PERCEPTION OF VCV's IN A FLUCTUATING NOISE ENVIRONMENT

Howard-Jones, P.A.

Department of Phonetics and Linguistics, University College London, Gower St., London, WC1E 6BT.

1. ABSTRACT

Two experiments are described which investigate the intelligibility of a small set of spoken VCV's when presented in fluctuating noise environments to normally hearing subjects. The masking effects of three noise environments, which differ with respect to the time-scale of the fluctuation of their power, are compared with those of white noise.

The first experiment compared the masking effectiveness of two samples of fluctuating broadband noise with that of white noise. One of these samples was 'cubed' noise whose power fluctuated on an instantaneous time-scale. The other fluctuating noise possessed random interruptions of 80 to 100's of milliseconds. The second experiment compared the masking characteristics of a sample of white noise with the same sample which is attenuated every other second, for a second. All the noise samples used in these experiments exhibited flat long-term frequency spectra. In both experiments, adaptive test procedures were used to determine the Speech Reception Threshold (SRT) associated with each sample of noise, and percentage-correct scores over a range of S/N levels were used to plot Performance-Intensity (PI) functions.

Results showed that, for this speech material, instantaneous fluctuations have little influence upon masking effectiveness, whilst the two noises which fluctuated over longer time-scales were less effective maskers than white noise (at the same r.m.s.).The third fluctuating noise, whose r.m.s. changed every second, produced a flattened PI function. The types of phonetic confusions associated with this noise were more random, and normally robust features, such as voicing, were more often undetected.

2. INTRODUCTION

Most studies of speech intelligibility in noise have involved the use of white noise maskers[1,2],speech babble[3,4] or cafeteria[5,6] noise. A common approach[7] is to measure the long-term spectrum of the speaker to be used in the tests, or of a multi-speaker speech corpus, and to filter white noise until it displays a similar long-term frequency content. However, all these noises are better behaved than many types of noise environments often encountered in everyday situations, when the background noise tends to suffer large fluctuations in power and frequency content.

Miller and Licklider(1950)[8] made an early study of the effects on speech perception of an interrupted broadband masker. Speech scores were measured for noises that were interrupted at various rates with various noise-time ratios. In efforts to improve and validate the Articulation Index as a procedure to estimate speech intelligibility, Kryter (62) was able to include the effects of interrupted noise maskers, such as those used by Miller and Licklider.

The investigations reported here derive from an attempt to study more closely the effects of fluctuating noise backgrounds on speech perception. Since Miller and Licklider, new types of audiometric tests, such as those based on adaptive procedures[9], have become popular in speech audiometry. The experiments discussed below were designed to test the feasibility of applying such procedures in studies involving fluctuating noise backgrounds, as well as to investigate further the types of effects that may be identified.

2. EXPERIMENT A

2.1 Noise and speech material used in Experiment A

There were three samples of noise used in Experiment A. The first was a 4 second sample of Gaussian white noise (Fig. 1(a)) sampled at 12.8 kHz. The second was 4 seconds of 'cubed' white noise (Fig. 1 (b)). This noise was created by digitally processing the white noise sample, raising its instantaneous amplitude by the power three, and then scaling the signal down. Such a noise signal retains its flat long-term power spectrum, but is more 'peaky' (on an instantaneous time-scale) than the white noise from which it is derived and sounds like a 'crackling' noise background. Fig. 1 (d) shows a detail of the cubed noise waveform compared with the white noise waveform that was used to produce it. The third noise was also created digitally, by treating a copy of the original noise sample in 80 ms time slices, a random selection of which were attenuated by 20 db (see Fig. 1 (c)). The signal was scaled such as to possess the same r.m.s as the original signal. Again, the processed signal (of 4 seconds duration) exhibited a flat frequency spectrum but this fluctuating noise sample contained interruptions with durations and spacings in the range 80 to 400 milliseconds.

The speech material used in these tests was restricted to one example each of twelve intervocalic consonants [m b p v f n d t z s g k] as in [ɑ'mɑ] and spoken by a female RP speaker. Data was sampled at 12.8 kHz.

