PERCEIVED SPECTRAL ENERGY DISTRIBUTIONS FOR EUROM.0 SPEECH AND FOR SOME SYNTHETIC SPEECH

Chaslav V. Pavlovic, Mario Rossi, and Robert Espesser

Institut de Phonetique, LA 261 CNRS, Université de Provence,
Aix en Provence, France
& (1st. author only) University of Iowa, Iowa City, Iowa, USA.

ABSTRACT

Normative data on speech energy variations over time has been obtained for five different languages (French, Dutch, English, Italian, and Danish) available on the EUROM.0 CD-ROM produced by the SAM partners.

1. INTRODUCTION AND METHOD

The purpose of this research was to obtain normative data on speech energy variations over time for five different languages (French, Dutch, English, Italian, and Danish) available on the EUROM.0 CD-ROM produced by the SAM partners [1]. Continuous passages spoken by four talkers (two of each sex) were analyzed for each language. The variable whose distribution is analyzed is termed "perceived spectral energy." It represents the energy of speech contained within a band 1-Hz-wide at the output of an exponential time window. The exponential weighting is performed to simulate the auditory filter. Five time constants (τ) of the window are used in the analysis: 13 ms, 30 ms, 80 ms, 125 ms, and 200 ms. For a given τ the perceived spectral energy is calculated as

$$E_\tau(t) = \int_{-\infty}^{t} p_f^2(\xi) e^{-(t-\xi)/\tau} d\xi$$

where p_f is the sound pressure density in pascals. The distributions were obtained for each critical band.

2. RESULTS

The results indicate that there are no significant main effects of sex and language. Therefore, the normative values are reported as the mean values across all talkers. Fig. 1 gives results for the τ of 13 ms. Each of the family of curves corresponds to the sound pressure level below which the perceived spectral energy occurs for the percentage of time indicated at the right of the curve. All the values are in reference to the long-term spectrum level specified in Table 1. No data below the 25% contour are depicted in Fig. 1 because they were contaminated by noise. Fig. 2 refers to the measurement with the τ equal to 200 ms. The apparent dynamic range of speech with this τ appears much reduced in comparison with the results obtained with the τ of 13 ms. The results for the time constants of 13, 30, 80, and 125 ms are given numerically in Tables 2 to 5, respectively.

3. DISCUSSION

These data will be useful in future studies trying to develop physical measures of speech quality. As an illustration of the possible utility of these measures Fig. 3 gives "range" values for speech synthesized by a commercially produced synthesizer (solid lines) and EUROM.0 speech (dashed lines). The variable "range" essentially measures the dynamic range of speech. In Fig. 3 the top

curve refers to the differences in the sound pressure levels below which the speech is 95% and 25% of time. The three curves below it refer, respectively, to the differences of 75% - 25%, 65% - 35%, and 55% - 45%. It would appear that the synthesizer somewhat compresses the dynamic range of speech. This analysis may provide a means to quantify this compression in various frequency bands and relate it speech quality. A more complete version of Tables 2 - 5, as well as the table for the τ of 200 ms are given in [2].

4. ACKNOWLEDGMENTS

This research is made possible by a grant from the EEC Esprit SAM project (Grant no. 2589).

5. REFERENCES

[1] Grice, M., and Barry, B. (1988). "EUROM.0 technical description," Doc.no.:SAM-UC-135. (SAM - ESPRIT Project 2589, Wolfson House, 4 Stephenson Way, London NW1 2HE).

[2] Pavlovic, C.V., Rossi, M., and Espesser, R. (1991). "Perceived spectral energy distributions for EUROM.0 speech and for some synthetic speech. Doc.no.: CP_03_91.AIX. (SAM - ESPRIT Project 2589, Wolfson House, 4 Stephenson Way, London NW1 2HE).

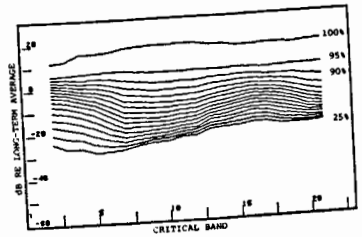

Fig.1 Perceived energy distribution for τ = 13 ms.

