DESIGNING AN ACOUSTIC SOURCE OF THE STIPA SIGNAL: HOW TO BUILD A GOOD TALKBOX

Sander J van Wijngaarden Embedded Acoustics BV, Delft, The Netherlands
Jan A Verhave {sander, jan}@embeddedacoustics.com

1 INTRODUCTION

A “talkbox” is a device for generating acoustic test signals, to be used as an artificial signal source instead of a live talker. Talkboxes are used in particular to produce STIPA signals for measuring the Speech Transmission Index [1]. The general aim is to reproduce the test signal in such a way that it resembles the sound coming from a live talker as closely as possible. The name “talkbox” in this context (not to be confused with an electric guitar effect with the same name) appears to have come into general use around the year 2000, with the release of STIPA. As far as the authors of this paper can trace back, the Gold Line STICis Talkbox [2] is the first device that was sold commercially with this name.

Figure 1. Gold Line STICis Talkbox

A talkbox always embodies design compromises: its fidelity usually falls short of a full-fledged Head and Torso Simulator (HATS) with an integrated artificial mouth. In particular, the directivity pattern and near-field behaviour of a talkbox is likely to deviate from a live talker. However, if one keeps the limitations of a talkbox in mind, a talkbox is easier to handle and more cost effective than a HATS.

A talkbox is essentially a combination of the following components:

- Loudspeaker
- Amplifier
- Signal storage
- Signal processing electronics
- Power source

In the case of the Gold Line STICis talkbox (fig.1), These components have simply been mounted together in a briefcase, the discrete parts clearly visible (e.g. CD player for playing back stored signals, and a multimedia loudspeaker box as the loudspeaker).

Talkboxes these days come with a variety of other test signals beside STIPA pre-installed, such as noises, sines, sweeps and speech. However, for the purposes of this paper we will focus on STIPA, as this still appears to be the primary application.

This paper draws from the experiences gained when developing the Bedrock BTB65 TalkBox (fig.2). We hope that it provides background information that contributes to the insight of anybody doing STIPA testing, but it is also intended for those who plan to build a custom talkbox of their own. A “home-built” talkbox, if properly designed and constructed, can perform just as well as any of the commercially available devices. Many sound engineers and acousticians will have the necessary knowledge and tools; the choice to buy a commercial product is often made for no other reasons that cost-effectiveness and efficiency.

Figure 2. Bedrock BTB65 Talkbox

2 USING A TALKBOX

2.1 Test setup alternatives for STI measurements

Since we are focusing on the use of talkboxes for Speech Transmission Index (STI) measurements, we will first consider some definitions in the context of the STI model. Although the STI is often loosely described as a method for measuring speech intelligibility, strictly speaking this is not accurate. Instead, the STI quantifies *the impact that a transmission channel has* on speech intelligibility. The transmission channel comprises everything between a talker and a listener, as defined by fig.3. The influence of the talker and listener are kept out of the equation.

Any procedure for measuring the Speech Transmission Index (STI) will involve replacement of the listener by an STI analyser (implemented either in hardware or software), and replacement of the talker by a test signal source. Given the fact that the STI is used across a wide variety of applications, it is perhaps unsurprising that many different embodiments of the test signal source have been implemented over the years. Whether or not it makes sense to use a talkbox depends, among other things, on the type of transmission channel that needs to be tested.

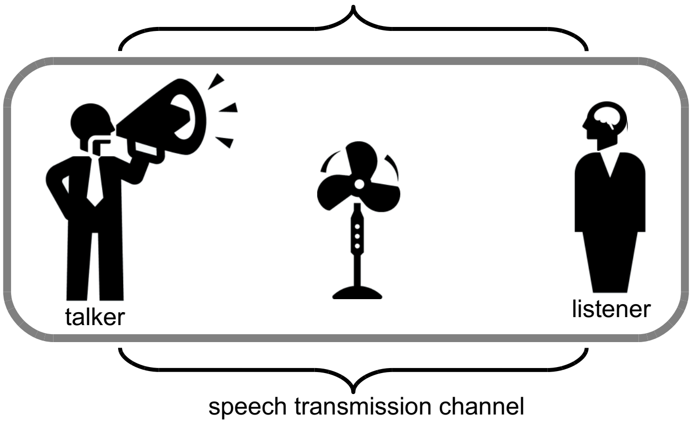

Figure 3. Schematic representation of the definition of a speech transmission channel. The channel comprises everything between the talker and listener that influences intelligibility, including noise sources and the acoustics of the environment, except for the talker and listener themselves.