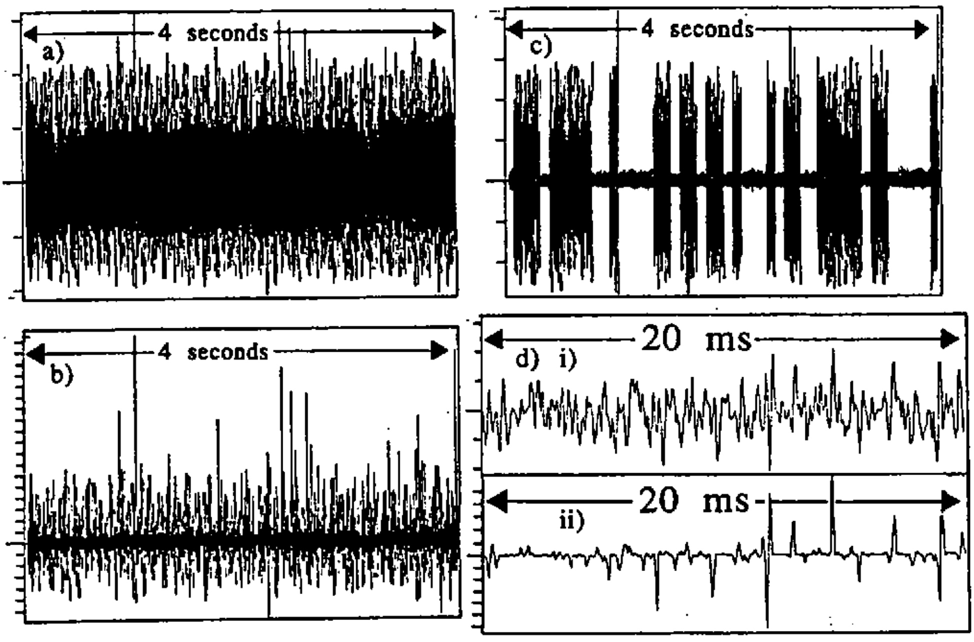

Fig. 1 Noise samples used in Expt. A: (a) the white noise sample, (b) the 'cubed' noise noise sample (see text) (c) the noise sample with interruptions (of duration and spacing of 80 - 400 ms) (d) (i) Detail of the white noise and (ii) of the 'cubed' noise which was derived from it. [Vertical gradations in Figs. 1 & 4 refer to same arbitary amplitude scale].

2.2 Methodology

The experiment consisted of two parts: determination of the SRT using adaptive test procedures, and measurement of the intelligibility of the same test material over a range of S/N levels to enable the plotting of a PI function.

2.2.1 Determination of the SRT using adaptive test procedures. An adaptive 'up-down' procedure was used to determine the SRT (threshold for 50% correct response). In such tests the level of the noise background is turned down or up, depending on whether the response was correct or not. The subject's performance should eventually settle at 50%, and the average S/N level can be used as an estimate of the SRT.

The adaptive computer-controlled test procedures used a MASSCOMP to play the speech (with noise) at 12.8 kHz (via a 5.6 kHz lowpass filter) through a pair of Sennheiser HD414SL headphones to a subject sitting in a quiet room. The subject was requested to enter his responses via a keyboard. During the test, each speech token,selected according to a randomised sequence, was mixed with the noise. The S/N ratio was controlled by scaling the noise sample prior to mixing it with the speech and outputting. Only one sample of each type of noise was used in these experiments, but each time the speech was mixed with the appropriately attenuated noise, a random start point in the four second noise file was chosen for the speech data. The level of the noise was increased by 1 dB when the subject selected the correct consonant from the list of twelve, and it was reduced by 1 dB when an incorrect response was given. Results from a pilot test were used to derive initial S/N levels appropriate for each type of noise. Before testing began, the level of the speech was set such that both Experiments A and B could be performed at a comfortable overall level.