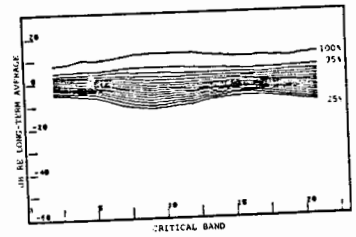

Fig.2 As Fig.1 but for τ = 200 ms.

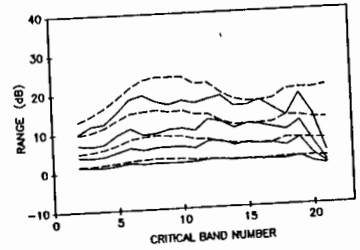

Fig.3 Range for a synthesized speech (solid lines) and EUROM.0 speech (dashed lines).

Table 1 The long-term rms of EUROM speech; 60 dB SLP overall

CRIT. BAND	CRIT. BAND C.F.(Hz)	RMS (dB SPL)
2	150	30.4
3	250	34.8
4	350	29.3
5	450	31.4
6	570	29.5
7	700	26.3
8	840	24.8
9	1000	22.7
10	1170	19.8
11	1370	18.2
12	1600	17.3
13	1850	15.0
14	2150	11.6
15	2500	12.5
16	2900	11.8
17	3400	10.6
18	4000	8.9
19	4800	7.8
20	5800	7.6
21	7000	5.9

Table 2 EUROM speech distribution for τ = 13 ms

CRITICAL	% OF SPEECH BELOW THE LEVEL IN THE TABLE									
BAND	25	30	40	50	60	70	80	90	95	100
2	-25.1	-21.9	-15.5	-9.6	-4.8	-1.2	2.0	4.9	6.5	12.9
3	-27.4	-22.7	-14.7	-8.8	-4.4	-1.3	1.6	4.7	6.7	12.4
4	-25.7	-22.2	-15.4	-9.8	-6.0	-2.7	0.3	4.1	6.7	15.5
5	-27.8	-24.4	-17.1	-10.9	-6.0	-2.5	0.8	4.3	7.0	15.5
6	-27.5	-25.2	-19.9	-14.4	-9.4	-4.7	-0.5	4.1	7.1	15.9
7	-27.7	-26.3	-22.5	-17.9	-13.0	-8.1	-2.8	3.6	7.6	17.1
8	-27.3	-26.0	-22.6	-18.5	-14.1	-9.2	-4.0	2.6	7.0	18.0
9	-25.5	-24.3	-21.5	-18.3	-14.5	-10.1	-5.0	1.7	6.6	18.0
10	-23.8	-22.8	-20.3	-17.3	-13.8	-9.7	-4.6	1.9	6.6	18.6
11	-22.6	-21.7	-19.5	-16.7	-13.3	-9.3	-4.7	1.6	6.1	19.3
12	-21.9	-20.8	-18.5	-15.7	-12.4	-8.6	-4.1	1.9	6.5	19.7
13	-19.4	-18.2	-15.6	-12.8	-9.9	-6.5	-2.5	2.9	6.8	18.3
14	-19.4	-18.2	-15.4	-12.4	-9.3	-5.9	-1.8	3.4	7.1	17.6
15	-18.5	-17.0	-13.9	-11.0	-8.1	-4.8	-1.0	3.6	7.0	16.9
16	-17.1	-15.5	-12.1	-9.1	-6.4	-3.6	-0.2	3.9	6.9	15.6
17	-16.6	-14.9	-11.6	-8.7	-6.2	-3.6	-0.8	3.0	6.3	17.2
18	-17.7	-16.1	-12.7	-9.5	-6.7	-3.9	-1.0	3.0	6.5	17.8
19	-18.7	-17.6	-14.8	-11.8	-8.7	-5.5	-2.0	2.9	7.0	17.1
20	-19.6	-18.7	-16.7	-14.3	-11.8	-8.7	-4.5	2.0	7.1	17.7
21	-19.1	-18.5	-16.7	-14.5	-12.1	-9.3	-5.3	1.8	7.3	18.2

