PROPOSED AVERAGE MALE AND FEMALE SPEECH SPECTRA USING HARVARD PHONETICALLY BALANCED SENTENCES

C. Nicolaides AMS Acoustics

1 OBJECT

The object of this report is to present the results collected from the analysis of phonetically balanced speech recordings.

2 SCOPE

The scope of this report is limited to a presentation of the results of the measurements collected from the analysis, and conclusions.

The scope of the results collected is limited to the recording of 9 test subjects. The method however, could be applied to a larger test populous to draw more meaningful averages.

3 APPARATUS USED

The following apparatus and software was used:

1. Marantz PMD670 Professional Solid State Recorder
2. Bruel & Kjaer Sound Level Meter (SLM) Type 2238
3. Bruel & Kjaer Sound Level Meter (SLM) Type 2260
4. Adobe Audition 3.0
5. D-Audio USB audio interface.

4 METHOD

For the recordings, Harvard phonetically balanced sentences were used. Some general information on Harvard sentences and the recording script is given in Appendix A.

Anechoic recordings of 9 subjects were made for the analysis (6 male, 3 female).

Recordings were made in the AMS anechoic chamber. The microphone of a Bruel & Kjaer 2238 SLM was used to record to a Marantz PMD670 solid state recorder (mono 44.1 kHz, 16bit PCM .wav).

Talkers were asked to read the sentences at a normal pace. Gaps were allowed between sentences, and they were asked to repeat any sentences where a mistake was made.

Editing of the digital audio files was conducted in Adobe Audition 3.0. The following processes of normalisation and analysis was undertaken on each of the recordings.

The normalisation process was conducted in the following way:

1. Each recording was split into individual sentences.

2. Each of these sentences was then normalised to 0dB peak.

3. The RMS of each of the sentences was set to that of the minimum present value.

4. The sentences were then reconstituted back into a passage of continuous speech.

5. The RMS of each of these passages was set to that of the minimum of all of the passages.

After the normalisation process was complete, for each talker's recording, the individual sentences have the same RMS. The RMS of each of the talker's entire passage is also matched.

Even though normalising to the RMS does not necessarily result in a constant perceived loudness, it worked well for these dry, un-equalised recordings.

Analysis of each file was undertaken using Adobe Audition 3.0 and a Bruel & Kjaer 2260 SLM.

From the 'Amplitude statistics' window in Adobe Audition, values for Peak (dB) and Total RMS (dB) were recorded. A screen dump of the spectral view of each waveform was also taken.

A Bruel & Kjaer 2260 SLM was used to collect data for each of the speaker's recordings. To conduct this analysis, the files were played back (looped) through Adobe Audition 3.0, and the output of the D-Audio soundcard was routed to the input of the 2260. Measurement control of; one measurement, over a 5 minute 30 second time period, was used to collect the following data:

- L_{eq} (octave bands, A weighted, and linear).
- L_{max} (octave bands, A weighted, and linear).
- L_{min} (octave bands, A weighted, and linear).
- L_1 (octave bands, and A weighted).
- L_{10} (octave bands, and A weighted).
- L_{50} (octave bands, and A weighted).
- L_{90} (octave bands, and A weighted).
- L_{99} (octave bands, and A weighted).

5 RESULTS

5.1 The data collected has summarised and presented in table 1 and figures 1 to 4. All results have been normalised to the highest present value.

5.2 Values for 'BS MALE' and 'BS FEMALE' were drawn from BS 60268-16, Table A.3. These values were normalized to the existing results, by setting their A weighted value to the same as the average L_{Aeq} of the results.

5.3 Analysis data was collected in octave bands between 31Hz and 8kHz. The 31Hz and 63Hz octave bands have been omitted from this report as they are not used in the STI standard (BS 60268-16).

