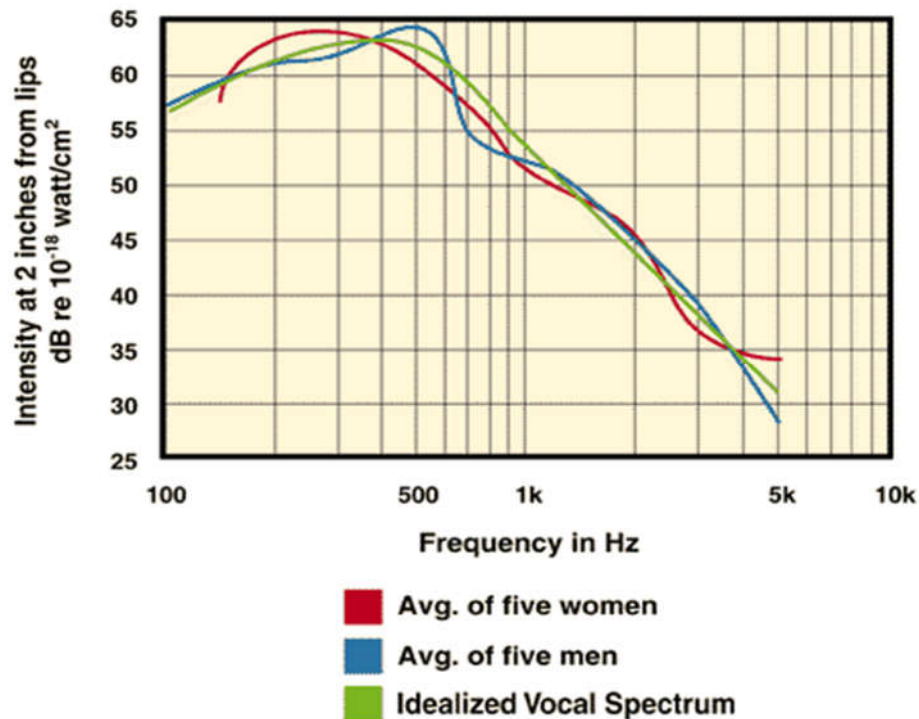


Fig. 1: Vocal spectrum graph for male and female speakers with an “idealized” human vocal spectrum superimposed ²

Diminished intelligibility is associated with a loss of information that is coded in a number of highly interactive elements, and many factors influence it. Background noises can mask the speech. Both the **direction of the source, relative to the listener, and the direction of the interfering noise can alter the degree of masking**. Intelligibility is also affected by the **predictability of the message, the speaker's accent/pronunciation and, not least, the sharpness of the listener's hearing**.

Factors That Affect Intelligibility in Sound Systems

The goal of a speech reinforcement system is to deliver the speaking voice to listeners with sufficient clarity to be understood. Given the complexity of the speech signal, the task of providing high-quality speech reinforcement in real-world, less-than-ideal conditions is doubly complicated.

Masking

The most common obstacle that speech system designers face is the intrusion of unwanted sounds that inevitably interfere with the speech signal. The effect is called “masking,” – a general term that covers a very wide variety of situations.

² French, N. R. and Steinberg, J. C. “Factors Governing the Intelligibility of Speech Sounds,” JASA vol. 19, no. 1 (1947)

Masking noise can come from acoustical sources such as ventilation equipment, traffic, crowds and commonly, reverberation and echoes. It can also arise electronically from thermal noise, tape hiss or distortion products. If the sound system has unusually large peaks in its frequency response, the speech signal can even end up masking itself. One relationship between **the strength of the speech signal and the masking sound** is called the **signal-to-noise ratio expressed in decibels**. Ideally, the S/N ratio is greater than 0dB, indicating that the speech is louder than the noise. Just how much louder the speech needs to be in order to be understood varies with, among other things, the type and spectral content of the masking noise.

So we could define it as the ratio between the strength of the desired speech signal and that of introduced noise, expressed in decibels. At 0 dB the two are of equal strength; negative values are associated with loss of intelligibility due to masking. Positive values are usually associated with better intelligibility.

“The most uniformly effective mask is broadband noise. Although, narrow-band noise is less effective at masking speech than broadband noise, the degree of masking varies with frequency. High-frequency noise masks only the consonants, and its effectiveness as a mask decreases as the noise gets louder. But low-frequency noise is a much more effective mask when the noise is louder than the speech signal, and at high sound pressure levels it masks both vowels and consonants”³.

