

SOUND POWER OF NORMAL SPEECH FOR BUILDING ACOUSTICS

C Hopkins Acoustics Research Unit, University of Liverpool, UK
S Graetzer* Acoustics Research Unit, University of Liverpool, UK
G Seiffert Acoustics Research Unit, University of Liverpool, UK

1 INTRODUCTION

The sound power of speech is often needed to assess speech intelligibility¹, privacy² and security³ inside buildings. This paper reports sound power measurements for normal speech from British English speakers in one-third octave bands from 63Hz to 20kHz.

ISO 3382-3¹ formalises the assessment of the acoustic performance of open-plan offices for an occupant speaking with a normal vocal effort. This Standard only gives octave band values from 125Hz to 8kHz for unisex speech, an average of the values from male and female talkers. However, when predicting sound transmission of speech from one space to another it is often necessary to use one-third octave bands instead of octave bands because of spectral features that commonly characterise the airborne sound insulation. Unfortunately, limited sound power data is available with phonetically balanced speech in one-third octave bands, particularly for male and female talkers below 160Hz where large differences can occur. In addition, there has been recent interest⁴ in the Extended High Frequency (EHF) range of speech (i.e. above 7kHz) and the information that it provides for speech perception and recognition (particularly with fricatives). However, little or no sound power data is available between 7kHz and 20kHz and whilst this frequency range is not usually critical for building acoustics, these data are included in this paper to add to the discussion on the EHF range.

2 MEASUREMENTS

Twelve talkers (six male, six female) were recorded in the ARU anechoic chamber. These talkers were native British English speakers with an accent similar to Received Pronunciation (Standard Southern English) and between 21 and 47 years of age.

Talkers produced the IEEE sentences⁵, which form 72 word lists in total (where each list comprises ten sentences), in a pseudo-random order. Before the recording session, the talkers were asked to “speak normally as you would in everyday conversation” to elicit a normal vocal effort. If the talker hesitated or made an error, s/he repeated the sentence.

The recordings from the on-axis microphone at a distance of 1m from the mouth were reported in previous publications^{6,7} and this ARU speech corpus is freely available for download at <https://datacat.liverpool.ac.uk/681/>.

Sound power measurements were based on the procedures in EN ISO 3745:2012⁸ for precision measurements in an anechoic chamber. The sentences were recorded using half-inch, free-field microphones into a Bruel and Kjaer LAN-XI Type 3050 front end and Bruel and Kjaer Time Data Recorder at a sampling frequency of 65.536 kHz. Sixteen microphones were arranged in a hemispherical array that surrounded the talker on their right side such that symmetry was assumed for the sound field on their left side. The centre of the hemisphere was at the mouth position and the radius was 1m. Talkers were seated, and no microphone was placed underneath the seat; hence this was taken into account when calculating the sound power.

*Currently at University of Salford, UK

3 RESULTS

3.1 Measurement uncertainty

The uncertainty in the sound power level in one-third octave bands due to spatial sampling with the sixteen microphones was calculated according to EN ISO 3745 and is shown in Figure 1. Below 1.25kHz, the uncertainty is typically below 1dB, and between 1kHz and 5kHz it is 1dB to 2dB.