Each test used each of the 12 consonants four times, randomly arranged into a sequence of 48 tokens. Each of 6 subjects was asked to perform two tests for each of the 3 noises (6 tests for each of the 6 subjects). The tests were arranged such that each subject experienced one of the six possible permutations of presentation order twice. A value for the SRT was derived from each test by calculating the average of the last 36 presentation levels.

2.2.2 Determination of the PI Function. This was facilitated using the same arrangement of hardware as outlined above. Here, however, each test was carried out at a constant S/N ratio, and the percentage correct score was recorded. Six new subjects each performed six tests with each of the three noises at +12, +6, +3, 0, -3 and -6 dB relative to a previously estimated SRT. The three batches of six tests were arranged amongst the subjects such that each subject experienced one of the six possible permutations of presentation order.

2.3 Results and Discussion (Experiment A)

The SRT's for each noise, from the adaptive test results averaged over the six subjects, are given in Table 1. The results of the second part of the experiment were combined to produce the PI functions shown in Fig. 2. (N.B. All S/N ratios reported in this document, unless otherwise stated, are with reference to the same arbitrary scale).

Table 1. SRT's for noises in Experiment A, derived using adaptive test procedures.

Noise Type	Speech Reception Threshold (dB)	Estimated Error (dB)
White	-17.45	+/- 0.51
'Cubed'	-17.00	+/- 0.32
Interrupted (80 - 320 ms)	-21.66	+/- 0.62

The results of the adaptive tests tend to imply that the masking effectiveness of the cubed noise is similar to that of white noise. The interrupted noise appears less effective than both. Inspection of the PI functions associated with the three noises confirms this difference in masking effectiveness. (Differences, in the absolute sense, between SRT's measured by the two methods may be due to subject group, but are more likely the result of practice effects: measurement of the PI function allows the subject to gain more familiarity with the test material, and performance measured by this method generally appears better.) Miller and Licklider

(1950) also found that, when compared at similar r.m.s. values, interrupted noise did not mask speech as well as continuous noise. They explained: "When there were 10 bursts of noise per second, the listeners were able to get several glimpses of every word and to patch these glimpses together .... ".

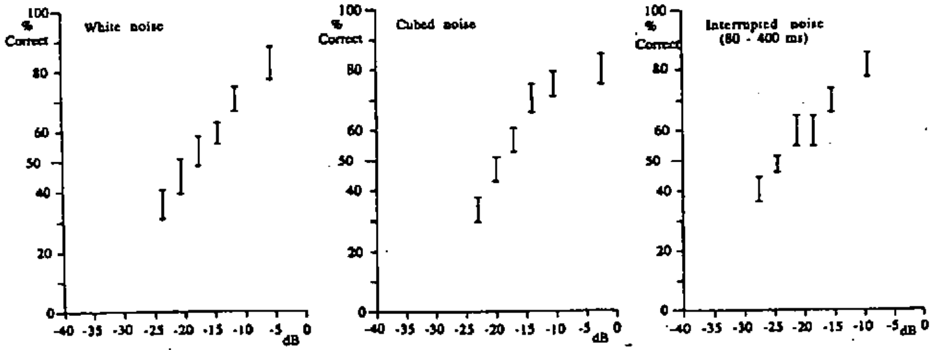

Fig. 2 The PI functions derived from combining results for six subjects listening to the test material masked by the three noise samples used in Experiment A.