5.4 Table 1, below, shows some general information on each of the talker's recordings:

Speaker	Sex	Length (s)	Peak (dB)	Tot RMS (dB)	Peak-Mean (dB)
Barry	M	25.284	-2.67	-21.6	18.93
Chris	M	24.331	-0.94	-21.6	20.66
Helen	F	24.312	-1.38	-21.6	20.22
Jacky	F	30.668	-1.93	-21.6	19.67
Janice	F	27.539	-3.77	-21.6	17.83
Ricky	M	19.548	0	-21.6	21.6
Tim	M	23.589	-0.72	-21.6	20.88
Tony	M	27.219	-1.11	-21.6	20.49
Xavier	M	22.125	-4.94	-21.6	16.66
AVERAGE	**---**	**24.957**	**-1.94**	**-21.6**	**19.66**

5.5 The male and female average L_{eq} from the results is plotted with the BS male and female results in Figure 1:

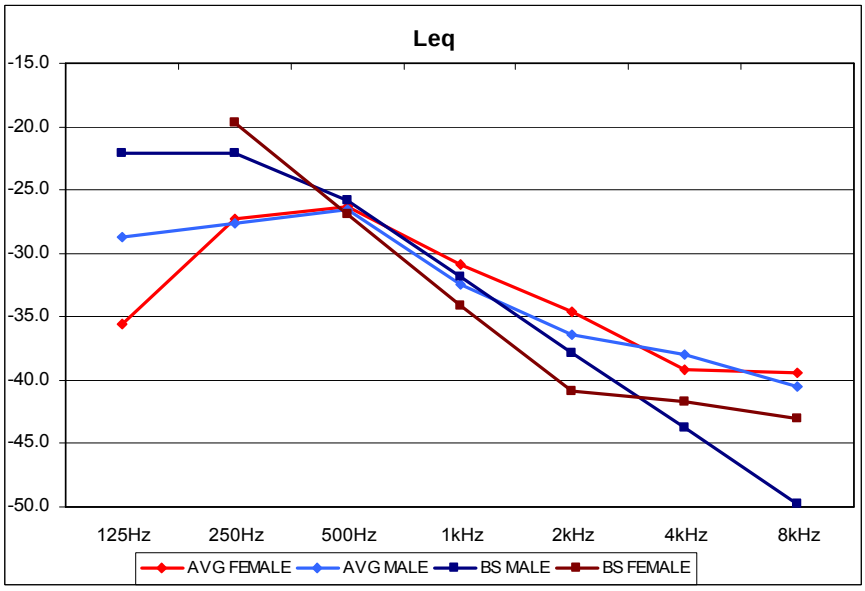

Figure 1

5.6 The male and female average L_{10} from the results is plotted with the BS male and female results in Figure 2:

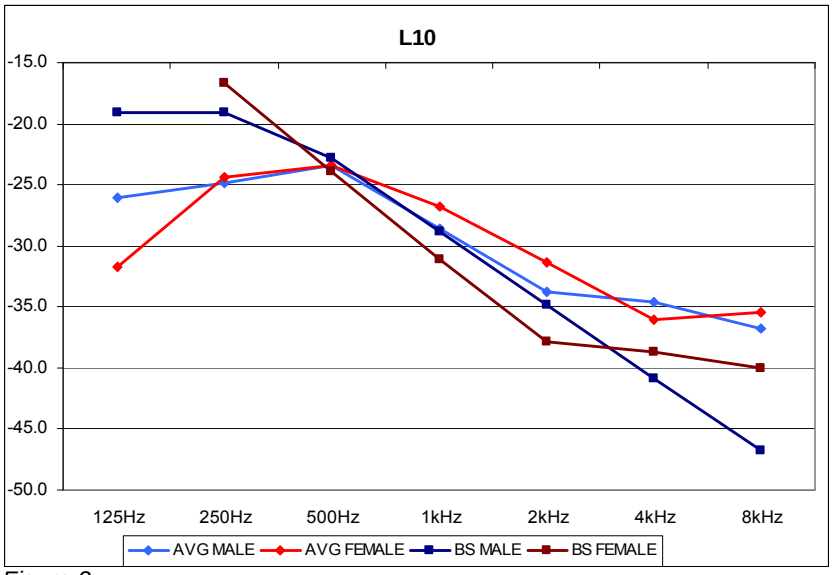

Figure 2

5.7 For information only, Figure 3 shows the Male Averages overlaid on the same axes:

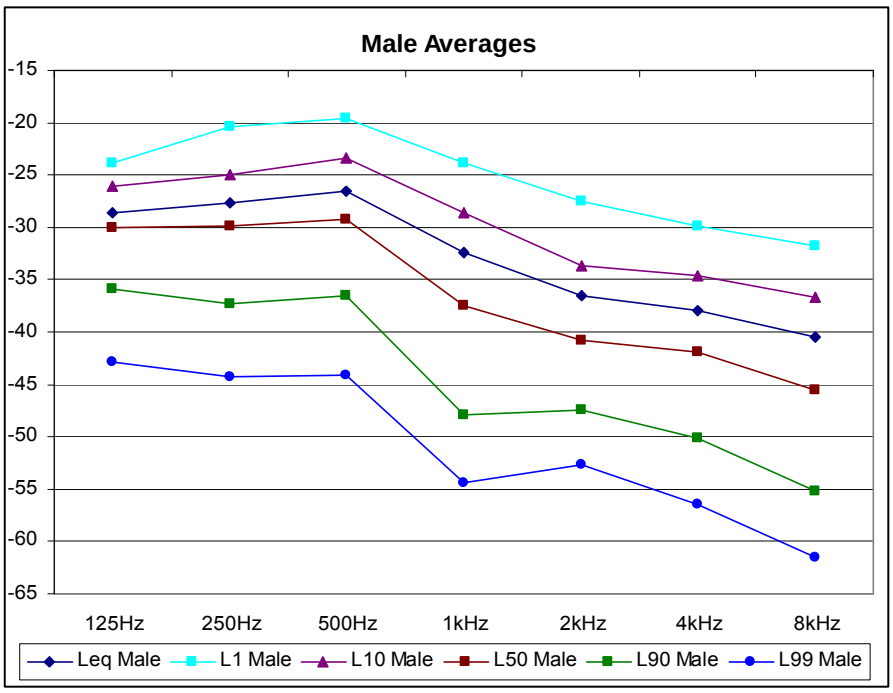

Figure 2

5.8 For information only, Figure 4 shows the Female Averages overlaid on the same axes:

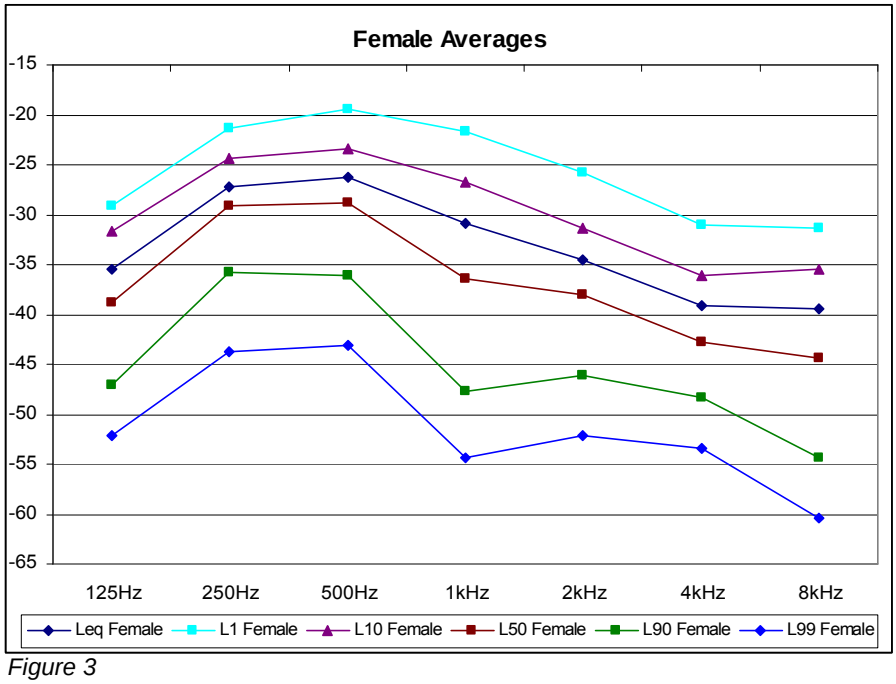

Figure 3

5.9 Screen dumps of spectral plots from Adobe Audition 3.0, showing frequency intensity v time for each recording, are shown in Appendix C.

