Audio Engineering Society

Convention Express Paper 238

Presented at the 156th Convention

2024 June 15-17, Madrid, Spain

An open source 3D printed talkbox for speech intelligibility measurements

Daniele Ponteggia[1]

[1] *Audiomatica Srl, 50136 Firenze (Italy)*

Correspondence should be addressed to Author (dp@audiomatica.com)

ABSTRACT

The measurement of Speech Transmission Index (STI) with an acoustic input channel requires the use of a source that simulates a human speaker. The IEC 60268-16:2020 standard suggests the possibility of utilizing a simple loudspeaker driver enclosed in a small box with a recommended opening size of less than 65 mm to mimic the directional behavior of a human voice. This device, referred to as a "talkbox" by the IEC standard, is commonly employed for such purposes. However, achieving a linear response within the octave bands ranging from 125 Hz to 8 kHz, as required by the STI measurement technique from such a small loudspeaker cone driver, is non-trivial. Commercial implementations often resort to equalization to linearize the response of the device. In this paper, we present the design and implementation of a low-cost talkbox leveraging 3D printing technology and we explore a method for linearizing the response of the talkbox using pre-distortion of the STI/STIPA signal.

1 Introduction

Speech Transmission Index (STI) has recently emerged as the main standard for objective evaluation of speech intelligibility, especially in the context of sound reinforcement systems. The IEC standard IEC 60268-16:2020, now in its fifth version [1], has an history of four decades during which refined up to the current version. When assessing systems with an acoustical input, such as sound reinforcement systems where the input is a microphone or even rooms without sound reinforcement, a standardized electro-acoustic transducer is required as a source.

The standard mandates the use of a source that mimics the characteristics of human speaker. Standard states that "[...]In the absence of an artificial mouth, a suitable transducer such as a small, single-source, high-quality loudspeaker with cone diameter or aperture not exceeding 65 mm, may be used [...]. In practice, a simple loudspeaker driver with a diameter less than 65 mm, enclosed in a compact box known as a "talkbox" can serve as an alternative to much more expensive mouth, head or head and torso simulators. However, the response of such a talkbox devices typically deviates from linearity in the required bandwidth and requires equalization to replicate the spectral characteristics of human speech, as specified in the IEC standard.".

This paper introduces a simple design for such talkbox device utilizing an rather inexpensive 2.5-inch full-range loudspeaker driver housed in a 3D-printed drop-shaped enclosure produced via fusion deposition printing method. The frequency response of the device shall fullfill the standards

requirements: "the shape of the test spectrum at 50 mm from the source shall not deviate from the defined STI spectrum shape by more than ±2,5 dB when measured at the specified reference point of 250 mm or 500 mm (as nominated by the manufacturer)[...]".
A small loudspeaker driver in an enclosure will rarely adhere to this specification and some linearization of the frequency response is mandatory [2]. In our simple approach we will use a pre-distortion of the STI/STIPA signal to linearize the frequency response of our 3D printed talkbox, further reducing the cost of the device which will not require a DSP processor but a simple amplifier. This will be used in conjunction with a commercial computer controlled measurement system [3], but can also be easily implemented in other measurement solutions either open-source or commercial.

2 Loudspeaker box design

We selected an easy to source full-range 2.5" speaker driver from a local manufacturer.

Figure 1: Sica 2,5 H 0,8 SL Z000855

The driver is listed as "Hi-Fi / Studio Monitor" with a nominal frequency range from 180 Hz to 20 kHz. The datasheet frequency response is also promising, even if measured in a 5.5 lt vented box, a solution which we would like to avoid in our case where a closed-box solution in a rather small volume is preferred to get a directivity pattern less influenced by the box itself.