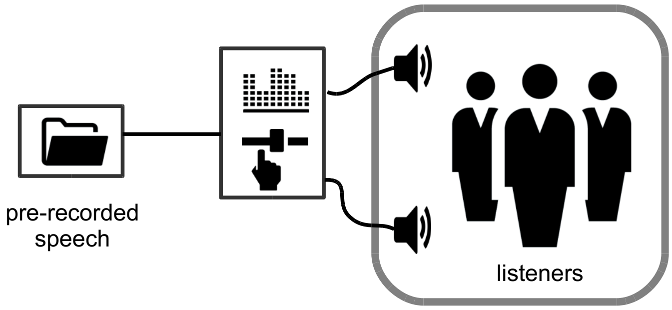

Figure 4. Schematic representation of a PA/VA system that uses pre-recorded messages. A talkbox is not needed.

Figure 4 shows a schematic representation of a Public Address (or Voice Alarm) system that uses pre-recorded messages. There is no live talker; instead, all messages are stored as files on the system. Situations like this do not call for a talkbox. Instead an audio file of the STIPA signal is usually uploaded to the system. In general, the compression technique can be seen as part of the transmission path. The STI may deteriorate due to the compression which will also be the case with the speech signal. Some compression techniques however will have a different effect on STIPA than on speech. The advice is always to cross check compatibility STIPA and compression with the standard.

If a live talker is involved, a paging microphone will be part of the transmission channel. In that case, it is recommended to use a talkbox, including the paging microphone as part of the tested transmission channel.

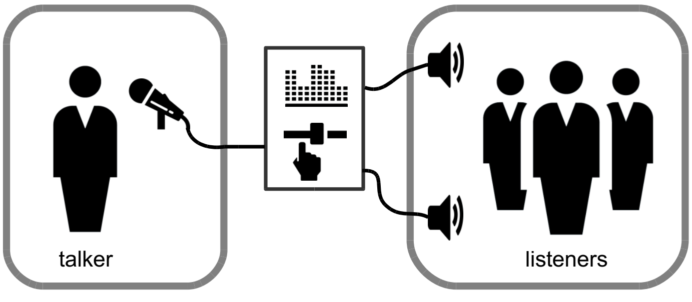

Figure 5. Schematic representation of a PA/VA system with a liver talker. Use of a talkbox is recommended.

The talkbox is set up in front of the paging microphone. The distance between talkbox and microphone should be the same as the distance from the mouth of a live talker to the microphone. The talkbox must be calibrated and configured in accordance with IEC-60268-16 [1] and ISO-9921 [3] to produce the level and spectrum that corresponds to a nominal talker.

Alternatively, the STIPA signal may be fed electrically into the PA system, bypassing the paging microphone (e.g. electrically into the microphone input, or into a line input of the PA system). This means that the influence of the paging microphone and the acoustic environment of the talker is excluded from the STI measurement, possibly leading to incorrect (usually overly optimistic) test results. Also, the electrical input level of the STIPA signal must be matched to the acoustic level of a live talker by hand (the procedure for which, as described in Annex J of IEC-60268-16, is not exactly trivial). For these reasons, the use of a pre-calibrated talkbox is often preferred.

2.2 Precautions when using a talkbox

Whenever a live talker is replaced by a talkbox, a number of precautions need to be taken to ensure the validity of the measurements. In general, one needs to make sure that the signals generated by the talkbox comply with the applicable standards, and that within the setup as a whole the talkbox works as intended.

The following precautions should not be necessary with a commercial (or previously validated home-built) talkbox, but are worth considering on newly rigged devices:

- Make sure that the actual signal level matches the selected vocal effort
- Make sure that the octave band spectrum matches the spectrum indicated in the standard.
- Make sure that the directivity pattern of the talkbox matches the directivity pattern of a human talker.
- Make sure that any distortion components produced by the talkbox are small enough that your measurements are unaffected.
- Compensate for small pauses in running speech which do not occur in the STIPA signal (Annex J)

The following paragraphs outline a series of steps and precautions that always need to be considered when using any talkbox.

2.2.1 Alignment of the paging microphone

The spectrum and the nominal level of the STIPA signal is defined by the IEC standard [1]. However, note that the specified spectrum is presumed to correspond to the main loudspeaker axis (perpendicular to the front panel of the talkbox). Off-axis measurements will yield a slightly different spectrum, due to the directivity pattern of the talkbox. Neither live talkers nor talkboxes will behave like omnidirectional radiators, especially in the higher frequency bands.