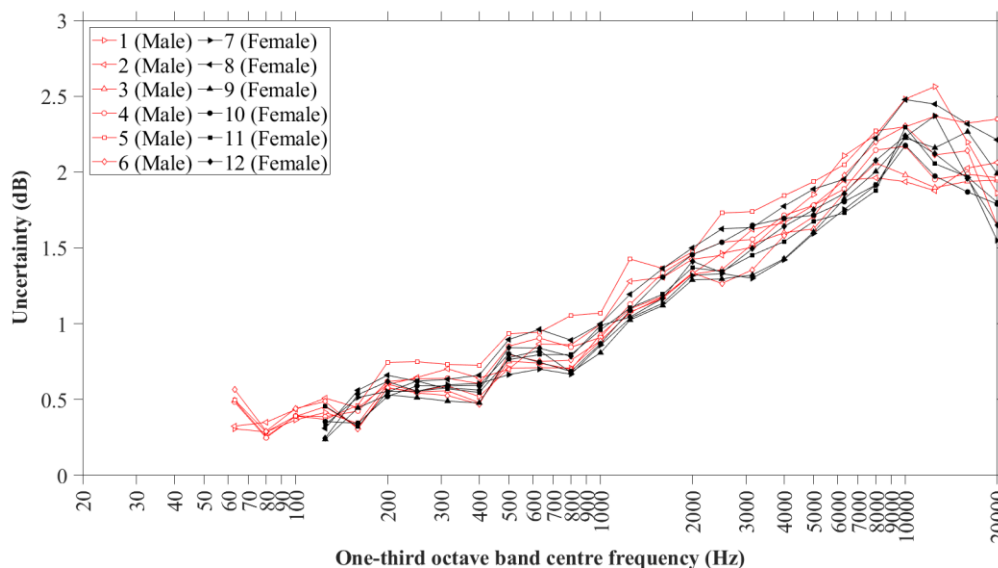


Figure 1. Uncertainty in the sound power calculated according to EN ISO 3745 for the 12 individual talkers.

3.2 Normalisation between talkers

Figure 2 shows the variation in the sound power using the average of 720 sentences from each of the 12 talkers that were asked to “speak normally as you would in everyday conversation”. There are large differences in the sound power between individual talkers. This is also evident in the sound pressure level in terms of the L_{Aeq} that was measured at 1m on-axis from the mouth (see legend). Speech was measurable down to the 63Hz band for male talkers and the 100Hz band for female talkers.

For the on-axis sound pressure level, a value of 60dB L_{Aeq} is assumed in ISO 3382-3 for normal speech in open plan offices and is commonly used as a rule-of-thumb for normal speech from adults in other situations. The average sound pressure level from the twelve talkers that were asked to ‘speak normally’ in an anechoic environment is seen to be lower than 60dB L_{Aeq} . Hence, to make meaningful comparisons between the spectral shapes of male and female talkers, the sound pressure levels measured at each of the sixteen positions are normalised such that the on-axis level is 60dB L_{Aeq} for all talkers before calculating the sound power.

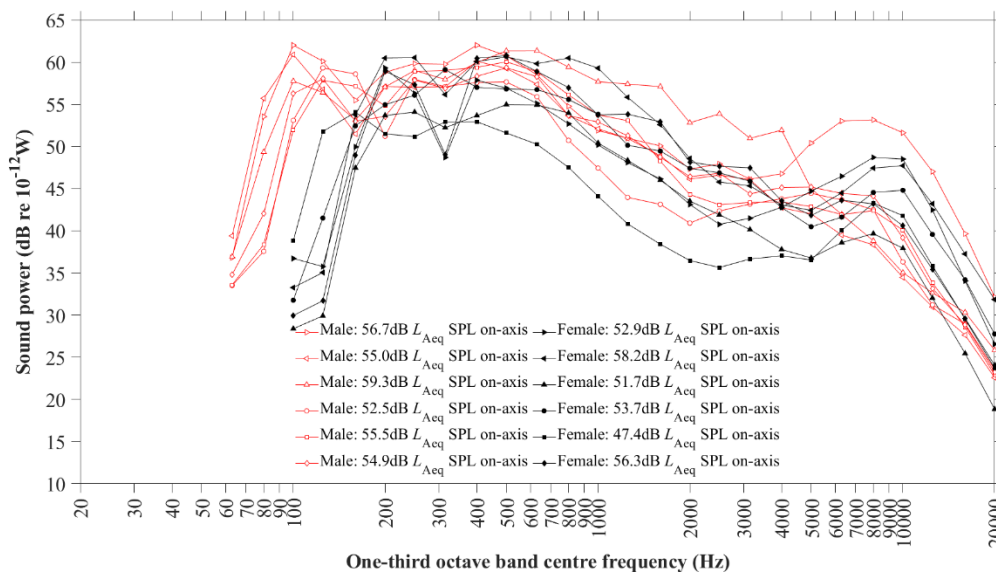


Figure 2. Sound power from the 12 talkers with the on-axis sound pressure level on terms of L_{Aeq} for each talker indicated in the legend.

