SONE-SCALED AND INTENSITY-J.N.D.-SCALED SPECTRAL 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 the 18 B.P. analysis filters were quantised at various multiples of the Sone scale and the intensity-j.n.d.-scale. The resulting synthetic speech was presented to a group of listening subjects and intelligibility scores were obtained for each type and level of quantisation. The results suggest that the Sone scale is preferable to the intensity j.n.d. scale at mid frequencies where many important speech cues are to be found.

1. INTRODUCTION

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 equivalence 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 each other and 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. 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. This may explain the tendency for people working with cochlear implants to quote implant performance for individual subjects in terms of j.n.d.'s relative to the overall dynamic range [1].

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]

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

(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 tokens were quantised at the output of the analysis demodulation LP filters at 4 different j.n.d. levels (1, 2, 4 and 8 j.n.d.s') and at 6 different sone levels (0.2, 0.4, 0.8, 1.6, 3.2 and 6.4 sones) as well as a "normal" 16 bit quantisation utilising the same filters and forming the benchmark condition. This gave 11 sets of data in all. The quantisation curves (at 1000 Hz) for the 4 j.n.d. and the 6 sone conditions are shown in figure 3.

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. 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 presented unmasked at 70 dB s.p.l. (ref. 20 uPa). The 20 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 for the 11 test conditions are shown in figures 4 and 5 for various phonetic classes. Figure 4 indicates that even the greatest levels of quantisation do not achieve a great deal of intelligibility loss for the vowels (as a class). The intelligibility of the vowels for the 3.2 and 6.4 sone conditions are nevertheless significantly lower the that of the 0.2, 0.4, and 0.8 sone conditions. It must also be noted that at about this level of quantisation the quality of the vowels deteriorates dramatically and they sound like they are spoken under water. It is interesting to note that some cochlear implant patients comment that the speech that they hear via their implant sounds like it is being spoken under water. For consonant intelligibility the 8 j.n.d. condition is significantly lower in intelligibility than the 16 bit condition whilst the 1.6, 3.2 and 6.4 sone conditions were significantly lower than the 16 bit and 0.2 sone conditions. An examination of both the curves and the above statistics suggests that the 8 j.n.d. condition may be equivalent in its effects on consonant intelligibility to either the 1.6 or the 3.2 sone conditions.

An examination of the results shown in figure 5 indicate fairly clear patterns for three of the classes (the stops being difficult to interpret). For the fricatives, 1.6, 3.2 and 6.4 sone results are significantly lower than the 16 bit and 0.2 sone results whilst the 8 j.n.d. results are significantly less than the 16 bit and 1 j.n.d. results. Examination of the curves suggests that for the fricatives the 8 j.n.d. condition seems to produce equivalent results to the 1.6 sone condition. For the nasals, similar results occur with significant drops in intelligibility at 1.6 sones and 8 j.n.d.'s and it would seem that the 0.8 sone and 4 j.n.d. conditions are equivalent in their effects upon intelligibility. In the case of the continuants the 6.4 sone condition is significantly lower than the 0.2 sone condition whilst all of the j.n.d. conditions are not significantly different. The equivalent points on these two curves appear to be the 3.2 sone and the 8 j.n.d. conditions.

In summary, for all phonetic classes there is no significant deterioration in intelligibility from 1 to 4 j.n.d.'s and from 0.2 to 0.8 sones. These conditions also show very little degradation (relative to the 16 bit case) in overall speech quality. Intelligibility deteriorates between 0.8 and 1.6 sones and between 4 and 8 j.n.d.'s and evidence from both the statistics and the intelligibility curves suggests that the 4 j.n.d. condition is approximately equivalent to either the 0.8 or the 1.6 sone condition. Neither the 0.8 sone nor the 4 j.n.d. condition ever display a significant drop in intelligibility relative to the 0.2 sone or the 1 j.n.d conditions respectively. These conditions can be considered the maximum levels of quantisation allowable before the intelligibility significantly deteriorates (at least for some classes) and they are also the coursest levels of quantisation that do not show a noticeable drop in speech quality. An examination of figure 3 indicates that the 4 j.n.d. curve intersects the 0.8 j.n.d. curve a little below 40 dB and that it intersects the 1.6 sone curve at about 50 dB (at 1000 Hz for a presentation level of 70 dB). This implies that the maximum allowable quantisation level is determined by the degree of quantisation down to about 40 dB and that about one quantisation step is all that is required below this level. It seems that the maximum degree of quantisation allowable before intelligibility and quality deterioration occurs is around 1 sone and that at the minimum intensity for which there appear to be significant cues (ie. down to about 40 dB) the j.n.d. curve which matches the 1 sone curve the closest is the 4 j.n.d. curve.

