Effect of Noise Reduction on Reaction Time to Speech in Noise

Mark Huckvale, Jayne Leak

Department of Speech, Hearing and Phonetic Sciences
University College London, London. U.K.
m.huckvale@ucl.ac.uk, j.leak@ucl.ac.uk

Abstract

In moderate levels of noise, listeners report that noise reduction (NR) processing can improve the perceived quality of a speech signal as measured on a typical MOS rating scale. Most quantitative experiments of intelligibility, however, show that NR reduces the intelligibility of noisy speech signals, and so should be expected to increase the cognitive effort required to process utterances. To study cognitive effort we look at how NR affects reaction times to speech in noise, using material that is still highly intelligible. We show that adding noise increases reaction times and that NR does not restore reaction times back to the quiet condition. The implication is that NR does not make speech "easier" to process, at least as far as this task is concerned.

Index Terms: intelligibility, quality, noise reduction

1. Introduction

Noise Reduction (NR) processing is being applied more frequently in communication systems, hearing aids and in forensic audio processing. However we still do not have a good scientific account of how human speech perception is affected by noise reduction. Counter-intuitively there are few experiments that have shown that NR improves the intelligibility of noisy speech signals. For example in a wide survey of techniques, Hu & Loizou [1] found that only one combination of algorithm and noise showed any significant improvement in intelligibility among 64 combinations tested. On the other hand, NR processing has been shown to improve listeners' subjective rating of quality on an MOS rating task. For example Hu & Loizou [2] show that MMSE processing [3] can increase overall subjective rating for speech in car-noise at +5dB SNR from a rating of 2.3 to 2.9 (out of 5).

Thus the main benefit from NR could be had in situations where the speech is already at high intelligibility, and where a reduction in any background noise might improve listening comfort, decrease listening effort and reduce fatigue. The question that needs to be addressed is whether NR *actually* confers these benefits, even if listeners in the MOS task report an improvement in signal quality.

We believe that MOS testing alone is insufficient to support the argument that NR improves listening comfort or reduces fatigue. There are many reasons why we should be sceptical of the results of MOS testing performed in the typical manner [4]: (i) only short utterances are used, not complete conversations; (ii) listeners are asked to combine multi-dimensional judgements about the signal into a single number, so we are not sure what aspects of the signal they are using to make their decisions; (iii) listeners are treated as experts in what is best for them, but they may not be aware of the cognitive effort required to process the utterances; (iv) the listening tests are short to explicitly prevent the effects of fatigue. In this study we try instead to measure the effects of NR on cognitive processing directly, by looking at reaction time to recognise spoken words presented in quiet, noise and noise-reduced conditions. Reaction time was chosen as a very simple measure of processing load, but we do not suggest that it is the only way in which such a test could be performed.

2. Effects of Noise on Language Processing

The study of human performance in noise has a long history, and the reviews of Broadbent [5] and Smith [6] generally agree that "noise has a definite effect on performance but that the precise nature of the effect depends on the type of the noise and the task being performed". We are particularly interested in the effect of noise on cognitive load in speech listening tasks where there is no problem with intelligibility. Most relevant studies however, have looked at performance on visual language tasks in noise. We briefly review these to obtain suggestions for how best to measure cognitive load in a language processing task in noise.

2.1. Effect of noise on memory

A number of studies have shown that our ability to remember a list of words is affected by the presence of noise. For example in [7] it was found that any kind of speech-like noise played while a listener attempted to remember a list of words degraded the accuracy of recall. This "irrelevant speech effect" seems a robust phenomenon presumably related to the fact that low-level phonetic and phonological processing systems are recruited by the interfering noise, and this diminishes the effectiveness of those systems to facilitate the memory of the words. It is also clear that the effect is not to do with recognition of the list items (which were presented visually in this experiment), since recall accuracy is also affected when the noise is played only during the remembering interval, and not while the items themselves are shown.

2.2. Effect of noise on comprehension

Although effects of noise on memory seem to be interfering at a phonological level, there is also evidence that noise interferes with comprehension. For example in [8] it was shown that the ability of subjects to answer questions about the contents of a read passage was affected by the presence of audio noise during the task. Both white noise and nonsense speech caused some effect on comprehension, but the largest effect was caused by interference with meaningful speech.