The direction, from which a masking sound arrives, relative to the direction of the speech signal, can affect the degree of masking. If the noise comes from the same place, the masking is greatest; it decreases as the distance between the noise and the speech increases because this makes it easier for the brain to discriminate between them. The masking effect is lowest when the presentation is through headphones, with the speech in one ear and the mask in the other. (Unfortunately, we can’t take advantage of that feature in sound reinforcement).

Hence we see that reverberation is so destructive of intelligibility, especially beyond critical distance. Being itself caused by the speech, reverb mimics the speech spectrum, but generally with greater low-frequency energy. Sufficiently long reverb and echoes such as are encountered in cathedrals and large sports arenas can actually function like multiple distractor voices. And by its nature, reverberant energy arrives from all angles, so it’s hard to separate from the speech using directional clues.

Machine Measure methods of Speech Intelligibility

Statistical tests using trained talkers and listeners are by far the most accurate and reliable methods for intelligibility testing. Unfortunately, they are complicated to set up, time-consuming to conduct and require extensive statistical analysis to interpret. Hence, consultants and acousticians have long sought an automated, machine-based test that could quickly and easily yield meaningful intelligibility scores for speech systems.

Nowadays, highly developed algorithms as SII (Speech Intelligibility Index) and various forms of the STI (Speech Transmission Index) allow measuring speech intelligibility. These techniques take care of many parameters which are important for intelligibility such as:

- Speech level
- Background noise level
- Reflections
- Reverberation
- Psychoacoustic effects (masking effects)

³ Section-2 : ‘Speech Intelligibility Papers’- Written by Ralph Jones. Edited by Rachel Murray P.E
<http://www.meyersound.com/support/papers/speech/>

In STI testing, speech is modelled by a special test signal with speech-like characteristics. Following on the concept that speech can be described as a fundamental waveform that is modulated by low-frequency signals, STI employs a complex amplitude modulation scheme to generate its test signal. The basic idea of STI measurement consists in emitting a synthesized test signals instead of a human speaker's voice.

The speech intelligibility measurement acquires and evaluates this signal as perceived by the listeners ear. At the receiving end of the communication system, the depth of modulation of the received signal is compared with that of the test signal in each of a number of frequency bands. Reductions in the modulation depth are associated with loss of intelligibility the Speech Transmission Index (STI) is a machine measure of intelligibility whose value varies from 0 (completely unintelligible) to 1 (perfect intelligibility).

STI is derived from the Modulation Transfer Function (MTF) in a room. MTF is calculated from a noise signal 125 Hz to 8 kHz octave bands with modulation frequencies between 0.63 Hz and 12.5 Hz (14 frequencies*7 octaves=98)

The MTF concept was proposed by Houtgast and Steeneken to account for the relationship between the transfer function in an enclosure in terms of input and output signal envelopes and the characteristics of the enclosure such as reverberation. This concept was introduced as a measure in room acoustics for assessing the effect of the enclosure on speech intelligibility.

To calculate STI :

$$(S / N) = 10 \lg \left(\frac{m(F)}{1 - m(F)} \right)$$

Where A weighting factor for each of the 7 octave bands is applied based on a standard speech spectrum, calculated from subjective testing (0.13, 0.14, 0.11, 0.12, 0.19, 0.17, 0.14) for 125 Hz to 8 kHz

$$\overline{(S / N)} = \sum_{oct} W_{oct} (S / N)_{oct}$$

Finally, the weighted mean signal to noise ratio is converted to STI giving a value between 0 and 1, 1 indicates perfect intelligibility.

$$STI = \frac{[(S / N) + 15]}{30}$$

<u>STI Range</u>	<u>Quality Rating</u>
>0.80	Excellent
>0.65	V. Good
>0.50	Good
>0.40	Fair
>0.30	Poor
0.30	Bad

Table 0.0 - STI- quality rating table

“A rising awareness for security issues, new technological means and the shortcomings of RASTI triggered the speaker manufacturer Bose and the research institute TNO to develop a new method for speech intelligibility measurements of PA installations. The result of these efforts is **STI-PA**, which allows quick and accurate tests with portable instruments. Like RASTI, STI-PA applies a simplified procedure to calculate the MTF. But STI-PA determines one MTF by analyzing all seven frequency bands, whereby each band is modulated with two frequencies.

