THE EFFECTS OF PRESENTATION LEVEL ON SONE, INTENSITY-J.N.D. AND DECIBEL QUANTISATION OF CHANNEL VOCODED SPEECH

R. Mannell

Speech, Hearing and Language Research Centre
Macquarie University, Sydney, Australia

ABSTRACT

Natural speech tokens were passed through a Bark-scaled channel vocoder simulation and the outputs of 18 B.P. analysis filters were quantised at various multiples of the Sone scale, the intensity-j.n.d.-scale and the dB scale. The resulting synthetic speech was presented to a group of listening subjects at 40, 50, 70 and 90 dB s.p.l. (ref:20 μPa.) and intelligibility scores were obtained for each type and level of quantisation. The largest step on the Sone or amplitude j.n.d. scales that did not result in a significant reduction in intelligibility was found to vary with presentation level. The largest dB step that did not result in a drop in intelligibility was, on the other hand, constant across the four presentation levels. When the Sone scale was transformed into a logSone scale it was found that the maximum allowable step on that new scale was constant across the four presentation levels. This suggests that steps of loudness doubling (and not steps of equal loudness) represent the appropriate scale for the amplitude dimension in speech perception and that the dB scale is a reasonable approximation of that scale.

1. INTRODUCTION

The deciBel scale was originally devised as a convenient scale for sound amplitude as it was considered to relate closely to human perception of loudness. This logarithmic relationship of intensity to loudness was not derived from a thorough empirical examination of human loudness perception. It derived instead from a general recognition that the relationship between human loudness perception and intensity was not a linear function of either intensity or pressure but rather something more closely approximating a logarithmic relationship. Since that time there have been a number of detailed examinations of human perception of intensity. There is more than one way of measuring human perception of sound intensity. Apart from the measurement of intensity thresholds, there are three main procedures. One procedure involves the measurement of just noticeable differences (j.n.d.'s or difference limens) [5]. The second procedure involves the examination of which intensities are equivalent at different frequencies (the Phon scale) [4] The third procedure asks what changes in intensity are required to produce a doubling (for example) in the perceived loudness (Sones) [11]. A fundamental question that has still not been fully addressed is how these measures relate to the perception of speech. It might be expected that the Sone scale would be more relevant to speech perception than intensity j.n.d.'s as the former can be derived from both complex sounds and pure tones whilst the latter was originally derived from pure tones. The dB scale would seem the least likely candidate as an appropriate scale for the amplitude dimension in speech perception as it is not directly derived from psychoacoustic measurements of intensity perception. Moore and Glasberg [8] argue that the loudness of even pure tones "depends upon the integration of loudness over a certain frequency region" (eg. 1 Bark or 1 ERB). The main disadvantage of the Sone scale is that it is very difficult to derive for individual subjects whilst it is relatively straightforward to determine the amplitude j.n.d.'s and very simple to derive the dB scale.

This experiment examines three intensity/loudness scales (dB, intensity-j.n.d., and Sone) at 4 presentation levels by quantising natural speech at increasing levels of coarseness. The aim is to determine which quantisation scale produces the most consistent intelligibility results across the four presentation levels and to thus determine which of the scales would be the best contender for representing the amplitude dimension in an auditory model of speech.

2. PROCEDURE

A channel vocoder simulation developed for another project [7] was modified to incorporate a quantisation module after the analysis BP and LP filters (see figure 1). The vocoder had identical analysis and synthesis filter banks consisting of 18 Bark-scaled filters the outputs of which were demodulated by identical 50 Hz LP filters.

Two quantisation procedures were utilised, one based on the intensity- j.n.d. scale (henceforth the j.n.d. scale) and the other based on the Sone scale. The j.n.d. scale was taken from Gulick [5] (p115) and the values were logarithmically interpolated in the frequency dimension to obtain approximate j.n.d. curves for each of the 18 centre frequencies of the BP filters. For each centre frequency the 0 j.n.d. point was set as the threshold intensity and the 1 j.n.d. point was determined to be the threshold plus the j.n.d. value at the threshold intensity. The 2 j.n.d. point was determined to be the intensity at the 1 j.n.d. point plus the j.n.d. value at that intensity and so forth to give curves similar to that depicted in figure 2. The Sone scale was developed in the following way. Firstly the Phon values were determined (after Robinson & Dadson [10]) for each of the filter centre frequencies. For 40 Phons and above Sone values were derived from phon values using the formula of Kinsler et al [6]

$$L = 0.046 \times 10^{(Ln/30)})$$

(where L is loudness in sones, and Ln is loudness level in phons)

Below 40 phons this relationship no longer holds accurately and so values were derived from the data given in Fletcher [3]. This procedure directly produces the sone curves for each of the filter centre frequencies similar to the curve given in figure 2.