2.3. Effect of noise on processing capacity

It is commonly assumed in psychological studies that the human information processing system has a limited capacity. Such a model is used to explain why cognitive performance degrades either when additional processing demands are placed on the system, or when the system itself becomes compromised. For example [9] uses such an account to explain why older listeners show worse comprehension of speech in noise than younger listeners. The assumption is that speech in noise makes additional demands on the processing system, which then has less capacity for higher linguistic tasks.

2.4. Effect of noise on attention

The effects of noise on attention are mixed. A number of studies have shown that small amounts of noise actually increase subjects ability to attend to the task. On the other hand it is clear that moderate to loud noise, particularly when experienced for a long period is detrimental to attention. [10].

2.5. Selection of measure

To assess the performance of noise reduction systems on the cognitive processing of speech in noise, we need to create some laboratory task that is at the same time: short in duration, easy to perform, and robust to a noise effect, but also: relevant to the kinds of processing that would occur in everyday speech tasks by users of NR systems. One way to get robust results in a short time is to stress the cognitive system, so that any increase in processing load leads to significant falls in processing accuracy or processing speed. Typically such loading has been performed in a dual-task paradigm, whereby both audio and visual modalities are occupied simultaneously. However such tasks are more difficult for subjects to understand, and subjects may choose different strategies to balance the demands of two tasks. Thus we have selected a single-task paradigm. Our research into previous methodologies has suggested the following possible tasks: (i) serial recall of words from memory – where subjects are played a word sequence and then asked to repeat the words from memory after some delay; (ii) comprehension of a read passage – where subjects are asked to answer questions about the events in some story hears some time earlier; (iii) a lexical decision task – where subjects must make a word/non-word decision as rapidly as possible after hearing the word; and (iv) a choice reaction time task – where subjects must choose between a small set of categories for a spoken word as fast as possible.

We have chosen here a choice reaction time (RT) task since it is the simplest of the options proposed. Although recognising a spoken word as fast as possible may not be a typical task for a user of an NR system, any delays in recognising words in an everyday conversational setting will inevitably lead to less processing time available for processing at higher language levels. Thus we believe that RT can stand as one possible "proxy" for user task performance.

3. Experimental Design

3.1. Materials

To reduce any learning effect we chose speech materials that were familiar to subjects and for which there was no ambiguity about response category. We used recordings of the digits 1-9 spoken quickly and with good articulation by one male speaker of British English. This allowed identification responses to be made on a numeric keypad.

We chose two types of interfering noise, typical for target applications of NR: car-noise and multi-speaker babble. The car-noise was recorded in the cabin of a Vauxhall Corsa travelling at 110kph. The babble noise was that provided on the NOISEX CD-ROM [11].

3.2. Conditions

There were five conditions: (i) quiet, (ii) babble, (iii) car noise, (iv) processed babble, (v) processed car noise. For the noise conditions SNR levels were chosen to make the noise clearly audible to the subjects without affecting intelligibility. We estimated the appropriate SNR levels to use by calculating the SII score [12] from the digit and noise recordings, and targeted an SNR which gave an SII score of 0.7, which corresponds to a predicted intelligibility of greater than 90% for this material [13]. This translated into an SNR of +6dB for the babble and –3dB for the car-noise. When mixing the digits with noise, the digits were maintained at a constant level. Five different noise sections were used to produce 5 noisy versions of each digit in each condition which were used randomly in the tests.

For the noise reduced conditions we applied the MMSE algorithm [3] as implemented in the Voicebox MATLAB toolkit [14]. The MMSE algorithm was chosen because of its demonstrated ability to improve MOS scores [2].

During the testing the noise was presented continuously, and not just during presentation of the speech. For the NR conditions, a recording of NR processed babble and NR processed car-noise was played in the background between the digits.

3.3. Measurement

To encourage the subjects to maintain attention during the test and to motivate them to respond as quickly as possible, the response time task was presented as a computer game called the Typometer, see Fig 1.