6 CONCLUSIONS

6.1 The Peak-Mean (dB) values shown in Table 1 lay within a 4dB margin. The average value, being 19.66dB is close to the value of 18dB as predetermined by the ITU (ITU-T Recommendation P56 - Objective Measurement of Active Speech Level).

6.2 When comparing the BS MALE curve to that of the AVG MALE in figure 1, we can see a good correlation between the 500Hz and 2kHz octave bands. The correlation is much weaker at 125Hz and 250Hz; the BS MALE curve having higher values. At higher frequencies (above 2kHz), the BS MALE curve drops off rapidly below that of the AVG MALE extending to a difference of 9dB at 8kHz.

6.3 In figure 1 a similar shape can be observed between the BS FEMALE and AVG FEMALE curves, but the values are only close (within 1dB) at the 500Hz octave band.

6.4 Figure 2 shows a similar relationship between the shapes of the BS MALE and AVG MALE curves. They are similar at mid frequencies, and do not correlate at low or high octave bands.

6.5 Looking at figures 3 and 4, it is clear that the shape of the percentile curves changes for L_{90} and $L_{99.}$ There is a visible dip at the 1kHz octave band as well as a steeper treble roll-off at 8kHz.

6.6 Upon perusal of the tables included in Appendix B it is worthy to note the values of average L_{Aeq} and average L_{A10}, which are -25dB and -22dB respectively. This shows that the average A weighted L_{10} level is 3dB louder than that of the average A weighted $L_{eq.}$

6.7 In conclusion, it can be observed that some of the data collected in this study conforms to preexisting ideas and standards. However, some discrepancies can be noted; the BS MALE curve seems to have disproportionate bass and treble response and it could be suggested that this is because the BS MALE shape is overly simplified.

7 REFERENCES

1. IEEE Recommended Practice for Speech Quality Measurement, vol.17, 239
2. Sami Lemmetty, Review of Speech Synthesis Technology, 84-85
3. BS EN 60268-18:2003 Sound System Equipment – Part 16: Objective rating of speech intelligibility by speech transmission index, 22
4. ITU-T Recommendation P.56, Objective Measurement of Active Speech Level, 3

8 APPENDIX A

8.1 HARVARD SENTENCES

Extract from “Review of Speech Synthesis Technology” by Sami Lemmetty

“Harvard Psychoacoustic Sentences is a closed set of 100 sentences developed to test the word intelligibility in sentence context. The sentences are chosen so that the various segmental phonemes of English are represented in accordance with their frequency of occurrence. The test is easy to perform, no training of the subjects is needed and the scoring is simple. However, when using fixed set of sentences, the learning effect is very problematic (Pisoni et al. 1980, Kleijn et al. 1998). The first five sentences of the test material are (Allen et al. 1987):

1. The birch canoe slid on the smooth planks.
2. Glue the sheet to the dark blue background.
3. It’s easy to tell the depth of a well.
4. These days a chicken leg is a rare dish.
5. Rice is often served in round bowls.

Nevertheless the number of sentences is large, the subject may also be familiar with the test material without listening to it. For example, the first one of these sentences is used in many demonstrations or sound examples.”

The 10 sentences from ‘List 1’ of the Harvard Psychoacoustic Sentences were used as a script for the recordings:

1. The birch canoe slid on the smooth planks.
2. Glue the sheet to the dark blue background.
3. It’s easy to tell the depth of a well.
4. These days a chicken leg is a rare dish.
5. Rice is often served in round bowls.
6. The juice of lemons makes fine punch.
7. The box was thrown beside the parked truck.
8. The hogs were fed chopped corn and garbage.
9. Four hours of steady work faced us.
10. Large size in stockings is hard to sell.