Mostly, talkers are expected to talk straight into the microphone, and the talkbox must be aligned at an angle of 0 degrees accordingly. In rare cases, incident speech from talkers and talkboxes will hit the microphone at an angle (e.g. in the case of microphone permanently fixed off-center on a desk). Either way, the talkbox must be oriented in the same direction as the live talker it replaces for the purposes of the test.

2.2.2 Distance between talkbox and paging microphone

The (electrical) level of the signal recorded by the paging microphone depends on the distance between the talkbox and the paging microphone. The talkbox must have a reference plane (usually the front panel), corresponding to a distance of 0.00m. This plane corresponds to the lips of a live talker similar to the mouth reference point (MRP) on a HATS. When measuring under anechoic conditions, the level at exactly 1.00m from the reference plane must correspond exactly to the nominal speech level (usually 60 dB(A); see also section 3.1 of this paper)

2.2.3 Avoiding near-field effects

A talkbox is usually calibrated relative to its free-field response. The spectrum in the near field is likely to be somewhat unpredictable, and may not conform to the spectrum of IEC-60268-16.

In practice, this means that there is a certain minimum "safe" distance between talkbox and paging microphone. Especially if you are testing a headset-microphone that is normally used at very short range (touching the lips), special care needs to be taken. Our procedure in such cases is to calibrate the talkbox specifically for exact position where we plan to place the microphone, making sure that the microphone remains fixed in place during measurements.

The minimum distance when such precautions have not been taken depends on the characteristics of the talkbox, and is hard to determine theoretically.

A pragmatic way to determine the minimum acceptable distance between microphone and talkbox is by simply measuring the frequency transfer characteristic as a function of distance and position. Based on our measurements (see also section 3.4 of this paper), we suggest as a general rule of thumb to maintain a minimum distance that is 3 to 4 times the diameter of the loudspeaker diaphragm.

3 DESIGNING A TALKBOX

As with any design process, there is no single best solution when it comes to developing a talkbox. Which approach is best depends to a large degree on the specific design goals. This goes beyond simply complying with the applicable standards. We propose that the following design goals are to be served:

1. Under nominal conditions, the talkbox must meet or exceed acoustic specifications defined by the applicable standards, as well as specifications following from commonly accepted standards of "good laboratory practice." This is the minimum requirement for the talkbox to be fit for its use.
2. The talkbox must be designed in such a way, that incorrect use by the end user ("user error") is inherently made less likely through the design of the system. This translates directly into the usability of the system.

3.1 Acoustic specifications

When setting target specifications for devices like a talkbox, a wide variety of different metrics can be used, ranging from common to quite exotic. Our aim is to arrive at the minimum set of specifications that guarantee the proper performance in this context (STIPA and related measurements). This results in the list of specifications given below.

3.1.1 Signal level

The nominal signal level for the STIPA signal is presumed by IEC-60268-16 to be 60 dB(A) at a distance of 1 meter. ISO-9921 defines corresponding levels for situations in which lower vocal effort is to be expected ("relaxed") and situations in which higher levels are expected ("Loud" up to "maximum shout.") These levels are defined with 6 dB steps. If the vocal effort is left undefined, 60 dB(A) at 1 meter is the default. At minimum, a talkbox needs to be able to produce test signals at this nominal level without distortion components that interfere with the measurement. Preferably, a talkbox should also be adjustable to lower (relaxed) and higher (loud) levels, so that the effect of variations in vocal effort can be studied. For this reason, we designed the vocal effort settings of the Bedrock BTB65 to be adjustable between 54 and 72 dB(A). Some application standards may also specify signal levels that deviate from the definitions of ISO-9921. For instance, NFPA-72 Annex D [5] requires the level to be set at 65 dB(A).

Any errors in the overall speech level will translate into systematic errors in STI measurements, especially when testing in ambient noise. We accept deviations of up to ± 0.5 dB, which translates into about 0.01 STI in practice. For factory calibration of commercial talkboxes, it will usually be feasible to determine and adjust the test signal level under anechoic conditions. When checking or adjusting the level in the field, the reverberant conditions of many rooms will not allow for measurements at 1m distance. Instead of checking for 60 dB(A) at 1 m distance, it is often more accurate to check for the equivalent 66 dB(A) at 0.5m.

3.1.2 Signal spectrum and frequency transfer function

IEC-60268-16 defines the octave band spectrum of the STIPA signal; see table I for the spectrum at the nominal level of 60 dB(A).

Table I. Octave band spectrum of the STIPA signal (rev 4) with an A-weighted level of 60 dB..