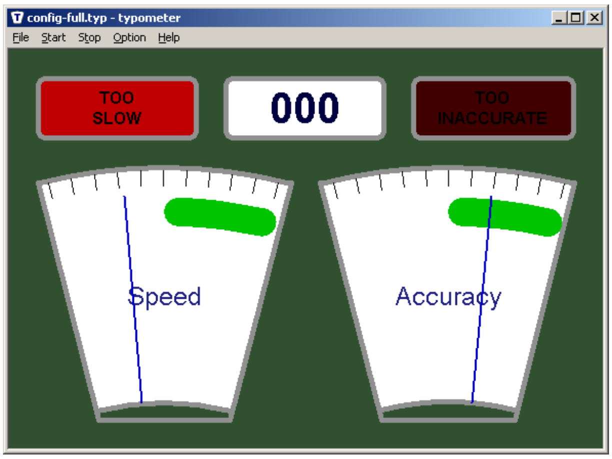

Figure 1: *Typometer user interface.*

The meters reported the subject's current speed and accuracy on the word recognition task. Subjects were encouraged to maintain a minimum speed and accuracy by the system only awarding points when both meters were in the green zone. For speed this meant the mean response time for the last 10 trials was within 1.1s (measured from the start of the digit), for accuracy this meant more than 50% of the last 10 trials were correctly keyed.

Presentations of the digits followed each keyed response with a random delay between 0 and 2.5s. Subjects had up to 2 seconds to respond after the start of the digit, otherwise the response was 'timed out'.

Each subject was tested across the five conditions in sequence within one session. The order of conditions was balanced across subjects using a double latin square which ensured that every condition and every condition dyad occurred in every position. Each condition was run until the subject had recorded a minimum of 10 correctly keyed repetitions of each digit. This took about 5-6 minutes per condition.

3.4. Subjects

Twenty subjects undertook the experiment. These ranged in age between 16 and 49 years (mean 29 years), 30% were male. The test was conducted binaurally over headphones in a quiet domestic environment at the same level for all subjects, which was between 68-76dBA depending on condition. Subjects were not tested for hearing loss, but all reported that they could hear the speech materials clearly.

4. Results

There was no difference in intelligibility across the conditions. The mean proportion of incorrect key presses in each condition, averaged over all subjects, is shown in Table 1.

Table 1. *Average incorrect key presses per condition.*

Condition	% Incorrect
Quiet	2.78
Babble	3.22
Car	2.33
Babble+NR	3.56
Car+NR	3.33

The mean RT for each subject for each condition for each digit was then calculated over the last 10 correct responses. The mean response time per condition is shown in Table 2, and the distribution of times is displayed in Fig.2.

Table 2. *Average mean reaction time per condition.*

Condition	Time (s)
Quiet	0.955
Babble	1.012
Car	1.000
Babble+NR	1.007
Car+NR	0.995

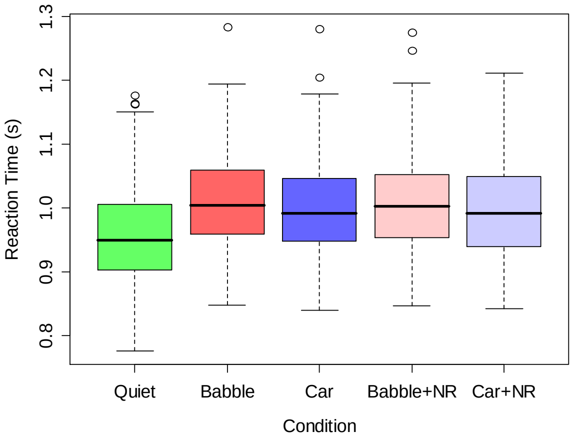

Figure 2. *Distribution of Reaction Time per Listening Condition*

A repeated measures analysis of variance across conditions (5 levels) and digits (9 levels) with subject as a random factor (20 samples), shows a significant effect of condition (F[4,76]=25.7, p<0.001), and of digit (F[8,152]=50.7, p<0.001) on mean reaction time, and also an interaction between condition and digit (F[32,608]=4.5, p<0.001). A post-hoc analysis shows that none of the noise conditions, processed or unprocessed, are significantly different from one another, but all differ significantly from the quiet condition.

To investigate differences across digits, the reaction times were normalised per subject and per digit by subtracting the mean response time for each digit in quiet from each measurement in the noise conditions. The distribution of reaction time increase across digits for babble and babble+NR is shown in Figure 3. The distribution of reaction time increase across digits for car noise and car noise + NR is shown in Figure 4.

