DOI: 10.61782/fa.2023.0597

forum acusticum 2023

SPEECH TRANSMISSION INDEX MEASURED USING ADULT AND CHILDREN HEAD AND TORSO SIMULATORS

Karin Loh* **Julia Seitz** **Hark Braren** **Janina Fels**
Institute for Hearing Technology and Acoustics, RWTH Aachen University, Germany

ABSTRACT

Acoustic measurements conducted using head and torso simulators (HATS) are considered to represent natural human hearing more realistically in comparison to measurements conducted using omnidirectional microphones. Traditionally, HATS are designed and built with respect to the anthropometric data of adults. Correspondingly, evaluation methods and metrics were primarily developed based on adults. Nevertheless, children are a major group of interest in learning spaces, and usually, they have different anthropometric head and torso dimensions than adults. This fact leads to the question of whether existing acoustic assessment methods are also valid for children. This work explores the differences in the speech transmission index (STI) derived from measurements using HATS with different anthropometric sizes with respect to children and adults.

Keywords: *speech transmission index, children, head and torso simulators, anthropometric size, binaural parameters.*

1. INTRODUCTION

Speech communication in educational buildings is important in children's daily school life. For successful development and learning, it is essential that all students can understand their teachers and fellows at every position in classrooms. However, two main problems contribute to poor speech transmission in this environment, namely background noise and unfavorable classroom acoustics [1]. Thus, it is more difficult for students to follow the lessons; they get tired more quickly, and consequently, their performance suffers [1].

To describe information loss within the speech transmission between speakers and listeners, a commonly used metric is the speech transmission index (STI) [2], which is derived from objective assessment methods. The STI is calculated by determining and further processing the Modulation Transfer Function (MTF), which contains the effects inside the transmission channel reflected by differences in the intensity envelope of the speech signal (sent signal vs received signal).

Traditionally the STI within rooms has been assessed using omnidirectional mono microphones without considering the effects of binaural hearing. Recently, several studies have investigated binaural models in combination with the STI measured using head and torso simulators (HATSs). They showed that using HATSs better represents human perception [3, 4]. A binaural STI (bSTI) comprises advantages of binaural hearing and, therefore, helps to minimize mismatches between the subjective speech intelligibility and the objective STI metric.

It is unknown whether these binaural models, developed based on investigations with adults, also apply to children with considerably different anthropometric sizes. Fels et al. showed that the different anthropometric sizes of children compared to adults result in a different amplification of sound on each octave band, respectively [5, 6]. Following this, a study in Italian classrooms [7] measured STI and interaural cross-correlation (IACC) using an adult HATS, a child HATS, and a mono microphone as receiver devices. They observed some differences in the STI between those three receiver devices. However, differences in IACC were more pronounced.

This observation was expected since the STI is developed as a single-value metric that does not account for individual differences in each octave band. Though modulation

*__Corresponding author__: karin.loh@akustik.rwth-aachen.de
Copyright: *©2023 Loh et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.*