The quantisation curves (at 1000 Hz) for 4 of the j.n.d. and 6 of the sone conditions are shown in figure 3.

The tokens were quantised at the output of the analysis demodulation LP filters at 6 different j.n.d. levels (1, 2, 4, 8, 16 and 32 j.n.d.s'), 7 different sone levels (0.2, 0.4, 0.8, 1.6, 3.2, 6.4, and 12.8 sones), and 6 different dB levels (1, 2, 4, 8, 16 and 32 dB). This quantisation process had to be repeated for each intended presentation level (40, 50, 70 and 90 dB s.p.l. (ref: 20μPa) for both the Sone and j.n.d. tokens as the size of any Sone or j.n.d. quantisation step increases as the presentation level is decreased. This can be seen in figures 4 and 5. The spectrum of the target of the vowel /3:/ is displayed on both diagrams at the four presentation levels. On figure 4 the four LPC spectra overlay the 1 Sone quantisation contours whilst on figure 5 the four spectra overlay the 1 j.n.d. quantisation contours. It can be clearly seen that the number of quantisation steps (and thus the fineness of the quantisation) increases as the presentation level increases for both the 1 Sone and the 1 j.n.d. scales. Not all sone or j.n.d. quantisation steps were possible at the presentation levels (40 and 50 dB) as the coarser steps at the lower presentation levels were sometimes equivalent to 1 bit quantisation and this often resulted in no signal over substantial portions of many tokens. This gave 68 sets of data in all.

The test items were 11 vowels in an /h_d/ frame and 19 consonants in a CV frame (V=/a:/) spoken by a speaker of Australian English. For reasons of space only the vowel data are presented in this paper. These tokens were recorded to professional audio standards in an echo free room digitised and vocoded on a VAX computer. The tests were conducted in a sound treated room using calibrated TDH-49 headphones with standard cushions and circumaural seals. The test tokens were presented unmasked at the four presentation levels referred to above. There were 20 listening subjects per condition with each subject being presented with all quantisation steps for a particular quantisation scale at two of the four presentation levels. The 120 listeners were all native speakers of Australian English and none had a history of hearing or speech pathology and all were screened with a speech discrimination test which ensured that they were reliably able to identify monosyllabic words presented at 40 dB s.p.l. Relevant pairs of intelligibility conditions and classes were compared using the chi square test and tested for significant difference at the 0.01 level

3. RESULTS AND DISCUSSION

The intelligibility results are displayed in figures 6, 7 and 8. It is clearly evident that dB quantisation has the most consistent effect across the four presentation levels whilst with the Sone and the j.n.d. quantisation vary in their effects on intelligibility at the different presentation levels.

For all 4 presentation levels, the coarsest dB quantisation step (figure 6) that is tolerated without any significant change in intelligibility is 16 dB. This seems surprisingly large but it is clear, upon examination of the vowel spectra in figures 4 and 5 that the dips between the formant peaks are of sufficient depth (typically around 20 dB) that 16 dB quantisation would still resolve the formant peaks.

The Sone quantised results (figure 7) show a consistent trend with respect to the presentation level. If the presentation level is increased by 10 dB (ie. from 40 to 50 dB) the coarsest tolerated Sone quantisation step doubles (from 0.8 Sones to 1.6 Sones). When the presentation level is increased by 20 dB (ie. from 50 to 70 dB) the coarsest tolerated Sone quantisation step quadruples (from 1.6 Sones to 6.4 Sones). These quantisation levels represent equivalent dB step sizes at each of the presentation levels. In other words, 0.8 Sones at 40 dB, 1.6 Sones at 50 dB, 6.4 Sones at 70 dB (and possibly 25.6 Sones at 90 dB) represent quantisation steps of about 16 dB in the vicinity of the major vowel spectral peaks.

The intensity-j.n.d. quantised results (figure 8) are similar to the Sone results in some respects. Clearly, coarser j.n.d. quantisation steps are tolerated as the presentation level increases, and further the coarsest tolerated j.n.d. steps represent quantisation steps of about 16 dB in the vicinity of the major spectral peaks. The relationship, however, between presentation level and the coarsest tolerated j.n.d. quantisation step is not as clear as it was with the Sone conditions. There is, for example, no doubling of the coarsest tolerated j.n.d. quantisation step as the presentation level is increased by 10 dB.