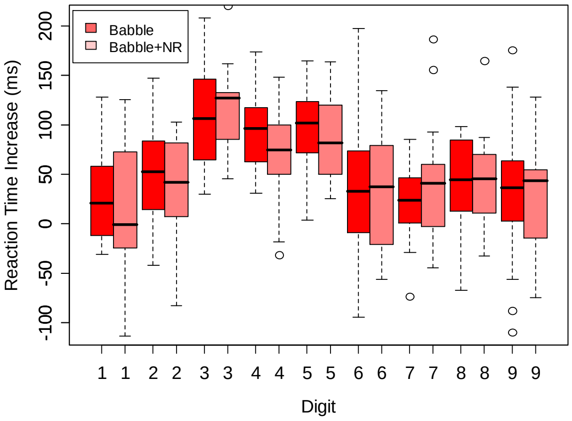

Figure 3. *Distribution of Reaction Time Increase per Digit across Babble conditions*

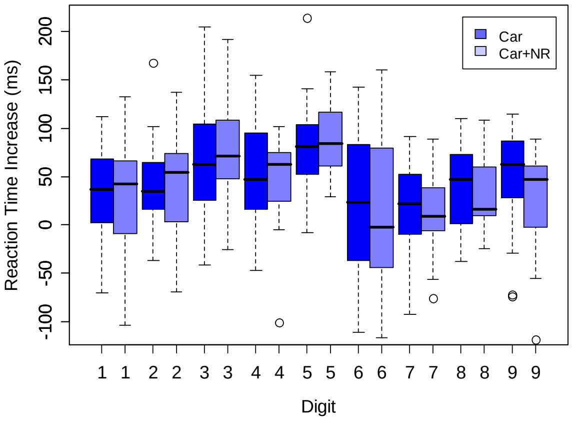

Figure 4. *Distribution of Reaction Time Increase per Digit across Car noise conditions*

Broadly speaking, the figures show that the addition of noise has an effect on all digits (mean RT increase=48ms), but that a larger effect is seen on digits 3,4 and 5. This is probably because these digits start with quiet voiceless fricatives which may have been energetically masked by the noise.

The question then arises whether the general increase in RT caused by the addition of noise is a real effect of additional cognitive load or just the result of the noise masking the start of the words – so that listeners took longer to notice that speech activity had begun. If we look at spectrograms of the digits in noise, we do see some masking of initial speech activity, but this tends to be of shorter duration than the RT increase. Also any masking is clearly reduced by the noise reduction processing although the RT is not. For example Figure 5b shows a spectrogram of the start of "eight" in babble, where the mean increase in reaction time is 44ms. Compared to the quiet condition shown in Fig 5a, it is possible to see that about 20ms of a rather quiet onset is somewhat masked. However compared to the noise reduced version in 5c, the onset is within 20ms of the quiet condition, but the mean RT increase is still 43ms. Thus while masking may be playing some role in the RT increase, there is also evidence of some general effect of the noise on speed of cognitive processing, an effect which is not reduced by the noise reduction process.

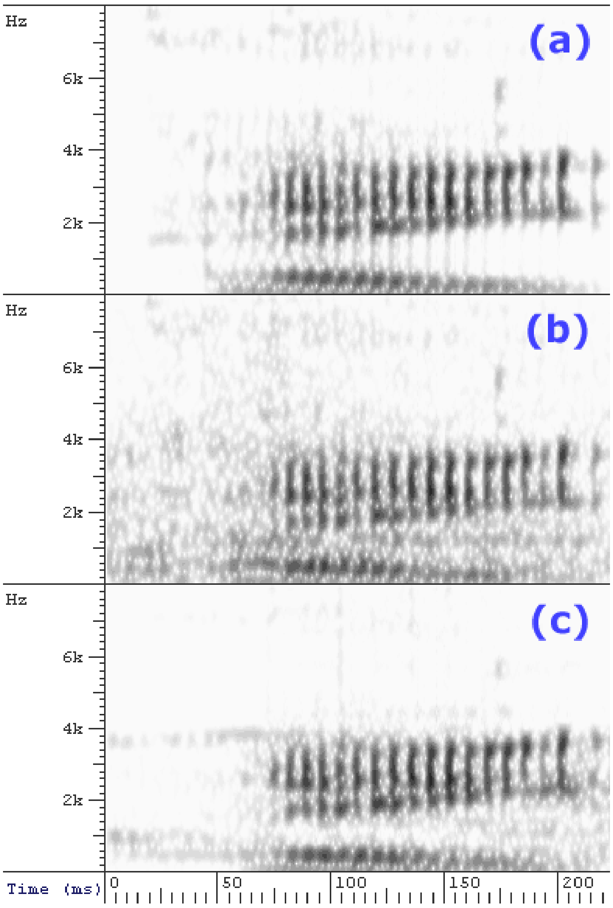

Figure 5. *Effect of Babble and NR on start of "eight". (a) quiet, (b) babble, (c) babble+NR*