Supposing that no severe impulsive background noise is present and that no massive non-linear distortions occur, **STI-PA** provides results as accurate as **STI**. If however impulsive background noise is present during the normal system operation hours, it is usually possible to mitigate the effects by also acquiring a measurement at a more favourable time e.g. under slightly different conditions in the area, or during the night time and to calculate an unbiased overall measurement by using the results of both test cycles.”⁴

A simplification can be applied to the test signal if the uncorrelated (speech-like) modulations, required for the correct interpretation of non-linear distortions, are omitted. This opens up the possibility of modulating and parallel processing of all frequency bands simultaneously, thus reducing measuring time. For each frequency band the modulation transfer is determined for two modulation frequencies. The STIPA method employs this simplification and takes 10 s to 15 s for a measurement (typically 12s).⁵

Instead of the 14 modulation frequencies applied to all seven octave bands as is the procedure for the full STI, the STIPA method applies, uniquely, to 12 modulation frequencies.⁶

But the unavoidable truth is that, as sophisticated as machine-based measurement systems may be, they cannot yet approach the complexity of the human ear/brain mechanism informed by a lifetime of experience decoding speech. We can only model those aspects of that exquisitely fine-tuned mechanism that we have come to understand.

For the procedure, 3 environments were chosen:

A normal lab room, a reverberant chamber and an anechoic chamber.

In each of the below environments, a class 1 sound level meter and a laptop with soundcard to send the synthesized test signals to the signal source through the power amplifier (Nor 280) with Cables to connect the signal source to the power amplifier & to the laptop was used.

In Lab room:

1. Nor 275 Speaker (Hemi-Dodecahedron) Speakers
2. Tivoli speakers

In Reverberation Chamber:

1. Balloon (for the RT measurement of the chamber by Impulse noise method)
2. Nor 275 Speaker (Hemi-Dodecahedron)
3. Tivoli speakers

In Anechoic chamber:

1. Yamaha powered monitor speaker model HS 50M
2. Tivoli speakers

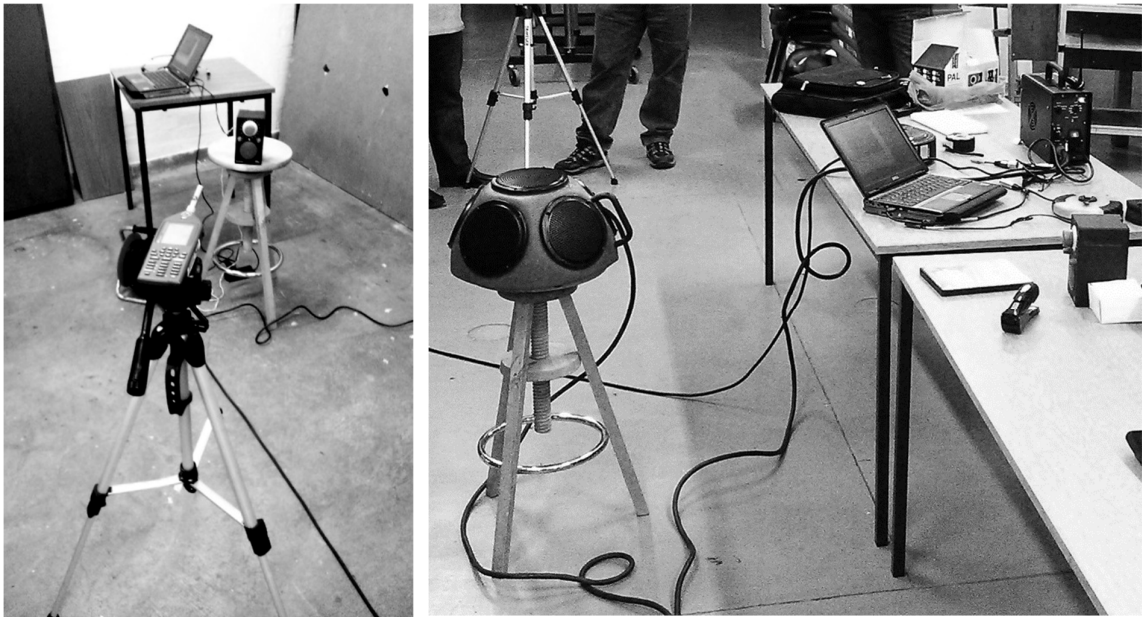
⁴ Introducing Speech Intelligibility

http://www.ntiaudio.com/Portals/0/Products/Minstruments/AL1/AppNotes/NTI_App_Note_Introducing_STI-PA.pdf

⁵ BS EN 60268-16:2003 - 4.4 : Sound system equipment - *Part 16: Objective rating of speech intelligibility by speech transmission index*

⁶ BS EN 60268-16:2003 - Annex -C : Sound system equipment - *Part 16: Objective rating of speech intelligibility by speech transmission index*

Procedure :



In the Lab room:

The apparatus as mentioned was set up in the Acoustics lab room.