Table 3 EUROM speech distribution for τ = 30 ms

CRITICAL	% OF SPEECH BELOW THE LEVEL IN THE TABLE									
BAND	25	30	40	50	60	70	80	90	95	100
2	-18.9	-15.7	-10.2	-6.4	-3.3	-0.5	2.1	4.7	6.2	11.8
3	-18.6	-14.7	-9.5	-6.0	-3.1	-0.7	1.8	4.6	6.4	12.0
4	-18.1	-14.9	-9.8	-6.7	-4.1	-1.8	0.8	4.1	6.5	14.8
5	-20.0	-16.0	-10.5	-6.8	-4.0	-1.4	1.2	4.3	6.7	14.2
6	-22.0	-19.0	-13.8	-9.7	-6.0	-2.7	0.5	4.3	6.8	14.8
7	-24.2	-21.9	-17.3	-12.9	-9.1	-5.3	-0.9	4.2	7.5	15.7
8	-24.4	-22.4	-18.1	-14.2	-10.4	-6.5	-2.3	3.3	7.1	16.9
9	-23.1	-21.6	-18.2	-14.7	-11.2	-7.5	-3.2	2.6	6.8	17.0
10	-21.9	-20.4	-17.4	-14.1	-10.8	-7.1	-2.9	2.8	6.7	17.4
11	-20.9	-19.7	-16.7	-13.6	-10.3	-7.0	-2.9	2.4	6.3	18.1
12	-20.0	-18.6	-15.7	-12.7	-9.5	-6.2	-2.4	2.8	6.5	18.2
13	-17.2	-15.7	-12.8	-10.1	-7.5	-4.6	-1.2	3.4	6.7	16.8
14	-17.2	-15.6	-12.6	-9.8	-7.1	-4.1	-0.6	3.8	6.9	16.1
15	-15.9	-14.2	-11.3	-8.6	-6.0	-3.1	-0.1	3.9	6.9	15.3
16	-14.4	-12.6	-9.5	-7.0	-4.7	-2.3	0.5	4.1	6.7	14.3
17	-13.7	-12.0	-9.0	-6.8	-4.7	-2.6	-0.1	3.4	6.3	15.7
18	-15.0	-13.1	-9.9	-7.3	-5.0	-2.7	-0.1	3.5	6.5	16.1
19	-16.7	-15.1	-12.0	-9.2	-6.7	-4.0	-0.7	3.8	7.0	15.7
20	-18.0	-16.9	-14.4	-12.0	-9.5	-6.4	-2.3	3.4	7.2	16.3
21	-17.9	-16.9	-14.7	-12.5	-10.1	-7.2	-2.9	3.3	7.5	16.7