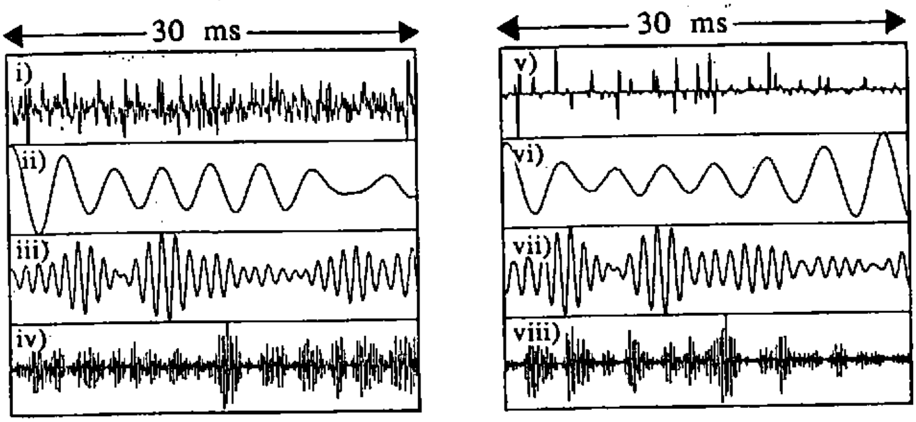

Fig. 3 The waveforms that result from passing white noise and cubed noise through digitally simulated auditory filters with various centre frequencies (f_c) The filtered waveforms are shown underneath the noise samples from which they are derived [(i) White noise and the same noise after filtering with f_c = (ii) 250Hz, (iii) 1 kHz and (iv) 5 kHz; (v) Cubed noise and the same noise after filtering with f_c = (vi) 250 Hz, (vii) 1 kHz and (viii) 5 kHz].

Both parts of Experiment A imply that the cubed noise behaves similarly to white noise as a speech masker, despite the severe (instantaneous) fluctuation of its power. This could be predicted using a model of the auditory system that consists of a parallel set of overlapping bandpass filters. Such filters have a temporal response which will 'smooth' out any rapid, instantaneous fluctuations, except, possibly, in the higher frequency channels, when the impulse responses of the filters are faster. Samples of the white and cubed noise used in this experiment were passed through a digital simulation of three auditory filters [10] (with centre frequencies at 250 Hz, 1000 Hz and 5000 Hz) and the resulting waveforms are shown in Fig. 3. Instantaneous fluctuation can be measured using the fourth moment of the waveform[11] and such an analysis was performed on the output of these filters. Only in the channel with the higher centre frequency of 5 kHz could differences be detected between the fluctuations caused by cubed noise and ordinary white noise. Such differences are unlikely to influence speech perception in this experiment (especially since all material was played out through a 5.6 kHz lowpass filter).

A rudimentary phonetic analysis of the results did not reveal any obvious differences between the types of confusions made by subjects listening against the three noise backgrounds. It had been thought that the interrupted noise might cause subjects to make confusions that were more disparate with respect to their phonetic features. An increase in the randomness of the subject's decision-making might also have led to flattening of the PI function. Since neither of these effects were observed, it was thought worthwhile to run a second experiment, using noise which fluctuated on an even longer time-scale.

3. EXPERIMENT B

3.1 Noise and speech material used in Experiment B

There were two samples of noise compared in the second experiment. The first was an 8 second sample of Gaussian white noise (Fig. 4 (a)) sampled at 12.8 kHz. The second was created from the first by attenuating every other one second section by 30dB (see Fig. 4 (b)).The speech material was identical to that used in the first experiment.

3.2 Methodology

As in Experiment A, the SRT was determined using adaptive procedures and PI functions were obtained using the same experimental protocol as before. This time, however, the PI functions were determined using 4 subjects listening to test material at +12, +6, +3, 0, -3, -9, -15 dB relative to a previously estimated SRT.

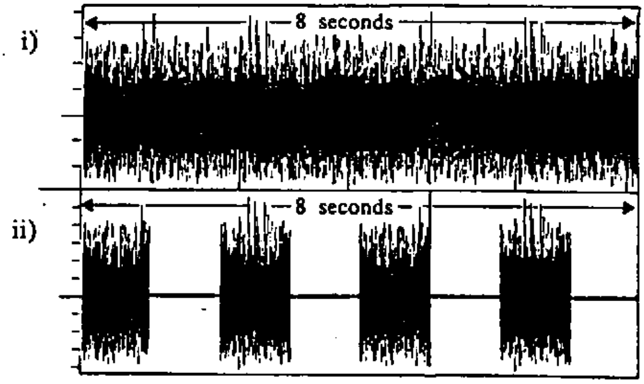

Fig. 4 Noise samples used in Experiment B: (i) White noise and (ii) the same noise with regularly spaced interruptions of one second duration, every other second.