The background noise of the room was measured.

The synthesized signal was played through the Nor 275 speaker (Hemi-Dodecahedron) which acts as a multidirectional signal source of sound. The SLM Nor 140 was used for the measurement of STI-PA- It was placed on a Tripod at a distance of about 3m from the signal source for its first measurement. The sound signal was generated and controlled from the laptop and the SPL & STI was noted down. The time set for the measurement was 12 seconds. This was repeated at distances of 1m(close distance measurement) & 9m(long distance measurement-far) from the sound signal source and measurements taken down respectively.

In the Reverberation Chamber

The apparatus was then set up in the reverberation chamber. This time the balloon burst method was carried out before the measurement

The Nor 275 hemi-dodec was placed at one corner of the rev.room and the SLM nor 140 on the opposite far end of the rev. room(@10m). The signal source was generated and the readings taken down. The closer distance measurements(close & medium) were carried out at almost $2/3^{\text{rd}}$ & $1/3^{\text{rd}}$ distances of the long distance measurement.(i.e @ 1m & 3m respectively)-This was repeated with the Tivoli speakers too.

Then the next stage involved opening the doors to the 10 m² area absorptive surface wall of the reverberation chamber. The experiment was repeated at close, medium & far distances from the signal source as before with the Nor 275 speaker & the Tivoli speakers.

The final stage of the experiment involved measuring the STI from the Tivoli speakers facing towards the absorptive surface of the wall. The measurements were taken at 1m & 3m respectively facing the direction of the speakers. This was to determine the measurement of the STI on grounds of effective sound localization

In Anechoic chamber:

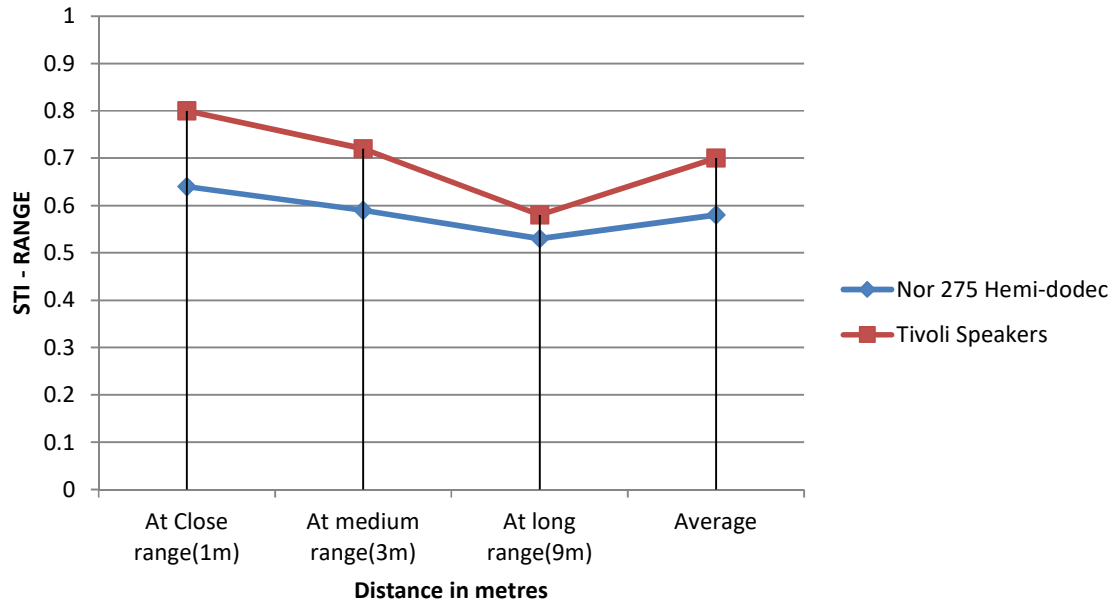
Here only the Tivoli speaker and a new addition - Yamaha powered monitor speaker model HS 50M were used. Both of them were tested at a distance of 3m from the signal source. The measurements were repeated several times and the STI results averaged to improve the accuracy of tests on the basis of repeatability.