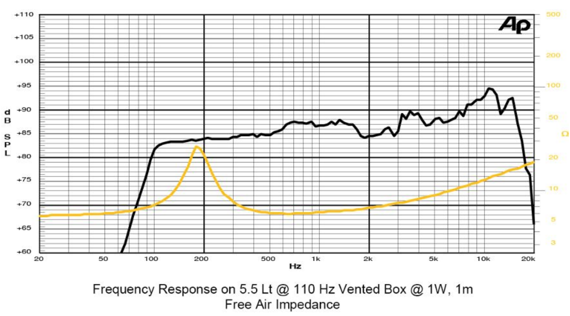

Figure 2: Sica Z000855 datasheet frequency response and free air impedance

This is a design choice which divert slightly from the one adopted by some commercial talkboxes available in the market, which are usually small shoe-box active loudspeakers. In our idea the loudspeaker box should be more similar to an artificial mouth, which is normally round shaped ideally to avoid diffractions from box edges. This is a design choice we made also from previous experience with a bigger 4 inch and a smaller 2 inch talkbox prototypes we used for several purposes in the past.
We then designed a drop-shaped box around the driver using a 3D CAD software, after few iteration we finalized the following design:

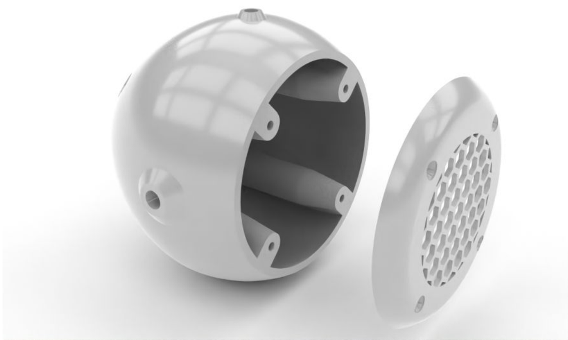

Figure 3: CAD rendering of the finalized design

The box can be made in two easy to print parts with the simple deposition technique. The two parts can be simply assembled with the loudspeaker driver using four M3 screws. Our design features an RCA connector and four small pieces of damping material inside the box.

Figure 4: All the parts needed for the build

The box has an internal volume of about 0.85 lt which seems adequate considering that the driver has a sufficient output level for the response to be corrected with active equalization. The loudspeaker needs also to be sufficiently compact to be easy to handle, in fact the design includes a 1/4-20 UNC thread to be mounted on a tripod or on a microphone stand using an adaptor.
The design is openly shared[1] as hardware through open source licensing, it can be easily replicated and/or adapted by researchers and practitioners interested in conducting acoustic STI measurements.
We built three units and tested the acoustical performances of the device.

1 The design is available on Audiomatica's web site https://www.audiomatica.com/wp/?page_id=3838

Figure 5: Actual picture of one the built devices

3 Talkbox measurements

We started the characterization of one the built devices by measuring the electrical impedance:

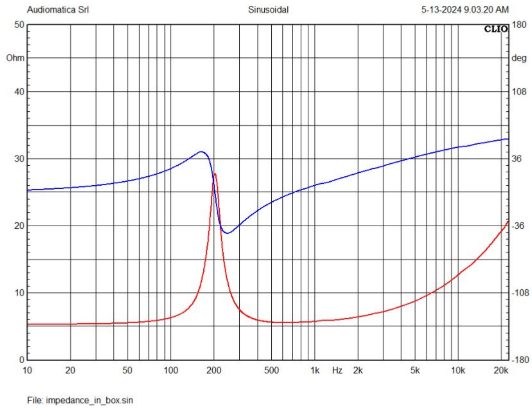

Figure 6: Electrical impedance of our talkbox speaker

We then measured the frequency response of the device at several distances with 1 V sinusoidal stimulus (Exponential Sine Sweep ESS):

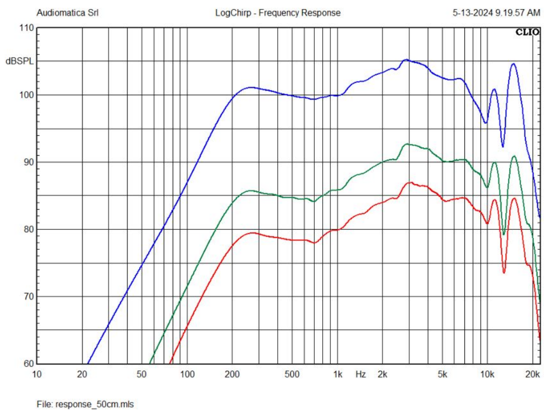

Figure 7: Frequency response at various distances

In detail the device frequency response in figure 7 reports the on-axis response at 5 cm (blue curve), at 25 cm (green curve) and at 50 cm (red curve).