3.3 Sound power

Figure 3 shows the sound power for the 12 talkers after normalization such that the on-axis microphone corresponds to 60dB L_{Aeq} . These normalised sound power data for averaged male, female, and male and female talkers are tabulated in Table 1.

At low frequencies, there are high sound power levels from male talkers in the 63, 80 and 100Hz bands where female talkers have little or no speech energy. Talker fundamental frequencies (F_0) were as low as ≈ 70 Hz for male talkers and ≈ 130 Hz for female talkers. Speech frequencies close to F_0 are not essential for speech intelligibility. For this reason, speech intelligibility parameters such as SII only use one-third octave bands from 160Hz to 8kHz where the lower bandedge of a 160Hz one-third octave band filter is ≈ 142 Hz, and STI uses octave bands from 125Hz to 8kHz where the lower bandedge of a 125Hz octave band filter is ≈ 81 Hz. However, low-frequency information is relevant to speech privacy and speech security because the availability of frequencies close to F_0 potentially allow identification of who is talking. Low frequencies may also play a role in distraction, such as in open plan offices, when estimating low-frequency background noise due to people talking and in informational masking. In these situations, it may be more appropriate to use the average male and/or average female spectra.

In the low-frequency range, there are large differences between individual male talkers (63Hz to 160Hz bands), and between individual female talkers (100Hz to 315Hz bands). Note that the uncertainty in these bands (refer back to Figure 1) is only ≈ 0.5 dB. For speech intelligibility (rather than speech privacy or security) these high levels of variation could be used to justify unisex talker (i.e. average of male and female) data at and above the 125Hz one-third octave band. In the mid-frequency range between 400Hz and 3.15kHz, the spectra for male and female talkers are similar (average values are within 2.1dB). At high frequencies (i.e. above 4kHz) there are large differences between individual talkers. However, the uncertainty is between 1.5dB and 2.5dB so for speech intelligibility assessments it is justifiable to use unisex talker (i.e. average of male and female) data at and above the 5kHz one-third octave band.

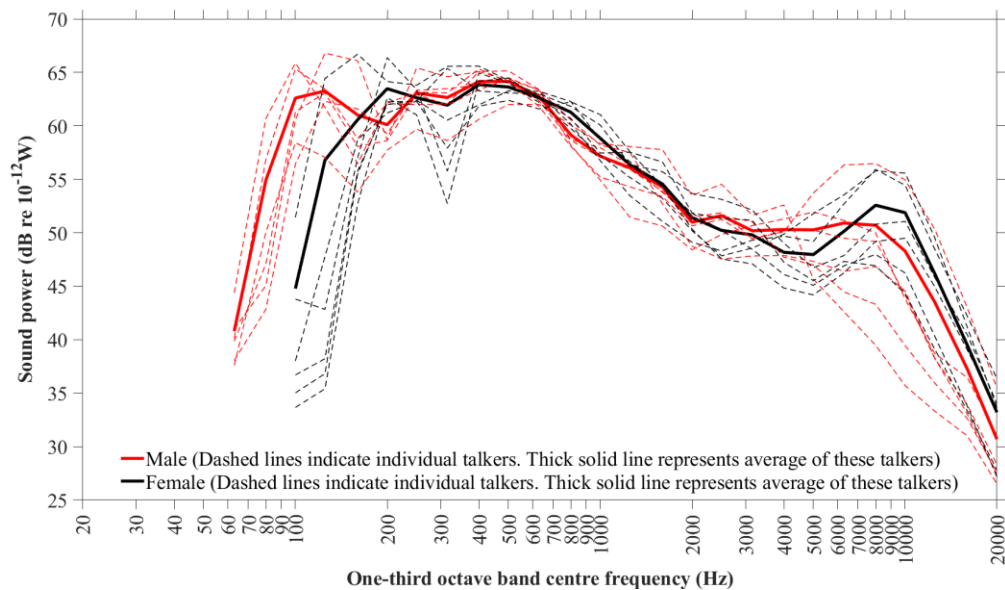


Figure 3. Sound power from 12 talkers after the levels from the 16 microphones were normalised so that the on-axis sound pressure level at 1m from the mouth is 60dB L_{Aeq} .