The time set for the measurement was 12 seconds.

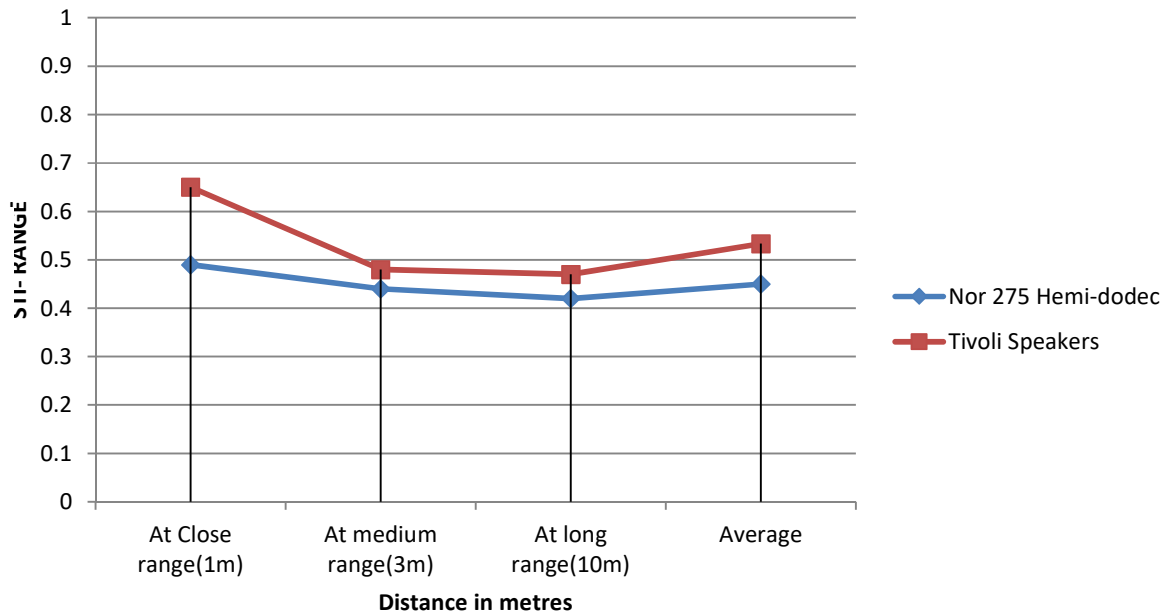
The background noise within was also measured & noted.

Results & Analysis:

Graph 1 (refer table A in annexure)
SPEAKER IN LAB ROOM with background noise of 36dBA



Graph 2 (refer table B in annexure)
SPEAKER IN REV. ROOM with background noise of 35dBA



Graph 3 (refer table C in annexure)
SPEAKER IN REV. ROOM with background noise of 35dBA & Absorptive Surface of 10m²