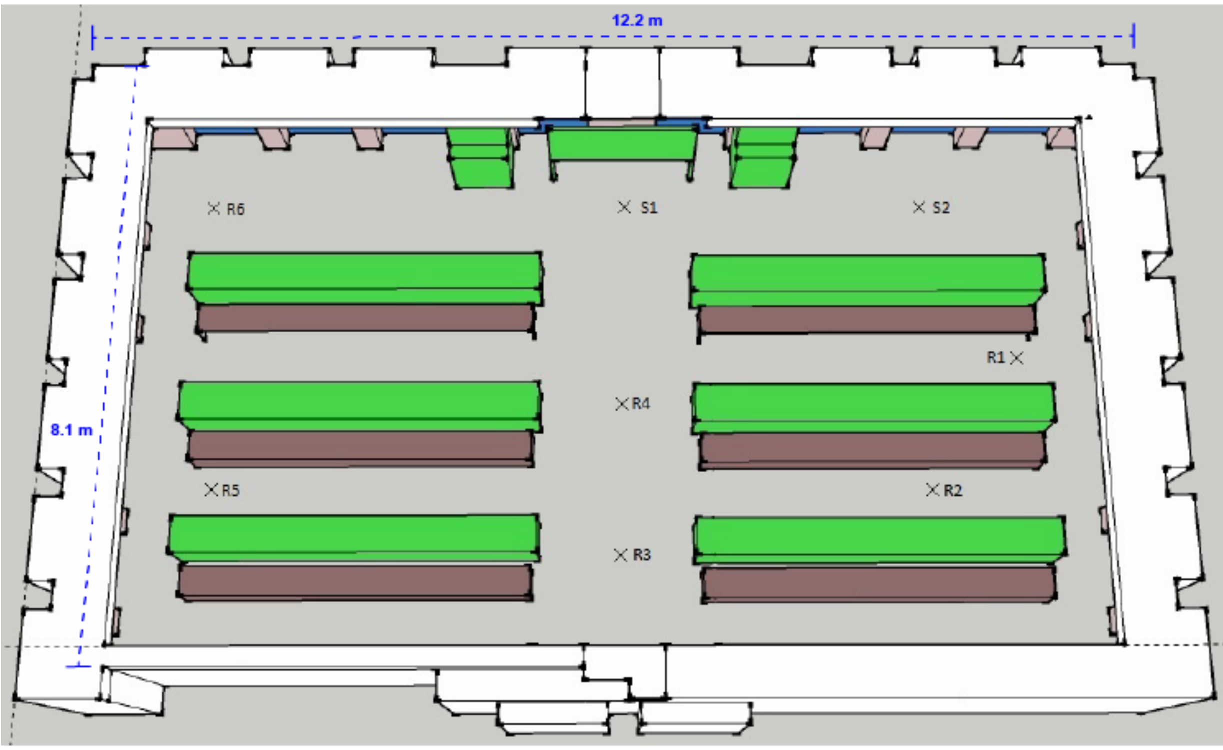

Figure 1. Visualization of the sender (S) and receiver (R) positions in the classroom.

transfer indices (MTIs) are calculated for the modulation transfer functions (MTFs) individually for each octave band, the MTIs are summed up finally to the single-value STI and interpreted as such [2].

Going a step further, there are different methods to account for the binaural hearing aspects within the computation of STI [8, 9]. The MTF coefficients for each octave band are derived for each ear separately, and before calculating the MTIs, specific MTF coefficients are chosen to represent both ears:

1. The mean can be calculated from the MTF coefficients from the left and right ear (mean method).
2. The better ear can be chosen, i.e., both ears are directly compared with each other, and the values of the MTF coefficients of the better ear are chosen (better ear method).
3. The binaural method follows the work by van Wijngaarden et al. [3], which chooses the better MTF coefficients of both ears separately for each octave band (band-by-band).

This study's objective was to explore the effects introduced by using HATSs compared to the mono microphone, including the different binaural evaluation methods and the resulting MTIs. We expect to find differences in the MTIs between the receiver devices, which might be lost in the final STI value and might be essential to account for children's speech intelligibility.

2. METHODS

Measurements were conducted in the seminar room (exemplary for a classroom) of the Institute for Hearing Technology and Acoustics (IHTA), RWTH Aachen University in Germany, following IEC 60268-16 [2] with respect to the indirect method to derive the STI. Impulse responses were measured with two sender and six receiver positions according to DIN 3382-2 [10], resulting in twelve sender-receiver position combinations (cf. Figure 1), using a sine sweep, covering the frequencies from 20 Hz to 20 kHz.

As receiver devices, an omnidirectional microphone (B&K ½" type 2669) as the reference mono microphone, the ITA adult HATS [11], and the child HATS [5] were used. The child HATS was designed based on a range of statistically analyzed anthropometric sizes of children. Therefore, it has a simplified shape and compared to the adult HATS, it has no elaborated ear shape. The head size is comparable to a kindergarten child at approx. three to six years old [5]. Both

HATSs' ear axes and the microphone were positioned at 1.15 m over the floor, representing a sitting adult's ear height. All HATSs were facing toward the sender, the omnidirectional dodecahedron loudspeaker (cf. Figure 2).