5. Discussion

In this experiment we investigated whether time taken to recognise digits in noise was improved by noise reduction processing. If we had seen an improvement this could have been taken as evidence that NR had reduced the cognitive effort required to process speech in noise. However we did not see any improvement in reaction time caused by NR.

The experiment did show that the presence of noise increased the time to recognise a spoken word. We proposed two reasons for this: either because the start of "speech activity" is harder to detect due to noise masking the onsets of the words, or because the processing of speech in noise is more effortful and requires more processing resources.

Taking the first explanation, if noise makes speech activity take longer to detect, why does noise reduction have no effect? Is this because noise reduction does not restore quiet sounds at start of words, or because noise reduction processing itself reacts slowly to speech onsets in the noisy signal?

Taking the second explanation, if the presence of noise affects speed of recognition, why does noise reduction have no effect? Is it because noise reduction does not restore redundancy of phonetic cues, or because noise reduction adds processing artefacts which are themselves distracting and responsible for increased cognitive load?

Further investigations of the effect of NR on cognitive load are required to make these issues clear. It would be interesting to compare the results here with a task in which reaction time to "speech onset" alone is measured. This would help us isolate the auditory from the cognitive processing effects.

We have shown however that just because NR processing improves the subjective quality of speech signals it is not necessarily the case that NR leads to a reduction in cognitive effort.

6. Acknowledgements

Thanks to Gaston Hilkhuysen for preparing the audio materials and for statistical advice.

7. References

[1] Hu, Y. and Loizou, P.C. "A comparative intelligibility study of single-microphone noise reduction algorithms", J Acoust Soc Am 122 (2007) 1777.

[2] Hu, Y., Loizou, P., "Subjective comparison of speech enhancement algorithms", Proc. ICASSP 2006, 153-156.

[3] Ephraim, Y., Malah, D. "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator", IEEE Trans. Acoustics Speech and Signal Processing, 32 (1984) 1109-1121.

[4] ITU-T. Subjective test methodology for evaluating speech communication systems that include noise suppression algorithms. ITU-T Recommendation P.835, November 2003.

[5] Broadbent, D., "Human performance and noise", in *Handbook of Noise Control*, ed. C. Harris, McGraw-Hill, New York, 1979.

[6] Smith, A., "A review of the effects of noise on human performance", Scandinavian Journal of Psychology, 30 (1989) 185-206.

[7] Miles, C., Jones, D.M., Madden, C., "The locus of the irrelevant speech effect in short-term memory", J. Experimental Psychology: Learning, Memory and Cognition, 17 (1991) 578-584.

[8] Martin, R., Wogalter, N., Forlano, J., "Reading comprehension in the presence of unattended speech and music", Journal of Memory and Language, 27 (1988) 382-398.

[9] Pichora-Fuller, K., "Processing speed and timing in aging adults: psychoacoustics, speech perception and comprehension", Intl. J. Audiology, 42 (2003) 59-67.

[10] Kahneman, D, *Attention and Effort*, Prentice-Hall, 1973

[11] NATO noises (audio recording). NATO: AC243/(Panel 3)/RSG-10. Soesterberg, NL: Institute for perception-TNO, the Netherlands.

[12] ANSI S3.5-1997 American National Standard Methods for the calculation of the speech intelligibility index, ANSI 1997.

[13] Bell, T., Dirks, D., Trine, T., "Frequency-importance functions for words in high- and low-context sentences", J. Speech Hear Research. 35 (1992) 950-9.

[14] Brooks, M., "VOICEBOX: Speech Processing Toolbox for MATLAB", Department of Electrical & Electronic Engineering, Imperial College, London UK, 2008. Available from: http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html