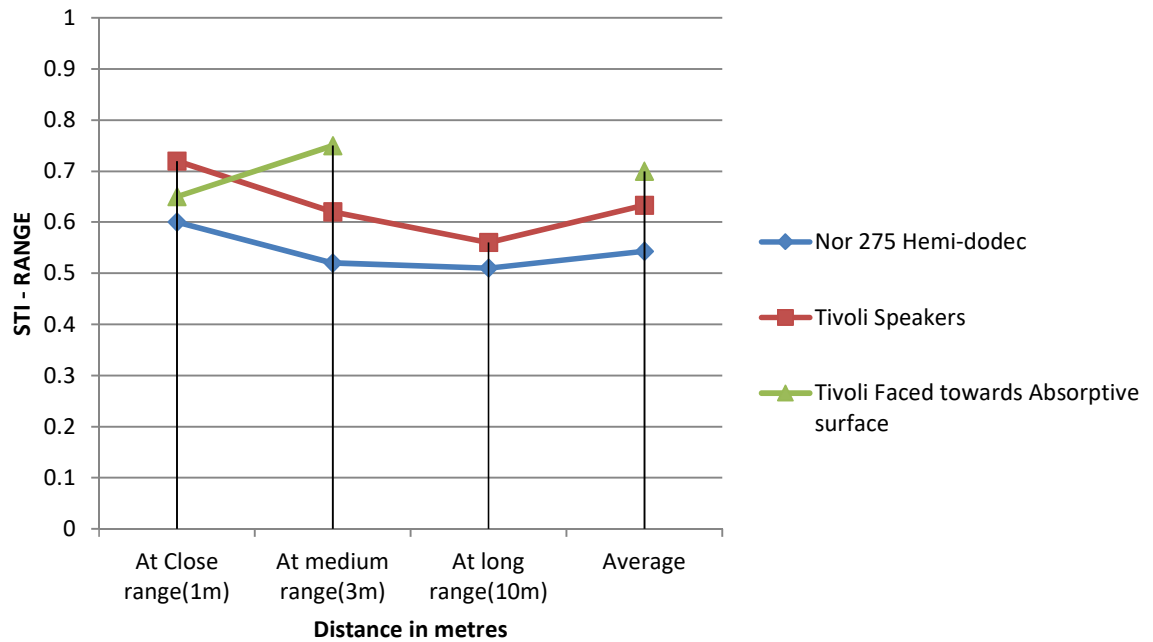


Table 1 : Anechoic chamber observations

SPEAKER IN ANECHOIC with background noise of 21.6dBA	Measurement @ 3m from signal source	STI-PA	Average.STI-PA	SPL in dBA	Avg.SPL
Tivoli	1	0.92	0.925	58.3	57.65
	2	0.93		57	
Yamaha HS 50M	1	0.86	0.906	61.2	68.8
	2	0.92		61.6	
	3	0.92		66	
	4	0.94		75.2	
	5	0.89		80	

In the Lab room, with a background noise of 36dBA; it is observed that the SPEECH Transmission Index(STI) has a gradual decline in its level as the distance between the signal source and the sound level meter is increased by 3m. That indicates that the clarity of intelligible speech goes on declining with increase of distance.

Moreover there is a difference in the STI of the **Nor-275** (hemi-dodecahedron) speakers & the **Tivoli** speakers. The Nor-275 shows a low STI average of **0.58**(which is termed to be intelligible speech as per the quality rating- **Table 0.0**) when compared to Tivoli speakers STI average of **0.7**(which is termed as very good). The sound power level output of both the speakers being almost the same at all distances/positions on a time measure of 12s.

Hence, this proves that the Tivoli speakers proved to be better than the Nor 275, this is also because, the Tivoli speakers were uni-directional in its output whereas the Nor-275 was emitting sound in all directions(multidirectional) and wasn't specifically directing sound towards the listener/sound level meter. - **Refer Table -A**

Table A - Lab room STI measurements

SPEAKER IN LAB ROOM with background noise of 36dBA		At Close range(1m)	At medium range(3m)	At long range(9m)	Average
Nor 275 Hemi-dodec	STI-PA	0.64	0.59	0.53	0.58
	SPL in dBA	70	65	60	65
Tivoli Speakers	STI-PA	0.8	0.72	0.58	0.7
	SPL in dBA	70	65	57.3	64.1

In the reverberant chamber, with a background noise of 35dBA; it is observed that the SPEECH Transmission Index(STI) has a gradual decline in its level as the distance between the signal source and the sound level meter is increased from the medium range distance from the SLM to the farthest position(10m). Whereas there is a **steep decline from the closest position(1m) to the medium range position(3m)** from the SLM. This is more prominently noted with the hemi-decahedron (Nor-275).This indicates that the multidirectional Nor-275 speaker acted more like a unidirectional source of sound when it is the closest to the SLM Nor 140. But when moved to farthest position from medium position it shows a increase of 1.4dB in the sound level (71.9 - 73.3dB).But the Speech intelligibility shows only a downward graph, indicating poorness of clarity in intelligible speech with the increase of distance.

Refer Table -B

Table B- Rev.room STI measurements

SPEAKER IN REV. ROOM with background noise of 35dBA		At Close range(1m)	At medium range(3m)	At long range(10m)	Average
Nor 275 Hemi-dodec	STI-PA	0.49	0.44	0.42	0.45
	SPL in dBA	74.1	71.9	73.3	73.1
Tivoli Speakers	STI-PA	0.65	0.48	0.47	0.533

Moreover there is a difference in the STI of the **Nor-275** (hemi-dodecahedron) speakers & the **Tivoli** speakers. The Nor-275 shows a low STI average of **0.45**(which is termed to be fairly intelligible speech as per the quality rating- **Table 0.0**) when compared to Tivoli speakers STI average of **0.533**(which is termed as good enough). The sound power level output of both the speakers being different at all distances/positions on a time measure of 12s. The Nor-275 emitted a higher level of sound but only created more masking of sound within the space.