Table 4 EUROM speech distribution for τ = 80 ms

CRITICAL	% OF SPEECH BELOW THE LEVEL IN THE TABLE									
BAND	25	30	40	50	60	70	80	90	95	100
2	-10.3	-8.4	-5.6	-3.5	-1.6	0.4	2.1	4.2	5.6	10.5
3	-10.5	-8.4	-5.6	-3.5	-1.6	0.1	1.9	4.2	5.8	10.7
4	-10.0	-8.3	-5.9	-4.1	-2.5	-0.7	1.3	3.9	6.0	12.7
5	-10.7	-8.5	-5.7	-3.6	-1.9	-0.2	1.6	4.1	6.0	11.9
6	-13.2	-11.0	-7.7	-5.1	-2.9	-0.8	1.4	4.2	6.2	13.1
7	-16.2	-13.9	-10.3	-7.4	-4.7	-1.8	1.0	4.7	7.0	13.5
8	-17.7	-15.4	-11.7	-8.8	-6.1	-3.3	0.0	4.1	7.0	15.0
9	-17.9	-15.9	-12.7	-9.8	-7.0	-4.1	-0.7	3.7	7.0	14.9
10	-17.2	-15.4	-12.3	-9.5	-6.6	-3.7	-0.5	3.6	6.8	15.4
11	-16.6	-15.0	-12.0	-9.3	-6.6	-3.8	-0.7	3.4	6.4	15.8
12	-15.6	-14.1	-11.0	-8.4	-5.8	-3.2	-0.5	3.5	6.6	15.8
13	-12.7	-11.3	-8.8	-6.6	-4.5	-2.3	0.4	3.9	6.4	14.4
14	-12.6	-11.1	-8.6	-6.3	-4.1	-1.7	0.8	4.0	6.6	14.0
15	-11.4	-10.0	-7.5	-5.3	-3.3	-1.2	1.1	4.0	6.4	13.1
16	-10.0	-8.6	-6.3	-4.4	-2.5	-0.7	1.4	4.1	6.1	12.1
17	-9.6	-8.3	-6.2	-4.5	-2.9	-1.1	0.8	3.6	6.0	13.4
18	-10.5	-8.9	-6.6	-4.8	-3.0	-1.1	0.9	3.9	6.3	13.5
19	-12.2	-10.7	-8.1	-6.0	-3.8	-1.6	1.0	4.3	6.5	13.3
20	-14.4	-12.9	-10.5	-8.0	-5.5	-2.7	0.6	4.3	6.8	14.1
21	-14.7	-13.4	-11.0	-8.8	-6.3	-3.2	0.2	4.5	7.0	14.5

Table 5 EUROM speech distribution for τ = 125 ms

CRITICAL	% OF SPEECH BELOW THE LEVEL IN THE TABLE									
BAND	25	30	40	50	60	70	80	90	95	100
2	-7.7	-6.3	-4.2	-2.5	-0.9	0.6	2.1	3.9	5.2	9.5
3	-8.0	-6.4	-4.3	-2.6	-1.0	0.3	2.0	4.0	5.5	9.8
4	-7.7	-6.5	-4.7	-3.2	-1.7	-0.2	1.5	3.8	5.7	11.7
5	-7.9	-6.4	-4.3	-2.6	-1.2	0.2	1.7	3.9	5.6	10.8
6	-9.9	-8.3	-5.7	-3.7	-1.9	-0.2	1.7	4.0	5.9	12.1
7	-12.6	-10.7	-7.8	-5.3	-3.1	-0.9	1.6	4.6	6.6	12.2
8	-14.2	-12.2	-9.2	-6.8	-4.5	-2.1	0.8	4.3	6.9	13.6
9	-14.9	-13.2	-10.3	-7.8	-5.3	-2.8	0.1	4.1	7.0	13.8
10	-14.4	-12.8	-10.0	-7.4	-5.0	-2.5	0.2	3.9	6.7	13.9
11	-14.0	-12.5	-9.9	-7.4	-5.0	-2.6	0.0	3.7	6.5	14.6
12	-13.1	-11.5	-8.9	-6.6	-4.4	-2.2	0.3	3.7	6.4	14.7
13	-10.5	-9.2	-7.0	-5.2	-3.4	-1.4	0.9	3.9	6.3	13.3
14	-10.4	-9.1	-6.9	-4.9	-2.9	-0.8	1.2	4.1	6.3	12.8
15	-9.2	-7.9	-5.9	-4.1	-2.4	-0.6	1.4	4.0	6.1	12.0
16	-8.1	-6.9	-5.0	-3.3	-1.8	-0.1	1.7	4.0	5.7	10.8
17	-7.8	-6.7	-5.1	-3.5	-2.1	-0.6	1.1	3.6	5.7	12.1
18	-8.4	-7.3	-5.4	-3.8	-2.2	-0.5	1.4	3.9	5.9	12.3
19	-10.0	-8.7	-6.6	-4.5	-2.7	-0.6	1.6	4.2	6.1	12.0
20	-12.1	-10.9	-8.3	-5.9	-3.7	-1.3	1.3	4.3	6.3	12.8
21	-12.5	-11.3	-9.0	-6.8	-4.3	-1.7	1.3	4.5	6.6	13.2