A 70 dB presentation level was chosen for several reasons. Firstly, it is a comfortable listening level corresponding to the level of normal conversation. Secondly, the shape of iso-response auditory nerve tuning curves have consistent shape up to about 70 dB but increasingly distort above that level as saturation occurs [9]. Further, Dowell et al [2] found 60-70 dB but not 80 dB to be good presentation levels for cochlear implants. These similar figures imply that auditory nerve saturation is the limiting factor for both normally-hearing and cochlear implant subjects. It is reasonable to assume that we have adapted our normal speech levels to make use of that part of the intensity range (70 to 40 dB s.p.l.) where there is both sufficient intensity to pick up important cues up to 30 dB below the speech level and yet the intensity is not so high as to cause distortion of those cues through auditory nerve saturation.

It must be stressed that the curves at 1000 Hz are a fairly good representation of the sone and j.n.d. scales between 1000 and 4000 Hz, however as the frequency drops to 200 Hz or rises to about 10,000 Hz the 1 sone and the 1 j.n.d. curves become almost equivalent over the range of 40 to 70 dB. Many cues occur, however, in the frequency range where the curves in figure 2 and 3 apply and so the number of quantisation levels available would need to be determined from either the 1 sone or the 4 j.n.d. scale.

When cochlear implant performance is defined in terms of the number of j.n.d.'s in an overall dynamic range (eg. [1] dynamic range 2.6 to 16.4 dB and difference limens 0.2 to 0.8 dB) the number of available quantisation levels may actually be one quarter that implied by the quoted figures. For example, a dynamic range of 16 dB with difference limens of around 0.8 dB seems to imply the existence of 20 quantisation levels whilst it may be that there are only 5 quantisation levels available.

4. CONCLUSIONS

It seems that for much of the frequency range the maximum amount of quantisation that will not result in significant drops in intelligibility for at least some phonetic classes is 1 sone. In this frequency range and for the range of intensities that appear to contain most speech cues (70 to 40 dB for presentation levels of 70 dB) the 4 j.n.d. curve appears to produce similar results to the 1 sone quantisation level. These results support the notion that the reference point in sone calculations (40 phons or 40 dB at 1000 Hz equal to one sone) is not an arbitrary reference point but may be related to the effective data quantisation that occurs in the process of human speech perception. This is not a surprising finding when one realises that there is a power relationship between sones and phons above 40 phons (1 sone) but not below that point. It is reasonable that we would adapt our speech perception to the intensities with a more stable relationship to loudness.

5. REFERENCES

[1] BUSBY, P., TONG, Y.C., & CLARK, G. (1990), "Psychophysical studies on cochlear implant patients with early onset of profound hearing impairment", paper given at *Tactile Aids, Hearing Aids and Cochlear Implants: An International Conference*, Sydney.

[2] DOWELL, R., SELIGMAN, P., & WHITFORD, L. (1990), "Speech perception with the 22-channel cochlear prosthesis: A summary of ten years development", paper given at *Tactile Aids, Hearing Aids and Cochlear Implants: An International Conference*, Sydney, May 1- 3, 1990

[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

[9] PICKLES, J.O. (1986), "The neurophysiological basis of frequency selectivity", in MOORE, B.C.J., *Frequency Selectivity and Hearing*, London: Academic

[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.