3.3 Results and Discussion (Experiment B)

The SRT's for the two noises, averaged over the six subjects used in the adaptive test, are given in Table 2. The PI functions for the two noises are shown in Fig. 5.

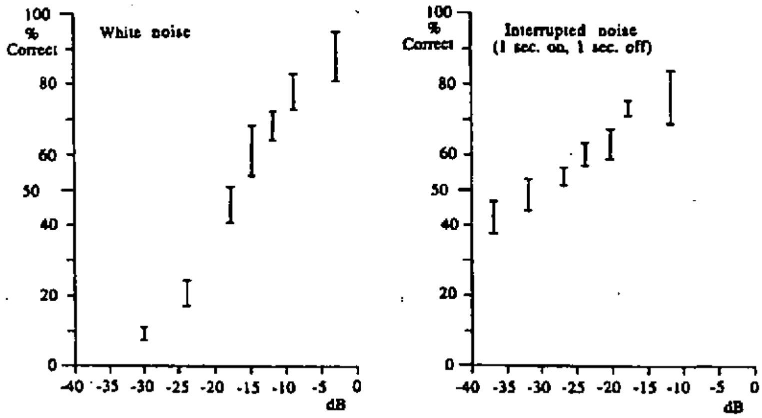

Fig. 5 The PI function for white noise and interrupted (1s. on/off) noise , derived by combining results for four subjects listening to the test material over a range of S/N levels.

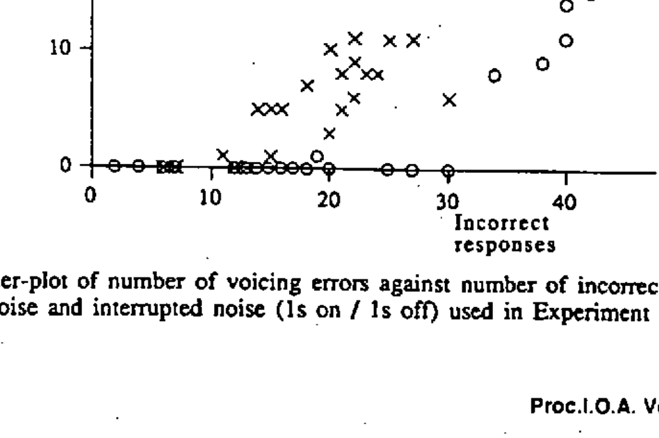

Fig. 6 Scatter-plot of number of voicing errors against number of incorrect responses for the white noise and interrupted noise (1s on / 1s off) used in Experiment B.

VCV's IN FLUCTUATING NOISE

Table 2. SRT's for noises used in Experiment B, derived using adaptive procedures.

Noise type	Speech Reception Threshold (dB)	Estimated error (dB)
White noise	15.05	+/- 2.73
Interrupted (1s on/off)	23.79	+/- 2.90

Again, the results of both parts of the experiment associate a release of masking with the interrupted noise. The 'glimpsing' effect, referred to above, can still be used to explain the general release of masking observed for this interrupted noise. In this case, it may be caused by a consonant still being intelligible, even when most of it falls within the louder burst, the lesser part of it proving sufficient for identification.

It should be pointed out, however, that some of the results of the adaptive testing were judged to be less reliable in Experiment B. Although the SRT's determined for each subject were not distributed over a very great range, the subject's progress during the adaptive test with this interrupted noise was markedly more random than with white noise. The levels prompted by the subjects' responses did not always settle very convincingly. Inspection of Fig. 5 reveals a flattening of the PI function for subjects listening to speech perceived in this fluctuating noise, which indicates that performance is less sensitive to changes in S/N level. This will detract from the efficiency of an adaptive test which may need to be extended, with respect to the amount of speech material presented to each subject, before an SRT can be confidently derived.