It is worth noting that a great deal of the measurement of human loudness scales has involved judgements of halving and doubling loudness and that such psychoacoustic measurements have then been used to develop scales of equal loudness. It may actually be more appropriate to utilise instead a scale of loudness doubling (analogous to the octave in pitch perception). When the sone scale is converted into a Sone doubling scale, or a logSone scale,(figure 9) it can be seen that there is an almost linear relationship between logSones and deciBels. The logSone contours shown in figure 9 have been derived by displaying the 0 Sone and the 0.1 Sone contours and then the 0.2, 0.4, 0.8, 1.6, etc... Sone contours (ie. sequential doublings of 0.1 Sones).

4. CONCLUSIONS

The above results strongly suggest that when developing an auditory model of speech little will be gained by utilising quantisation scales based upon Sones or amplitude-j.n.d.'s as the actual size of the quantisation step (in terms of the units on those scales) is highly dependent upon presentation level.

It seems highly likely that the logSone scale represents the most valid quantisation scale because its effects are constant at different presentation levels and the scale itself is derivable from psychoacoustic measurements. The deciBel scale is, however, a very close approximation of the logSone scale and it has the great virtue of being very easy to determine.

It is necessary, however, to point out that the quantisation scales so selected are only indications of the degree of quantisation fineness required at intensities in the vicinity of the major spectral cues. The strength of the raw Sone scale is that once the position of the peaks (in terms of amplitude and with respect to the proposed or actual presentation level) has been determined it is possible to select a quantisation step that provides the appropriate information at the formant intensities and yet doesn't "waste" quantisation steps at intensities below the peaks.

5. REFERENCES

[3] FLETCHER, H. (1953), *Sound and Hearing in Communication*, Van Nostrand.

[4] FLETCHER, H. & MUNSON, W.A. (1933), "Loudness, its definition, measurement and calculation", *J.A.S.A.* 5, 82-108.

[5] GULICK, W.L. (1971), *Hearing: Physiology and Psychoacoustics*, Oxford.

[6] KINSLER, L.E., FREY, A.R., COPPENS, A.B. & SANDERS, J.V. (1982), *Fundamentals of Acoustics* (3rd edn.), New York: John Wiley.

[7] MANNELL, R.H., & CLARK, J.E. (1991), "A comparison of the intelligibility scores of consonants and vowels using channel and formant vocoded speech", in *Proc. XII ICPhS*

[8] MOORE, B.C.J. & GLASBERG, B.R. (1986), "The role of frequency selectivity in the perception of loudness, pitch and time", in MOORE, B.C.J., *Frequency Selectivity and Hearing*, London: Academic Press

[10] ROBINSON, D.W. & DADSON, R.S. (1956), "A re-determination of the equal-loudness relations for pure tones", *British J. Applied Physics* 7, 166-181.

[11] STEVENS, S.S. (1938), "A scale for the measurement of psychological magnitude: loudness", *Psychol. Rev.*, 43, 405-416.

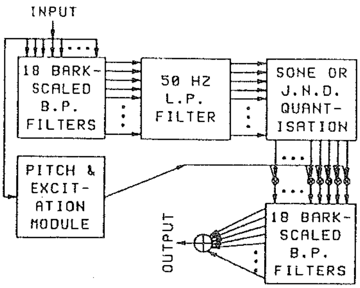

FIGURE 1. CHANNEL VOCODER

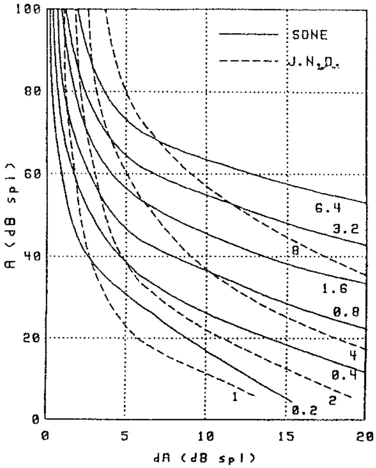

FIGURE 3. AMPLITUDE QUANTISATION CURVES (REFERENCE LEVEL [A] VS. QUANTISATION STEP [dA]) FOR VARIOUS LEVELS OF SONE AND AMPLITUDE J.N.D. QUANTISATION