Figure 2. Measurement setup in the seminar room: omnidirectional dodecahedron source (front left), singer HATS source (front right), adult HATS receiver (center) and child HATS receiver (rear by the windows).

Furthermore, a singer HATS simulating a natural speaker was used to playback modulated pink noise for a direct measurement method for STI (as shown in Figure 3). However, these results are not discussed in this work and are objects to future studies.

Figure 3. Child (left), adult (right) and singer (center) head and torso simulators (HATSs).

Measurements as well as the post-processing and evaluation of the resulting measurement signals were conducted using MATLAB and the ITA toolbox [12].

3. RESULTS

Measurement results from in total twelve sender-receiver position combinations were averaged to receive one representative STI value for the room. STI results for all three receiver devices were generally obtained between 0.704 and 0.717 (cf. Table 1). Binaural STI measured using the HATSs were most comparable to the STI from reference microphone when the binaural STI was calculated using the mean method (cf. Figure 4). The binaural method (band-by-band) yielded highest differences for the HATSs compared to the reference microphone (cf. Figure 5).

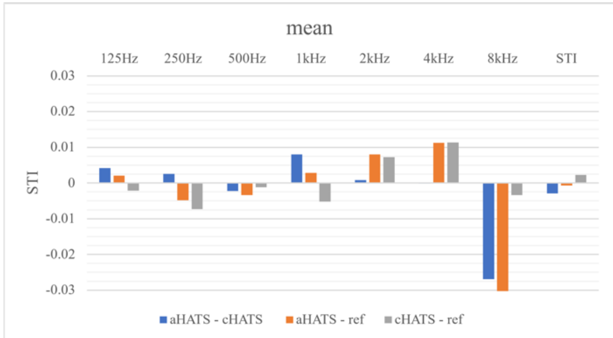

Figure 4. Differences between the receiver devices evaluated using the mean method.

To account for possible effects from the head and torso simulators, the measurements results were evaluated according to the MTI values before they were summed up to the single-value STI. With this, it is possible to examine effects on the individual octave bands. All detailed MTI values for each octave band, receiver device and binaural evaluation methods are summarized in Table 1.

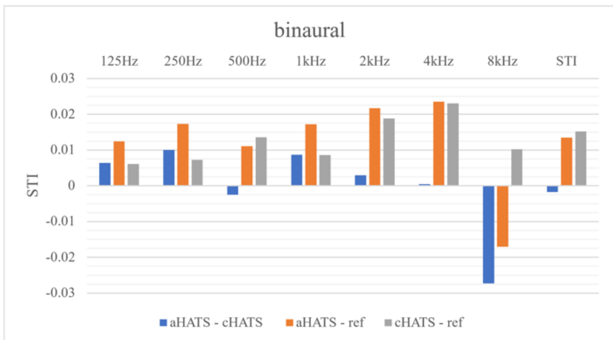

Figure 5. Differences between the receiver devices evaluated using the binaural method.

Table 1. Modulation transfer and speech transmission indices (MTIs and STIs) according to receiver devices and binaural evaluation methods.

			MTI							STI
			125Hz	250Hz	500Hz	1kHz	2kHz	4kHz	8kHz	
binaural	aHATS	*M*	0.662	0.679	0.678	0.731	0.716	0.711	0.765	0.717
		SD	0.039	0.040	0.033	0.030	0.043	0.049	0.039	0.035
	cHATS	*M*	0.655	0.669	0.681	0.722	0.713	0.710	0.792	0.719
		SD	0.039	0.043	0.039	0.037	0.037	0.045	0.041	0.036
betterEar	aHATS	*M*	0.654	0.675	0.672	0.729	0.712	0.710	0.765	0.715
		SD	0.042	0.044	0.036	0.031	0.046	0.050	0.039	0.036
	cHATS	*M*	0.650	0.661	0.674	0.718	0.710	0.708	0.792	0.715
		SD	0.040	0.046	0.042	0.040	0.038	0.046	0.042	0.036
mean	aHATS	*M*	0.651	0.657	0.664	0.716	0.702	0.698	0.752	0.703
		SD	0.040	0.034	0.037	0.030	0.039	0.048	0.036	0.034
	cHATS	*M*	0.647	0.654	0.666	0.708	0.701	0.698	0.779	0.706
		SD	0.040	0.044	0.037	0.032	0.036	0.039	0.037	0.033
	ref	*M*	0.649	0.662	0.667	0.714	0.694	0.687	0.782	0.704
		SD	0.039	0.044	0.040	0.030	0.040	0.038	0.035	0.033

Note. aHATS = adult head and torso simulator, cHATS = child head and torso simulator, ref = reference microphone. Mean (*M*) and standard deviation (*SD*) is calculated for $N = 12$ sender-receiver positions.