According to standard requirements the STI spectrum reproduced at 5 cm should not deviate more than 2.5 dB from the one reproduced at the "reference point" either at 25 cm or 50 cm as stated by the manufacturer.

We can either post-process the above measurements or reproduce the STI spectrum, but since the response of the device is non-linear as a function of frequency we could only compare differences betweeen measured spectums. Looking at the difference among the frequency response at 5 cm and the response at 25 cm we can see that in fact over the required range the spectrum is just inside a 5 dB interval. We can then assume that the requirement is loosely met.

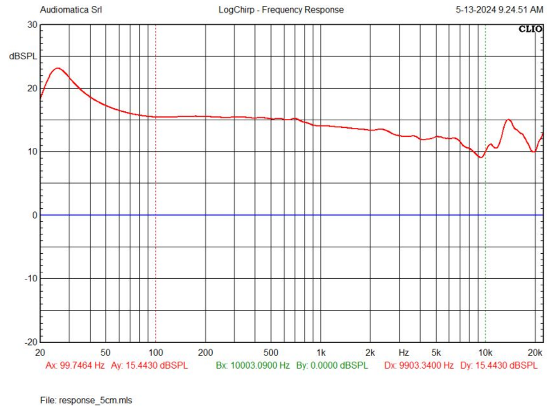

Figure 8: Difference in response at 5 cm compared to 25 cm (red curve).

We also measured the polar pattern of the device using a computer controlled turntable [4].

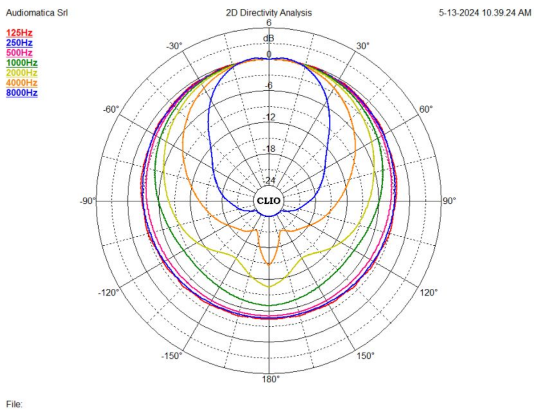

Figure 9: Device polar pattern

Thanks to its drop shape there are no visibile sidelobes.

4 Response linearization

Once the response of the device is measured, it is possibile to apply an equalization to linearize the device response. In fact the IEC standard does not require the talkbox response to be linear but instead it asks that the male spectra reproduced is the one reported in the table A.4 of the standard itself.

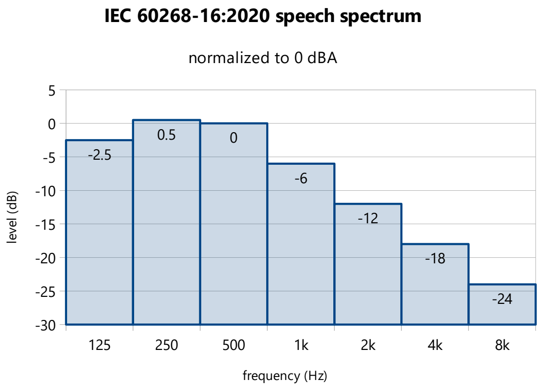

Figure 10: Normalized speech spectrum as per IEC 60268-16:2020

As an example a FIR filter inverting the band-pass filtered response of the device can be simply employed:

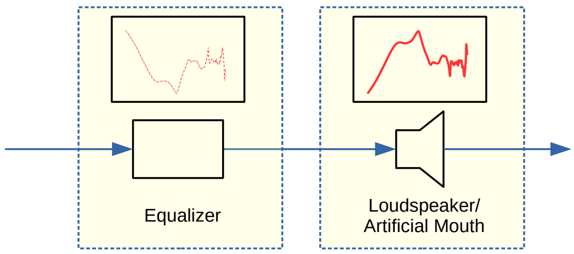

Figure 11: Linearization of the source by equalization

Here we can take advantage of the fact that the signal to be reproduced is known and it is either the male speech spectra STI noise or the STIPA modulated noise signal, depending on which measurement method is used: indirect or direct.