Octave band frequency (Hz)	125	250	500	1000	2000	4000	8000	A
Octave band level (dB)	62.9	62.9	59.2	53.2	47.2	41.2	35.2	60.0

Revision 4, the current version of the standard [1] does not explicitly specify at what distance from a talkbox to measure this spectrum. Although, the overall level of 60 dBA suggests this spectrum should be met a 1.0m distance, it is known that this test signal spectrum was obtained using speech material recorded from talkers at a shorter distance to the microphone. Due to the directivity pattern of the talkbox, the spectrum will inherently vary with distance. We normally assume the spectrum of table I to apply at a normal working distance of about 0.25m.

We specify that the STIPA spectrum from a talkbox is to correspond to table I with a margin of no more than ± 1.0 dB at a distance of 0.25m. The next upcoming STI revision will most likely require the test signal spectrum to be met at a specified distance (0.5m).

Achieving the specified accuracy for the signal spectrum will require signal processing (i.e., some form of filtering or equalization). It is unlikely that the frequency transfer function of any real-life loudspeaker will be sufficiently flat (on its own) to play back the STIPA signal without any further processing. Without further going into the means of equalization (we apply narrowband 1/n-th octaveband equalization), we simply specify the "flatness" after correction as follows: when measured in 1/3 octave bands, no band may deviate more than ± 1 dB from the mean level within the frequency range from 80 Hz to 16 kHz as per standard. Note that this specification leaves room for certain (narrow) resonance peaks and does not specify phase behaviour. This would probably not be a good way to specify loudspeaker behaviour for general sound reproduction. In this case however, these specifications ensure that the frequency transfer of the talkbox does not degrade the accuracy of our measurements, while keeping the measurements needed for validation and calibration as simple as possible.

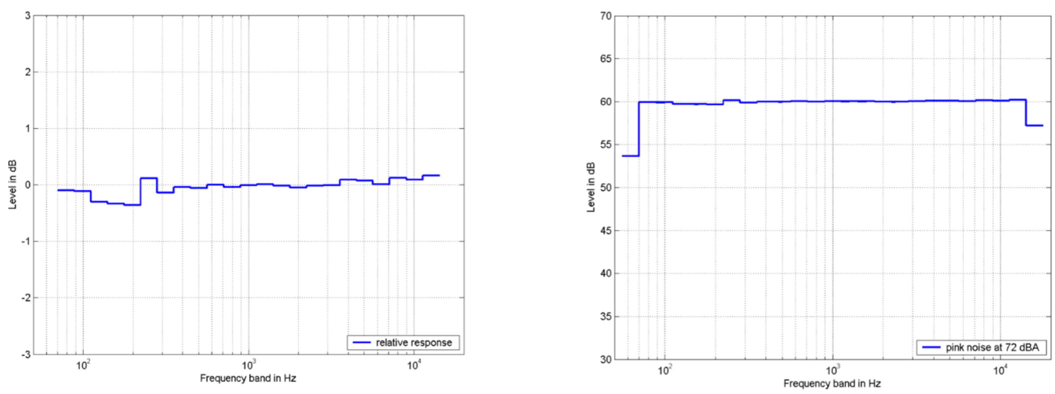

Figure 6. Typical example of the (corrected) frequency transfer function of a Bedrock BTB65 talkbox (left), meeting the specifications as defined above. The right panel shows the spectrum when playing back pink noise at a level of 72 dB(A)

3.1.3 Non-linear distortion

As long as sufficiently sensitive test equipment is used, distortion products are nearly always detectable in any system's impulse response. When specifying high end audio equipment the aim is usually to make sure that distortion products are small enough that they are not perceived by the listener. This requirement is usually translated into specifications in terms of distortion metrics for harmonic distortion (such as THD+N) and intermodulation distortion. For the talkbox, we adopt a different approach towards assessing distortion.

The difference between the design of a talkbox and of a loudspeaker (e.g. for music playback), is that talkbox signals are *not* primarily intended for human listeners, but rather as test equipment. Keeping distortion components below the threshold for human perception is still desirable, but not strictly necessary, as long as the validity of our measurement is not at stake. Lesser importance is assigned to creating a perceptually distortion-free device, also keeping in mind that:

- We aim to use a loudspeaker with a small-diameter cone to obtain the required directivity characteristics (see below); these smaller loudspeakers are inherently inefficient for low frequencies;
- We need to correct the frequency transfer so that it is flat throughout a wide frequency range, which means that we will need to boost the low-frequency end of the spectrum;
- We prefer the maximum sound pressure levels at which the test signals can be played back to be as high as possible.