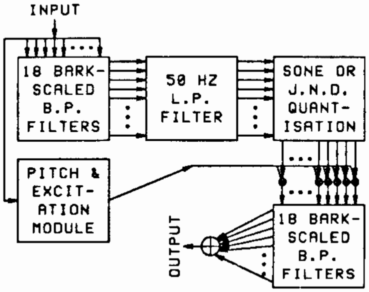
INPUT
18 BARK-SCALED B.P. FILTERS
50 HZ L.P. FILTER
SONE OR J.N.D. QUANTISATION
PITCH & EXCITATION MODULE
18 BARK-SCALED B.P. FILTERS
OUTPUT

FIGURE 1. CHANNEL VOCODER

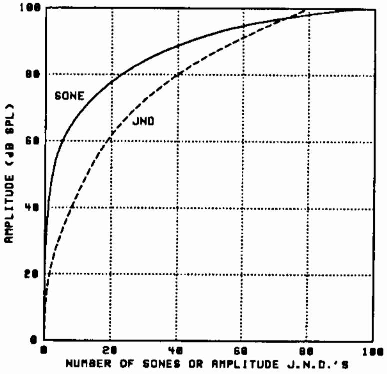

FIGURE 2. AMPLITUDE VERSUS NUMBER OF SONES OR AMPLITUDE J.N.D.'S BETWEEN EACH AMPLITUDE AND THRESHOLD AT 1000 HZ

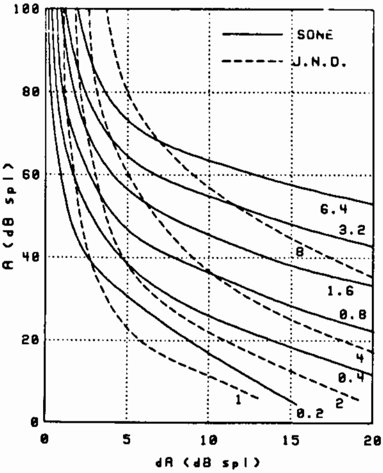

FIGURE 3. AMPLITUDE QUANTISATION CURVES (REFERENCE LEVEL [A] VS. QUANTISATION STEP [dA]) FOR VARIOUS LEVELS OF SONE AND AMPLITUDE J.N.D. QUANTISATION

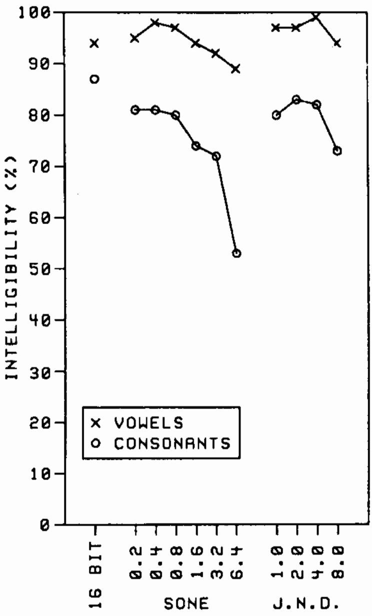

FIGURE 4. INTELLIGIBILITY SCORES FOR QUANTISED VOWELS & CONSONANTS

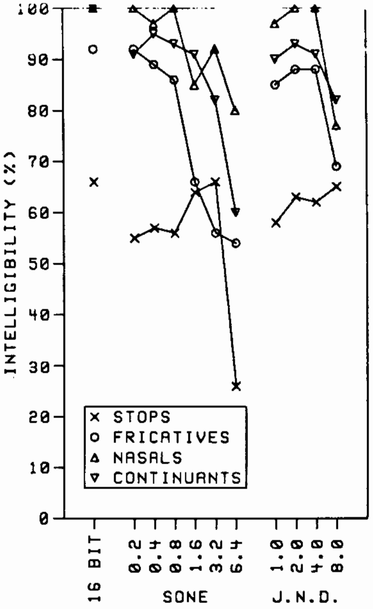

FIGURE 5. INTELLIGIBILITY SCORES FOR STOPS, FRICATIVES, NASALS, AND CONTINUANTS