This tecnique can be generally used with any loudspeaker when carrying out STI/STIPA measurements, not only with the proposed loudspeaker build. Instead of using an equalizer in series with the non-linear loudspeaker we pre-distort the signal to be reproduced.

The method can be summarized as in the following figure:

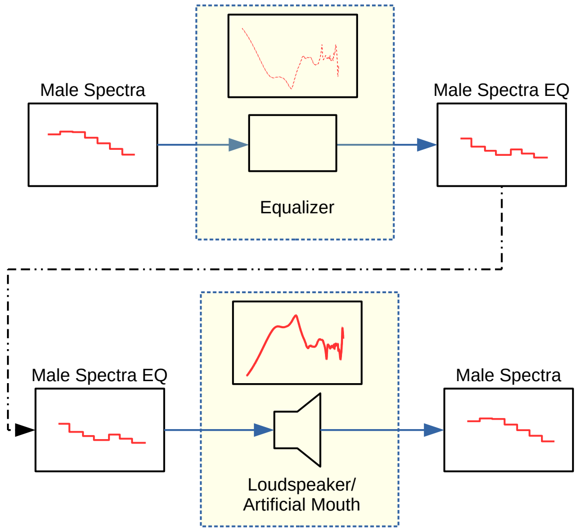

Figure 12: Linearization of the source by pre-distortion of input signal

This tecnique has been implemented in the Audiomatica CLIO system as STIPA EQ function. The method requires a complex frequency response measured with the MLS&LogChirp method, i.e. derived from an impulse response of the system to be equalized and as a result applies a FIR filter with the inverse of the response to a given signal of same lenght of the measured IR.

The tecnique does a simple inversion of the IR to match a target response which is a band-pass function with low and high cutoffs set to the band edges of the full octave frequency bands covered by the STI/STIPA method: 125 Hz and 8 kHz.

As an example we show here how a linearization of the response of our talkbox at 5 cm distance, pre-distorting the STI male level signal.

Here are shown the results of a one octave RTA measured on-axis at 5 cm from the device with pre-distorted IEC STI speech shaped pink noise (red curve) compared to the non distorted signal:

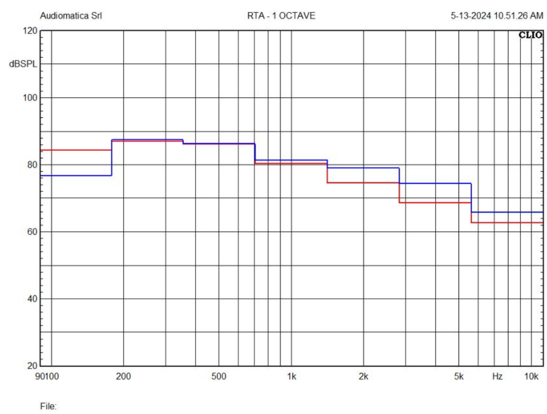

Figure 13: RTA of speech spectra without and with pre-distortion

We can compare this spectrum to the IEC definition and see that the difference is negligible and well under the suggested 0.5 dB deviation:

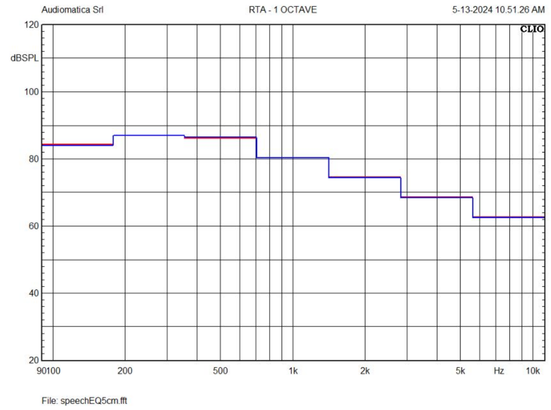

Figure 14: Speech spectrum reproduced by device (red curve) compared to standard requirements (blue curve)