As table I shows, the lower bands in the STIPA contribute largely to the overall sound pressure level. We need to be able to play back this signal at levels up to at least 66 dB(A), but preferably 72 dB(A). Almost every loudspeaker with a 50-60mm cone having a sufficiently wideband response, will have to be operated close to its limits, and non-linear distortion products will be higher than one might accept for high end audio applications.

Our criterion for accepting nonlinear distortion products is that no measurable influence on STIPA results occurs. STIPA measurements are sensitive not only to noise and reverberation, but also to non-linear distortion. Non-linear distortion components that are strong enough to have any impact speech intelligibility are penalised. There are situations in which distortion components may be audible without showing up in STIPA results. In those cases, the components are strong enough to be perceived, but not strong enough to have an impact on speech intelligibility.

The STIPA-based validation measurements for distortion are straightforward: simply carry out STIPA tests under noise-free and anechoic conditions throughout the entire range of sound pressure level settings for the talkbox, and make sure that the individual m-values per octave band exceed 0.95 (m>0.95). This implies that the overall STI is not affected by distortion; hence, the intelligibility of any speech produced by the talkbox is also unaffected.

Validation tests as described above were also done for the Bedrock BTB65, making sure that non-linear distortion components did not interfere with STIPA tests (see table II). Similar (functional) tests were done to validate other types of tests commonly done with this device (such as sweep-based frequency transfer measurements). Yet under some specific conditions (at the maximum sound level settings, mostly when playing back speech signals under perfect listening conditions), distortion components cross the threshold of audibility for trained listeners.

Table II. Average m-value measured for the BTB-65, across all band and modulation frequencies, at a distance of 0.50m, as a function of sound pressure level.

LAeq (dB)	54	57	60	163	66	69	72
Average m-value	0.99	0.99	0.99	0.99	0.99	0.99	0.99

3.1.4 Free-field behaviour

As stated in section 2 of this paper, it is important that the distance between the talkbox and the (paging) microphone under test is large enough to observe free-field (or far-field) behaviour. Near-field effects will cause the reproduced signal spectrum to be inaccurate. The level in the lower frequency bands are usually higher than intended.
A pragmatic way to determine the minimum distance that must be observed, is by simply measuring the frequency transfer function as a function of distance between talkbox and microphone. Near-field effects will show up as erratic deviations that increase as the distance decreases, most dominantly in the higher frequency bands.

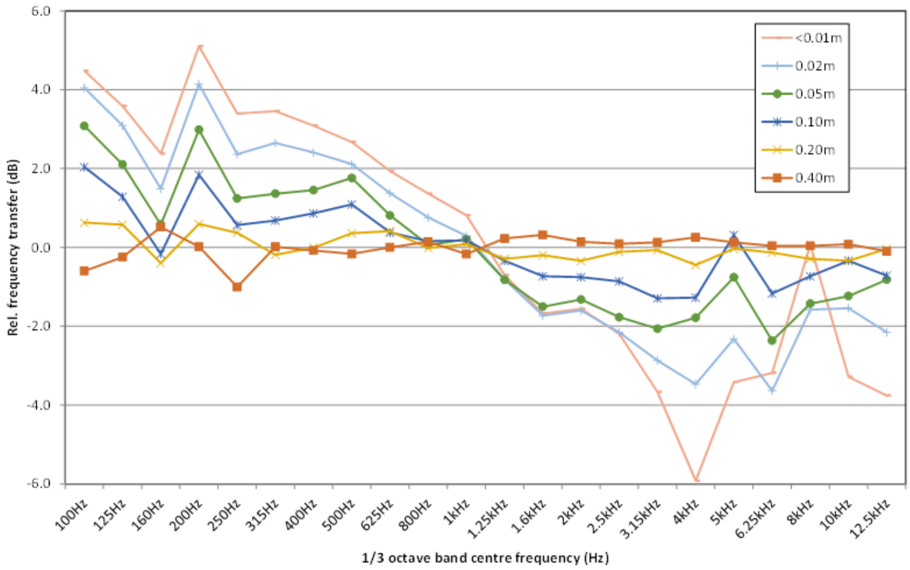

Figure 7. Frequency transfer function of the Bedrock BTB65 Talkbox as a function of distance (SPL 60 dB(A) at 1 meter distance).