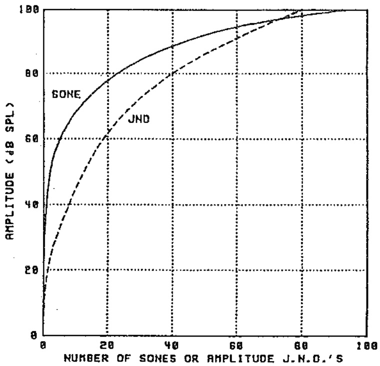

FIGURE 2. AMPLITUDE VERSUS NUMBER OF SONES OR AMPLITUDE J.N.D.'S BETWEEN EACH AMPLITUDE AND THRESHOLD AT 1000 HZ

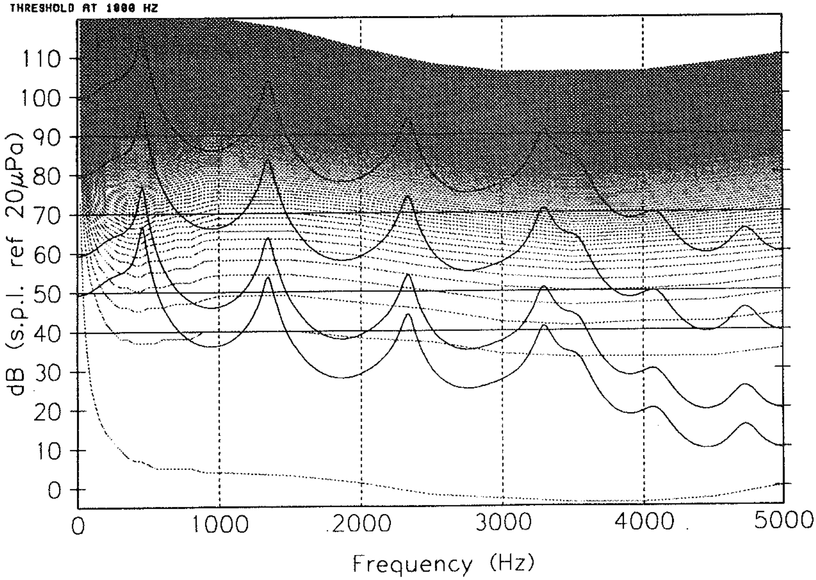

Figure 4 LPC spectra at 4 presentation levels (40, 50, 70, 90 dB s.p.l. ref 20μPa) compared to 1 Sone contours.