The nature of the fluctuating noise sample used in this experiment is such that for a range of overall S/N levels, half of the speech can be almost completely masked whilst that half of the speech occurring between the louder bursts remains reasonably intelligible. Depending upon the difference in level between the 'on' and 'off' parts of the noise, a plateau in the PI function can form around the 50% intelligibility region and cause levels in the adaptive test procedures to vary across a wide range. (The criteria for a correct response depending chiefly upon whether the speech occurs during a quiet section.) The flattening of a PI function by fluctuations in the masker was reported by Speaks et al.(1967)[11] who noticed that the PI function was less steep when he used a competing message as a masker rather than white noise. He explained that this was due to the 'random masking and disruptive features of competing speech'.(Such an effect has unfortunate consequences for anyone trying to improve the S/N level of speech in a noise environment fluctuating in this manner, as it may not greatly improve intelligibility).

Since, at lower S/N levels, many of the speech tokens were completely masked by the louder sections of the fluctuating noise masker, the types of phonetic confusions observed for these conditions were predictably more random. A scatter-plot of number of voicing errors against number of incorrect responses is shown in Fig. 6. This diagram illustrates how the normally robust feature of voicing, can be less easily detected in noise that is fluctuating on such a time-scale than in white noise, when both noises are considered at S/N levels that allow comparable degrees of overall intelligibility.

4. SUMMARY OF RESULTS AND CONCLUSIONS

Several types of effects associated with the masking of speech by a fluctuating noise background have been identified (see Table 3) and it has been shown that adaptive test procedures can be practicably applied as part of such investigations. Care should be taken, however, with noises that fluctuate on some longer time-scales, as a flattening of the associated PI function can reduce the efficiency of adaptive procedures and detract from the significance of any results obtained.

Table 3 Summary of results for Experiments A and B.

Noise	Time-scale of fluctuation	Masking	PI slope	Phonetic effects
Cubed	Instantaneous	<------------As white noise----------->		
Interrupted	80ms - 400 ms	4dB less effective than white noise	As white noise	As white noise
Interrupted	1 sec	Approx. 10dB less effective than white	Flattened	More indiscriminate confusions

5. REFERENCES

1. HAWKINS, J.E. and STEVENS, S.S 'The masking of pure tones and of speech by white noise.' J. Acoust. Soc. Am. 22, 6-13 (1950).

2. KEITH, R.W. and TALIS, H.P. 'The effects of white noise on normally hearing and hearing-impaired listeners' Audiology 11, 177-186 (1972).

3. KALIKOW, D.N. and STEVENS, K.N. 'Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability' J. Acoust. Soc. Am. 61(5), 1337-1351 (1977).

4. GORDON-SLANT, S. 'Phoneme feature perception in noise by normally-hearing and hearing-impaired subjects' J. Sp. Hear. Res., 28, 87-95 (1985).

5. COOPER, J.C., CUTTS, B.P. 'Speech discrimination in noise' J. Sp. and Hear. Res. 14, 332-337 (1971).

6. DIRKS, D.D.,MORGAN, D.E., DUBNO, J.R. 'A procedure for quantifying the effects of noise on speech recognition' J. Sp. Hear. Disorders, 47, 114-123 (1982).

7. FOSTER, J.R. and HAGGARD, M.P. 'The four alternative auditory feature test (FAAF) - linguistic and psychometric properties of the material with normative data in noise' Br. J. Aud. 21, 165-174 (1987).

8. MILLER, G.A. and LICKLIDER, J.C.R. 'The intelligibility of interrupted speech' J. Acoust. Soc. Am., 22(1), 167-173 (1950).

9. LEVITT, H. 'Transformed up-down methods in psychoacoustics' J. Acoust. Soc. Am. 41, 467-477 (1971).

10. PATTERSON, R.D.,CUTLER, A. 'Auditory preprocessing and recognition of speech' in 'Cognitive Psychology - Research Directions in Cognitive Science' eds. A. Baddeley and N. O. Bernsen (1989).

11. HARTMAN, W.M. and PUMPLIN, J. 'Noise power fluctuations and the masking of sine signals' J. Acoust. Soc. Am. 83(6), 2277 - 2289 (1988).

12. SPEAKS, C. KARMEN, J.L. and BENITEZ, L. 'Effect of a competing message on synthetic sentence identification' J. Sp. Hearing Res. 10, 390 - 396 (1967).