Hence, this proves that the Tivoli speakers again proved to be better than the Nor 275, this is also because, the Nor-275 was emitting sound in all directions creating more reverberation and disturbances.

The RT of the reverberation chamber was measured by the impulse noise method and noted in **Table D**. It is observed that the reverberation was highest within the lower frequency range mainly. So this infers that the STI would be further affected by masking noise consisting of lower frequencies(63k-1Khz) rather than higher frequencies within the space.

(for eg: machinery,equipment, or similar functions within a space that is highly reverberant can prove really bad for speech communication. Refer Table -D

Moreover when compared to the Lab room, it is seen that the STI levels of both signal sources have come down considerably when measured in the reverberation chamber. Thus the Lab room is much better in terms of a better communicative environment for speech.Although with the addition of the 10sq.m of absorptive surface on one entire wall of the reverberant chamber did enhance the audible environment of the chamber to a good level.

Refer Table -C

Table C - Rev.room STI measurements (with absorptive wall surface)

SPEAKER IN REV. ROOM with background noise of 35dBA with Absorptive Surface of 10m ²		At Close range(1m)	At medium range(3m)	At long range(10m)	Average
Nor 275 Hemi-dodec	STI-PA	0.6	0.52	0.51	0.543333
	SPL in dBA	71.8	69.2	70.4	70.46667
Tivoli Speakers	STI-PA	0.72	0.62	0.56	0.633333
	SPL in dBA	68.8	66	65.9	66.9
Tivoli Faced towards Absorptive surface	STI-PA	0.65	0.75		0.7
	SPL in dBA	64.9	68.9		66.9

Not only that but when the Tivoli speakers were directed straight towards the absorptive surface and the signal source measured, that too did a lot of good to its STI average which increased from 0.63 to a whopping 0.7 suddenly. This can be very well observed in **Table C**.

Table D - RT of the rev.chamber at various frequencies

Frequency in Hz	63	125	250	250	500	1K	2K	4K
RT in rev.room by bubble Burst with 2 people in Rev.chamber	3.51	2.73	2.96	2.89	3.01	2.58	2.11	1.42

In the Anechoic chamber, a new speaker was also brought in (Yamaha) which had a better configuration than the Tivoli speakers.(and more expensive too).

But the results of STI after emitting the signal source of sound showed a different picture; where the Tivoli gave an STI of 0.925(Excellent quality of intelligible speech) whereas the Yamaha speakers gave a lesser level of 0.906. Repetitive tests were carried out just to confirm if this result was correct or if there were fluctuations each time. However this proved to be concrete that the Tivoli speakers had a better speech intelligibility index.

The above observations give us a fair picture on directional sound sources being used within a space and the effect of the environment on the same. We could assess the quality of the room/environment for its communicative sharpness and clarity with this process. Here the anechoic chamber proved to be the best and clearest environment for speech; then came the Lab room which proved superior to the reverberant chamber as it contained less reverberant sound/reflected sound waves.

It is necessary that the STI tests & checks be done on a fixed interval basis especially in public gathering spaces like stations, undergrounds and auditoriums etc to maintain the quality of speech in its public address systems with the advent of time. This would ensure maximum safety and less confusion in announcements being made through these PA systems.

References:

Ref.: 'Introduction to Speech Intelligibility'

source:http://www.ntiaudio.com/Portals/0/Products/Minstruments/AL1/AppNotes/NTI_App_Note_Introducing_STI-PA.pdf

Ref.: Houtgast, T. and Steeneken, H.J.M., "The modulation transfer function in room acoustics as a predictor of speech intelligibility", *Acustica* 28, 1973, p.66-73.

Ref.: Bradley, J. S. "Predictors of Speech Intelligibility in Rooms," *JASA* vol. 80, no. 3 (1986)

Ref.: 'Correlation of Speech Intelligibility in Reverberant rooms with Three Predictive Algorithms' by Kenneth D.Jacob (Bose Corporation, Framingham, MA 01701, USA)

source: http://pro.bose.com/pro/technical_papers/tp_speech_intell_product.pdf

Ref.: 'Speech Intelligibility Papers'- Written by Ralph Jones. Edited by Rachel Murray P.E

source :<http://www.meyersound.com/support/papers/speech/>