A limitation of the equalization tool present in the CLIO software is the fact that the FIR filter is implemented by FFT, then the size of the IR must match the size of the signal to be pre-distorted. While this is not an issue for the STI level pink noise signal, which is a quasi-stationary noise of 64k samples length, it might become an issue if we consider the STIPA signal which has a 100 seconds length (4,800,000 samples). A possible solution in this case is to use the tool available in the CLIO software to apply pre-distortion to a delta signal and then convolve the STIPA signal with the distorted delta using a Scilab script. Another possibility is to use a Scilab script to implement the inversion and then the convolution with the designated signal.
As far as quality of the reproduced speech pink, the standard generally asks for limited amount of distortion, if we setup the level to be 60 dBA at 1 m we can derive the RMS level of the pink noise. This allows to measure the distortion using a sinusoidal signal with same RMS level. The response and distortion is measured again at 5 cm in the near field with a sinusoidal level comparable to the speech level of 60 dB at 1 m:

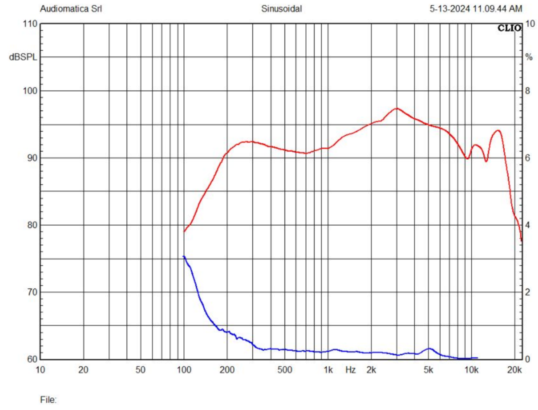

Figure 15: Distortion at a level equivalent to 60 dBA speech level at 1 m

It can be seen that the distortion is under 3% and well above that level over 200 Hz. This is related mostly to the fact that the in-box resonance frequency of the driver is at 200 Hz.
The device can cope well with a level of speech of 60 dBA at 1 m, and can also be used without thermal issues at higher levels.

5 Conclusions

In this paper we presented a build of a low-cost 3D printed talkbox which can be effectively used to inject the acoustical input in many practical cases of STI tests. We also showed a tecnique to pre-distort the STI/STIPA signal to further reduce the device cost. The same tecnique can be applied to any loudspeaker which fullfills the requests for the acustical source of the IEC standard.
The measurements demonstrate the effectiveness of the proposed talkbox design and the outlined linearization methods, offering researchers and practitioners a valuable tool for accurate speech intelligibility assessments in various acoustic environments without the need to invest in expensive equipments.
The same measurement and linearization tecnique used to pre-distort the device can also be applied to a DSP which can be used in series with the talkbox, using the same FIR filter to drive the device instead to pre-distort the signal. A brief research on this topic revealed that it is possible to use small power amplifier boards featuring an ADAU1701 DSP on-board. A further research which will follow this article will feature a possible low cost FIR implementation using such board.

References

[1] IEC 60268-16, *Objective rating of speech intelligibility by speech transmission index* 2020.

[2] S. Wijngaarden, J. Verhave, *"Designing An Acoustic Source Of The STIPA Signal: How*

To Build A Good Talkbox", Proceedings of the Institute of Acoustics, Vol. 38. Pt. 2. 2016

[3] D. Ponteggia, *"Speech intelligibility assessment using CLIO 11"*, 2016, https://www.audiomatica.com/wp/wp-content/uploads/APPNOTE_013.pdf

[4] D. Ponteggia, *"An Open Source Turntable for Electro-Acoustical Devices Characterization"*, E-brief presented at the 151st AES convention, 2021, on-line