Table 1. Average sound power levels for speech after the levels from the 16 microphones were normalised so that the on-axis sound pressure level at 1m from the mouth is 60dB L_{Aeq} .

One-third octave band (Hz)	Sound power level (dB re 10^{-12} W)		
	Average of six male and six female talkers	Average of six male talkers	Average of six female talkers
63		40.8	
80		54.9	
100	59.7	62.6	44.8
125	61.2	63.3	56.8
160	60.8	61.0	60.6
200	62.1	60.1	63.5
250	62.9	63.1	62.6
315	62.3	62.6	61.9
400	64.0	64.1	63.9
500	63.9	64.2	63.6
630	62.7	62.8	62.6
800	60.4	59.2	61.3
1000	58.1	57.2	58.8
1250	56.3	56.1	56.4
1600	54.5	54.3	54.6
2000	51.2	51.0	51.4
2500	51.0	51.6	50.2
3150	50.0	50.2	49.8
4000	49.4	50.3	48.2
5000	49.3	50.3	48.0
6300	50.5	50.9	50.1
8000	51.8	50.7	52.6
10000	50.5	48.3	51.9
12500	45.1	43.5	46.2
16000	38.5	37.2	39.5
20000	32.1	30.7	33.2

4 CONCLUSIONS

One-third octave band sound power measurements have been carried out for normal speech in one-third octave bands from 63Hz to 20kHz. The sound power data were for six male and six female British English talkers averaged over 720 sentences for each talker. A question which arises is whether these data could be used for other languages. Byrne *et al*⁹ measured speech produced with a normal vocal effort in many different languages but only used a microphone that was 20cm from the mouth at an angle of 45° in the horizontal plane that was level with the talker's mouth. It was found that when there was a difference between languages, this was not consistent between male and female talkers. Byrne *et al* concluded that the Long Term Average Speech Spectrum (LTASS) is similar over a wide range of languages, with no single language being significantly different from any other. Hence in the absence of other data it seems reasonable to apply the unisex sound power data for British English in this paper to other languages.

5 REFERENCES

1. EN ISO 3382-3:2022 Acoustics - Measurement of room acoustic parameters - Part 3: Open plan offices.
2. ASTM E2638-10 Standard test method for objective measurement of the speech privacy provided by a closed room. (2010)
3. M. Robinson, C. Hopkins, K. Worrall, and T. Jackson. Thresholds of information leakage for speech security outside meeting rooms. *J. Acoust. Soc. Am.* 136(3), 1149-1159. (2014)
4. E. Jacewicz, J.M. Alexander and R.A. Fox. Introduction to the special issue on perception and production of sounds in the high-frequency range of human speech. *J. Acoust. Soc. Am.*, 154(5), pp.3168-3172. (2023)
5. IEEE, "Recommended practice for speech quality measurements," *IEEE Transactions on Audio and Electroacoustics*, 17(3), 227-246 (1969).
6. S. Graetzer and C. Hopkins. Intelligibility prediction for speech mixed with white Gaussian noise at low signal-to-noise ratios. *J. Acoust. Soc. Am.* 149(2), 1346-1362. (2021)
7. S. Graetzer and C. Hopkins. Comparison of ideal mask-based speech enhancement algorithms for speech mixed with white noise at low mixture signal-to-noise ratios. *J. Acoust. Soc. Am.* 152(6), 3458-3470. (2022)
8. EN ISO 3745:2012+A1:2017 Acoustics - Determination of sound power levels and sound energy levels of noise sources using sound pressure - Precision methods for anechoic rooms and hemi-anechoic rooms.
9. D. Byrne, H. Dillon, K. Tran, S. Arlinger, K. Wilbraham, R. Cox, B. Hagerman, R. Hetu, J. Kei, C. Lui and J. Kiessling. An international comparison of long-term average speech spectra. *J. Acoust. Soc. Am.*, 96(4), pp.2108-2120. (1994)