In general, it was observed that the differences between the head and torso simulators and the reference microphone were most significant for the 8 kHz octave band. Especially in this octave band, distinguished differences between the adult and child HATSs were found (cf. Figure 4 to Figure 6). Binaural results obtained using the mean method were lower than 0.01 on all octave bands except the 8 kHz band. Most pronounced differences between the reference microphone and the HATSs were yielded within the binaural (band-by-band) evaluation method. The differences between the two HATSs were lower than 0.01 on all octave bands. Also here, the 8 kHz octave band was particularly noteworthy. Similar observations were found for the better ear evaluation method.

However, it must be mentioned that all differences were lower than 0.03, which is considered as the just-noticeable difference for the STI [13].

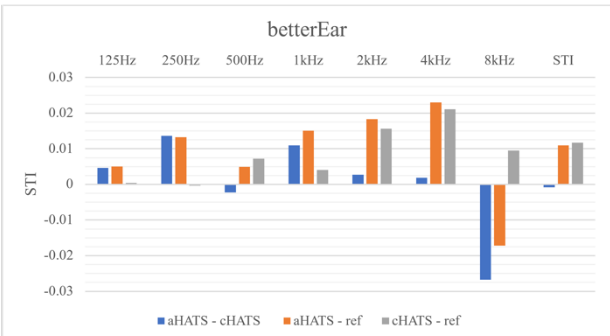

Figure 6. Differences between the receiver devices evaluated using the better ear method.

4. DISCUSSION

With respect to the work by Fels et al. [5, 6], who found that the differences in anthropometric head size led to different amplifications on the individual octave bands, it can be assumed that the STI as a single value may not sufficiently reflect the differences on the individual octave bands. Results from this work show that in the mean method, when calculating the binaural STI by averaging MTIs over both ears to result in the final binaural STI, no differences between the binaural STI of both HATSs were found in comparison to the reference microphone. When choosing a method, which considers simple binaural effects such as evaluating the octave bands individually or the better ear effect, it was observed that the binaural STI reflects differences between both HATSs compared to the reference microphone.
However, differences between the adult and child HATSs could not be found in the binaural STI independent from the binaural evaluation method. In this case, the investigation of the MTIs revealed more insights into the differences between the HATSs. Especially, the 8 kHz octave band reflected the differences mentioned by Fels et al. [5] stating differences due to anthropometric head sizes to be primarily found for higher frequencies.
Results of this study were considered preliminary since the MTIs and (binaural) STIs were derived following the indirect method for STI calculation. This method accounts for existing noises and directivity of the speaker only in a limited way which might be essential for binaural effects [9]. In this case, each ear would be positioned slightly differently within the directivity of a speaker, which might not be the case if an omnidirectional source like the dodecahedron is used. Furthermore, with regard to the work by Liang et al. [14], a strong dependency of the bSTI to the angle and positioning of the HATSs within the room and towards the sound source is expected. This effect was reduced in this study by common facing of the HATSs towards the omni-directional source, which explains the small differences between the left and right ear resulting in little differences between the bSTI and the reference STI. Differences found in this work were below the just-noticeable difference [13] of STI based on investigation including adult participants. It can, therefore, be assumed that the differences are negligible. However, it is unknow whether the just-noticeable difference also holds for children.