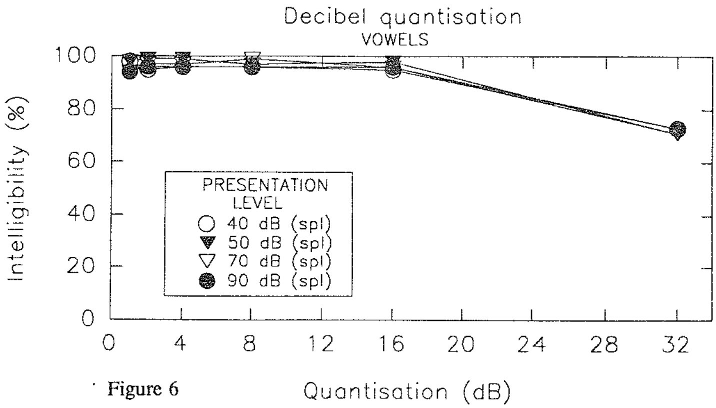

Figure 6

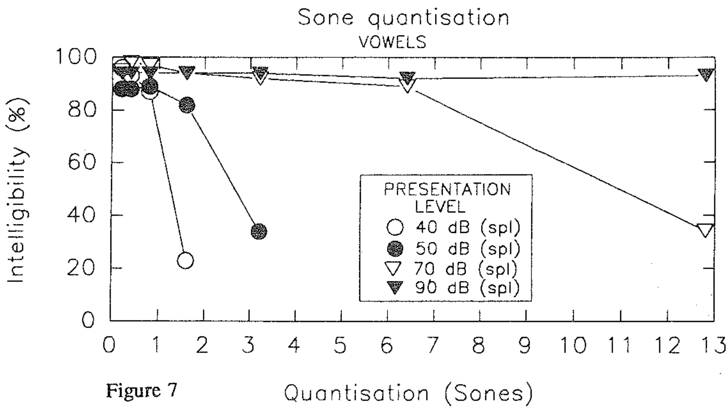

Figure 7

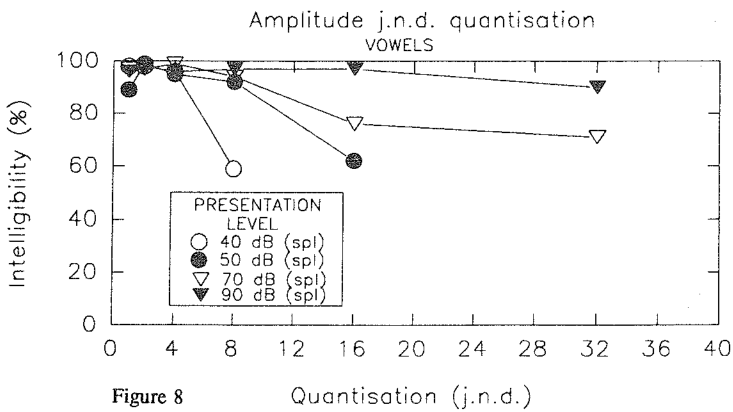

Figure 8

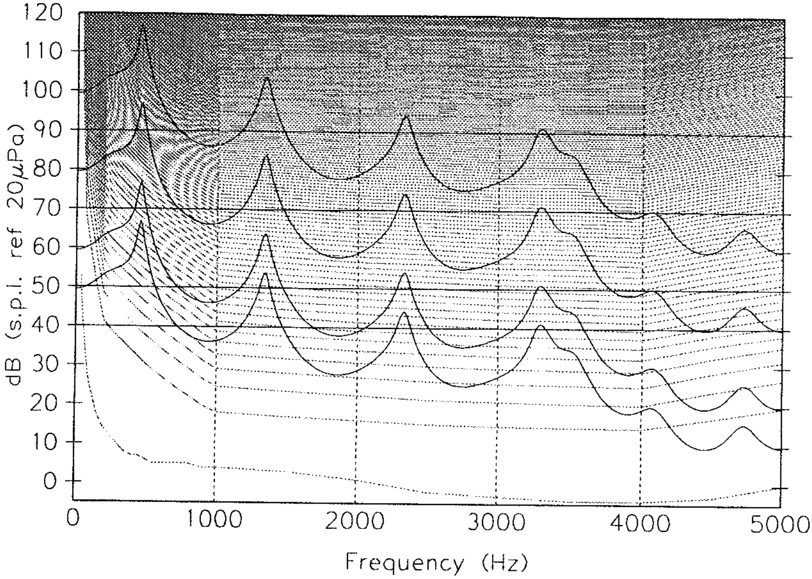

Figure 5 LPC spectra at 4 presentation levels (40, 50, 70, 90 dB s.p.l. ref 20μPa) compared to 1 j.n.d. contours.

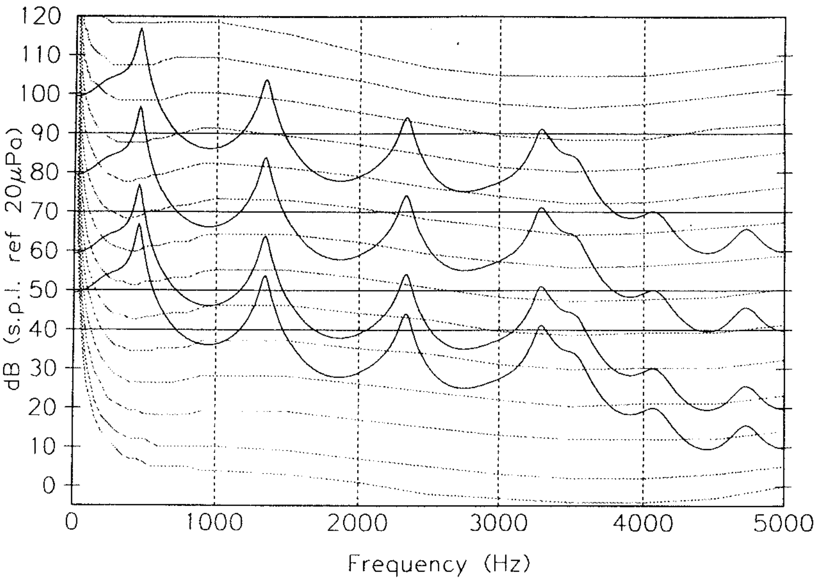

Figure 9 LPC spectra at 4 presentation levels (40, 50, 70, 90 dB s.p.l. ref 20μPa) compared to logSone contours created by sequentially doubling the 0.1 Sone contour.