As fig. 7 shows, evidence of near-field effects do not start showing up for the Bedrock BTB65 until distance is smaller than 0.20m. Above that distance, all 1/3 octave band levels are within the specified ± 1 dB margin. As the distance is reduced below 0.20, we observe two different phenomena:

- A gradual shift of energy from lower to higher frequencies. This also occurs at larger distances – simply because the directivity is higher in the high frequency bands, and the talkbox is calibrated at a single specific distance. However, the shorter the range, the more pronounced the effect.
- Erratic, position dependent deviations in the high frequency bands.

As long as the transfer function remains flat within ± 1 dB, we know that the impact on STIPA results will be negligible. Fig.7 then reveals that the Bedrock BTB65 is safe to use for STIPA measurements at distances down to 0.20m. Down to 0.10m or even 0.05m, deviations are smaller than 2 dB, and the overall error in STIPA measurements is still quite unlikely to exceed the usual statistical measurement error of up to 0.03. Below 0.05 cm, a dedicated calibration of the talkbox will be needed (such as keeping the microphone position fixed while applying the predicted level for the intended operating distance).

Note that near-field behaviour will also be observed with live human talkers, and ideally a talkbox should match the sound field around the human mouth. This is not feasible in practice – for one thing, there is a lack of objective near-field measurement data on live talkers. We consider this problem beyond the scope of this paper. In practice, the main concern with close-range measurements is that the irregularity of the directivity pattern will increase, and small differences in microphone positioning may translate into very significant differences in signal level. The slotted grille of the Bedrock BTB65 (see Fig.2) was designed to modify this characteristic of the near field as much as possible; the elongated slots varying in length and orientation are intended to produce varying contributions to the overall near-field directivity pattern, that average out in the overall pattern. When applying a loudspeaker without taking such measures, the acceptable minimum safe distance is usually a little greater.

3.1.5 Directivity pattern

The directivity pattern of the talkbox depends mostly on the dimensions of the entire device (box) and the loudspeaker, and to a lesser degree to the surface materials of the device. The general aim should be to replicate the directivity pattern of a live human talker with a reasonable degree of accuracy. It will have to be accepted that the directivity pattern will never be identical; if not, the design process will end up with a Head and Torso Simulator rather than a talkbox, and the complexity and cost will rise accordingly. For one thing, the fact that a talkbox does not feature a (simulated) torso will have an impact on the directivity in the bottom half of the vertical plane (the lower front area of the directivity balloon).

It makes sense to base a talkbox on an enclosure that roughly has the dimensions of a human head. In the case of the Bedrock BTB65, we chose a box that is somewhat smaller than a human head (180x150x130mm), since we subjectively found the directivity with a smaller box to match a human head somewhat better. We attribute this to the fact that an artificial box has flat reflective surfaces, compared to the curved and softer surface of the human head.

The choice of the loudspeaker diameter is the main compromise at the basis of any talkbox design. The current IEC-standard recommends a loudspeaker diameter not larger than 100mm, hence 100mm became more or less standard due to low frequency demands. However, Mapp [4] concluded that the directivity of 100mm loudspeakers is too high, and recommends using 50-60mm drivers. The drawback (as already noted above) is that a price needs to be paid in terms of maximum SPL or distortion (or both). For the Bedrock BTB65, we selected a loudspeaker with a 59mm diaphragm behind a 60mm aperture, with a slotted grille in front of it.

Over the years, a number of studies on the directivity of human talkers have been reported [6-9]. These studies are generally in agreement, with some differences resulting from the experimental conditions. Fig.8 shows how the Bedrock BTB65 compares to data from these studies.