5. SUMMARY

This work revealed insights into the importance of binaural measurement methods when assessing speech intelligibility within classrooms. Especially in the case of children's perception, this work showed that the usage of HATSs with the anthropometric sizes of children can provide additional insights into the children's speech intelligibility within classrooms, which might be beneficial when interpreting the suitability of rooms for educational purposes.
Furthermore, this work highlighted the importance of examining the individual octave band, indicating the differences in the higher frequency bands of the HATSs and between the HATSs, which might not be reflected in the overall single-value STI.

6. ACKNOWLEDGEMENTS

This research received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 874724 (Equal-Life).
Furthermore, the project was supported by Deutsche Forschungsgemeinschaft (German Research Foundation) under the project number: 368482240/GRK2416.

7. REFERENCES

[1] B. M. Shield and J. E. Dockrell, "The effects of noise on children at school: A review," *Building Acoustics*, vol. 10, no. 2, pp. 97–116, 2003, doi: 10.1260/135101003768965960.
[2] International Electrotechnical Commission, "IEC 60268-16 - Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index." 2012.
[3] S. J. van Wijngaarden and R. Drullman, "Binaural intelligibility prediction based on the speech transmission index," *The Journal of the Acoustical Society of America*, vol. 123, no. 6, pp. 4514–4523, Jun. 2008, doi: 10.1121/1.2905245.
[4] A. Schlesinger, J.-P. Ramirez, J. van D. Schuitman, and M. M. Boone, "Report on a binaural extension of the Speech Transmission Index method for nonlinear systems and narrowband interference," *Proceedings of the International Symposium on Auditory and Audiological Research*, vol. 2, pp. 353–362, Dec. 2009.
[5] J. Fels, P. Buthmann, and M. Vorländer, "Head-Related Transfer Functions of Children," *Acta*

Acustica united with Acustica, vol. 90, no. 5, pp. 918–927, Sep. 2004.

[6] J. Fels and M. Vorländer, “Anthropometric Parameters Influencing Head-Related Transfer Functions,” *Acta Acustica united with Acustica*, vol. 95, no. 2, pp. 331–342, Apr. 2009, doi: 10.3813/aaa.918156.

[7] N. Prodi, A. Farnetani, S. Yiluya, and J. Fels, “Investigating Classroom Acoustics by means of Advanced Reproduction Techniques,” in *Audio Engineering Society Convention 122*, 2007.

[8] A. W. Bronkhorst, “The cocktail-party problem revisited: early processing and selection of multi-talker speech,” *Atten Percept Psychophys*, vol. 77, no. 5, pp. 1465–1487, Jul. 2015, doi: 10.3758/s13414-015-0882-9.

[9] M. Lavandier and V. Best, “Modeling Binaural Speech Understanding in Complex Situations,” in *The Technology of Binaural Understanding*, J. Blauert and J. Braasch, Eds., in Modern Acoustics and Signal Processing. Cham: Springer International Publishing, 2020, pp. 547–578. doi: 10.1007/978-3-030-00386-9_19.

[10] Deutsches Institut für Normung e.V., “DIN 3382-2 - Acoustics -- Measurement of room acoustic parameters -- Part 2: Reverberation time in ordinary rooms.” 2008.

[11] A. Schmitz, “A New Digital Artificial Head Measuring System,” *Acta Acustica united with Acustica*, vol. 81, no. 4, pp. 416–420, 1995.

[12] M. Berzborn, R. Bomhardt, J. C. Klein, J.-G. Richter, and M. Vorländer, “The ITA-Toolbox: An Open Source MATLAB Toolbox for Acoustic Measurements and Signal Processing,” in *Fortschritte der Akustik - DAGA 2017*, Kiel: Deutsche Gesellschaft für Akustik e.V. (DEGA), Mar. 2017. [Online]. Available: https://publications.rwth-aachen.de/record/687308

[13] J. S. Bradley, R. Reich, and S. Norcross, “A just noticeable difference in C50 for speech,” *Applied Acoustics*, vol. 58, no. 2, pp. 99–108, 1999, doi: 10/fph3jz.

[14] L. Liang and G. Yu, “Binaural speech transmission index with spatialized virtual speaker in near field: Distance and direction dependence,” *The Journal of the Acoustical Society of America*, vol. 148, no. 2, pp. EL202–EL207, Aug. 2020, doi: 10/gg9fk5.