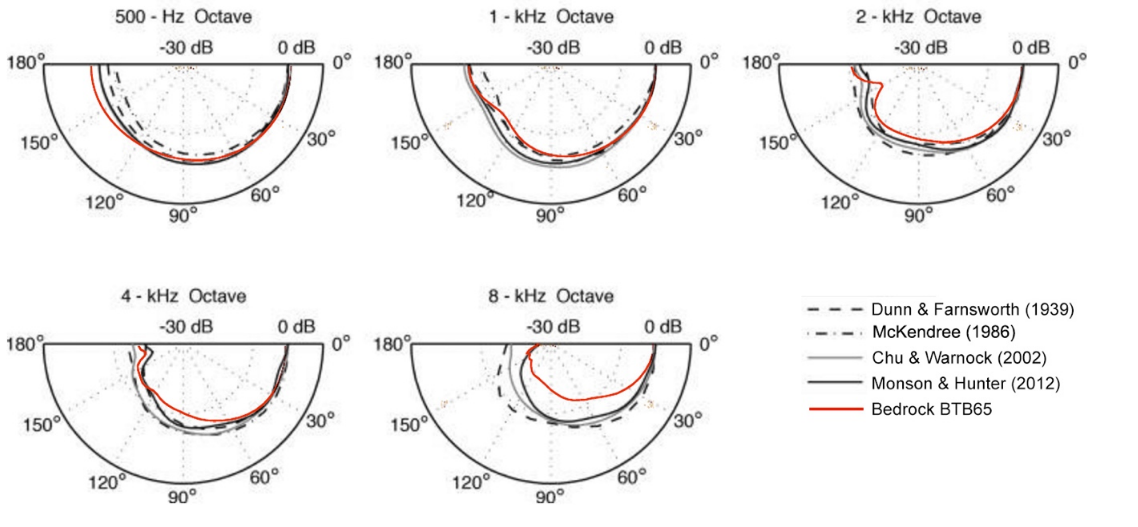

Figure 8. Directivity of the Bedrock BTB65 in the horizontal plane (red curve), compared with the directivity of human speech according to 4 different studies (after [6]). Data for lower octave bands have been omitted, since the BTB65 as well as human talkers show near-perfect omnidirectional behaviour.

Fig. 8 shows that the directivity of the BTB65, for frequencies up to 2 kHz, falls within the various studies of human directivity. At higher frequencies (most clearly in the 8kHz octave bands), the directivity of the BTB65 is notably higher than of human speech. This indicates that the diaphragm diameter of 59mm (although much smaller than the 100mm currently recommended by the standard) is still higher than optimal in terms of directivity.

Noting that the weight of the 8 kHz octave band in the Speech Transmission Index model is considerably lower than the other bands shown in Fig.8, and that most transmission channels poorly reproduce signals in the 8 kHz band anyway, the error introduced by the mismatch due to inadvertent off axis positioning in the 8 kHz band is judged to be minor.

3.2 Usability requirements

In our view, a talkbox should be designed as an easy-to-use device, that can be set up and used quickly without the need for lengthy calibrations or setup procedures. This adds to the appeal of the device and reduces the likelihood of user error.

Table III contains a list of requirements has been drafted to enhance the usuability of a talkbox, along with the way these requirements have been addressed in the Bedrock BTB65.

Table III. Proposed list of talkbox usability requirements, with a description how each requirement was addressed in the Bedrock BTB65 TalkBox.

Requirement	*Implementation in Bedrock BTB65*
The default state (in which the device starts up after storage and when it is first used) matches the most likely use cases.	The device requires no user input when booting up, and its default settings correspond to STIPA measurements at the (calibrated) nominal level. In other words, in the most common test scenario, a user only has to press "start."
The device needs to be as light, compact and portable as possible. A talkbox is of no value if it is left at the lab because it was too inconvenient to bring along.	The dimensions of the device were kept at the minimum needed to achieve the desired directivity pattern. Strong, light-weight plastics were used, adding mass internally only where needed for acoustic reasons.
Whenever a setting is changed (signal type, level, etc), the user should get clear and unambiguous feedback.	All controls are implemented through a touch display. Whenever a button is pressed, the new settings is immediately displayed in clear wording and numbers.
The talkbox should be easy to place/position and should stay in position while used.	A standard UNC tripod mount is included; rubber feet are mounted for forward-facing and upward-facing placement.
The reference point should be clearly indicated, to assist with alignment of the microphone and measuring the distance.	The reference point is unambiguously indicated as a dot in the middle of the loudspeaker grille. A laser pointer is embedded in the reference point, to assist in the on-axis alignment of the microphone.
The user should have flexibility to make adjustments, but at the same time it must be impossible (or at least very unlikely) that settings are changed by mistake.	Settings that affect the overall behaviour of the talkbox (such as adjustment of the level calibration) are accessible through a settings menu. This settings menu cannot be accessed while the device is in operation, only while the device is booting.

The list of requirements in Table III is certainly not exhaustive. When building a (custom) talkbox, the main idea is that all the design-time effort to optimize the device for its purpose only needs to be invested once, while the benefits (easier use and reduced risk of operator error) are enjoyed with every use of the device.

3.3 Other requirements

We need a talkbox to comply with our specifications not only when it is built; the device needs to meet these specs whenever it is taken out to be used. There needs to be confidence in the long-term stability of the device. The specification for the BTB65 is that it remains within its acoustic specifications for a minimum of 24 months, assuming that the device is kept protected against external influences (such as shock and impact, fluids or excessive humidity, and prolonged exposure to temperatures below freezing or above 50 deg C). True confidence in the stability of a talkbox only comes from prolonged testing (or use of a prototype in practice).

In terms of stability, the loudspeaker itself is the critical component. We found that most other failure modes of the device, such as wear/errors on flash memory used to store data, do not result in gradual degradation of its specification, but rather in immediate (and clearly noticeable) failure. The loudspeaker type used in the BTB65 is susceptible to some drift in its specifications (frequency transfer and sensitivity) in the first few hours of use. As a measure to eliminate this effect, a 12-hour burn-in period is included as part of the manufacturing process.

Another requirement concerns electric signal outputs. Since a talkbox already contains a signal generate (or a device for playback and storage of prerecorded signals), it makes sense to provide a separate line-out jack, to be used in cases where the test signal needs to be injected into the tested system electrically. On the BTB65, the loudspeaker and the line-out jack can be controlled separately (level setting, mute). The level on the line-out jack is not calibrated relative to any acoustic reference, but in dBU. A balanced line-driver is used to make sure that the flat frequency transfer and level calibration are preserved independent of the input impedance of the system under test.

4 CALIBRATION PROCEDURE

The calibration procedure is arguably one of the most complex part of designing and preparing a talkbox. In particular, making sure that the frequency characteristic of the device meet specifications can take a lot of time and effort. Not having to do any complex calibrations yourself is probably one of the main selling points of commercial talkboxes.

Potentially, there are three ways to “tune” the frequency transfer function of a talkbox to achieve the specifications (flat within 1 dB):

1. Implement a dynamic calibration procedure, including a sense microphone. The major advantage of this approach is that the device is adjusted for any drift in loudspeaker performance throughout its lifecycle. In practice, the benefit is limited (since there may also be drift in the sense microphone) and the implementation is complex and expensive.
2. Based on lab measurements, the specifications for a compensation filter for the loudspeaker’s frequency transfer is measured. This is then implemented as a FIR filter or though and embedded equalizer.
3. Instead of including a FIR filter or EQ in the talkbox, all signals are inverse-filtered for the loudspeaker characteristic.

The Bedrock BTB65 is capable of methods 2 and 3. Our production firmware uses method 3, since we are able to get the best performance that way. The compensation filters are individually measured for each talkbox, and processed offline in floating point resolution on a powerful computer. Method 2 relies on the embedded fixed-point processor. In theory, the same performance obtained when using method 3 can be achieved by embedding a floating point DSP in a talkbox – but at the expensive of additional cost and power consumption, without any noticeable benefit to the end user.

The input to our algorithm for calculating the compensation filter is a measurement on the transfer function of the (uncorrected) talkbox, obtained by playing back pink noise and recording the signal with a class 1 microphone. The signal is analysed with 1/36 octave resolution. Based on this measurement, the inverse filter is calculated.

5 CONCLUSIONS

The design and construction of a talkbox for STIPA measurements can be time consuming, but the process is fairly straightforward. The STIPA method itself can be used to obtain a quick and easy first impression on the performance and suitability of a talkbox.

When selecting a loudspeaker, a type with a small diameter (<60mm) is preferred to limit near-field effects and to achieve a directivity pattern that matches live talkers. For these smaller transducers, generating sufficiently high sound levels at acceptable distortion levels, while keeping the (corrected) frequency transfer function flat, is the main challenge.

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. Talkbox signal source Users Manual; Gold Line corporation, Redding, CT, USA. http://www.gold-line.com/pdf/manual/pa_talkbox.pdf
3. International Standards Organisation. " ISO 9921:2003, *Ergonomics – Assessment of speech communication"*
4. Mapp P. "Simulating Talker Directivity for Speech Intelligibility Measurements." Audio Engineering Society convention paper 9156, Los Angeles, USA, 2014 October 9-12.
5. NFPA72"National Fire Alarm and Signaling Code," National Fire Protection Association, Quincy, Massachusetts, USA (2013).
6. Monson BB, Hunter EJ, Story BH. Horizontal directivity of low- and high-frequency energy in speech and singing. *The Journal of the Acoustical Society of America*. 2012;132(1):433-441. doi:10.1121/1.4725963.
7. Dunn H. K., and Farnsworth D. W. (1939). " Exploration of pressure field around the human head during speech," J. Acoust. Soc. Am. 10, 184–199.
8. McKendree F. S. (1986). " Directivity indices of human talkers in English speech," Proceedings of Inter-Noise 86, Cambridge, pp. 911–916.
9. Chu W. T., and Warnock A. C. C. (2002). "Detailed directivity of sound fields around human talkers,"Technical Report, Institute for Research in Construction (National Research Council of Canada, Ottawa ON, Canada), pp. 1–47.