ARCHIVES OF ACOUSTICS
33, 3, 323–330 (2008)

LISTENERS' REACTION TIME RESPONSE TO SPEECH-IN-NOISE MATERIAL

Magdalena A. BŁASZAK

Adam Mickiewicz University
Institute of Acoustics
Umultowska 85, 61-614 Poznań, Poland
e-mail: blasaku@gmail.com

(*received March 30, 2007; accepted March 18, 2008*)

The paper addresses the problem of choosing speech material for the experiments concerning measurements of the composed reaction time (CRT). A comparison was done of mean reaction times measured for a group of subjects exposed to Polish vowels /a, e, i, o, u, y/[(1)] and to non-words recorded on a dummy head against traffic noise (European Standard EN 1793-3) generated from an open window. The results of this experiment, analyzed for various signal-to-noise ratios and different reverberation conditions, indicate that the mean reaction time was greater for non-words (when the subjects were exposed to more complex signals) rather then for vowels. However, differences in the relative growth of the reaction time values with decrease of signal to noise-source output difference level (SNS) were relatively small.

Keywords: reaction time, cognitive factors, speech, room acoustics.

1. Introduction

The reaction time, understood as the delay of the beginning of the listener's response in relation to the end of the generated signal, is one of the subjective measures of "listening difficulty" [1, 3, 5]. This is an important parameter, for example among other things, in synthetic speech perception and in the acoustical design of special enclosures, such as classrooms for children or for foreign language learning, because it takes into account the fact that a *man*, not a microphone, is the receiver.

In many experiments, based on the measurements of the reaction time [2], words or sentences are used as the speech material. This choice seems justified taking into regard the cognitive processes taking place in the human mind. On higher levels of processing logical and semantic relations are of great importance. Temporal information decoding is directly related to the semantic information and hence durations of mental processes

[(1)] IPA symbols: /a, ɛ, i, ɔ, u, ɨ/.

show high variation. The result depends not only on the acoustic parameters such as signal-to-noise ratio, the number of modulated bands (in the synthetic speech) and reverberation, but also on the listener's intelligence, rate of associations and familiarity with a given word (Fig. 1). To eliminate the effect of these factors on CRT, it seems more reliable to apply in the experiment the speech sounds without any semantic meaning. However, the question appears whether the use of isolated vowels or consonants would bring reliable results. Should not the acoustic signals used be more complex and more resembling sounds of speech like non-words (logatomes)?

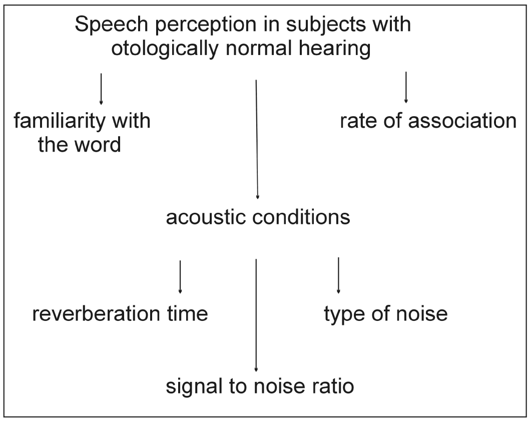

Fig. 1. Factors influencing the reaction times scores.

2. Experiments

The aim of the experiment performed in this study was to check the effect of the choice of the speech material on the delays in the subjects' responses, expressed in terms of identification scores in noise and as a function of response times in different reverberation conditions. Two types of speech materials were tested: the vowels and non-words (male speaker). The speech material was recorded in two rectangular rooms of the same geometry and different reverberation characteristics (see Fig. 2a, b), and in the anechoic chamber. The noise-generating source was mounted at the height of a window to imitate natural conditions, while the source of the sounds tested (vowels and non-words) was mounted at the place where usually a lecturer or speaker would stand. Both types of signals were recorded binaurally using a dummy head Neumann KU100 placed in the middle of the room at the same distance (3 m) to both sound sources. The speech and noise-source output difference levels (SNS) were: -3.0 and 3 dB. Also clean speech signals without the masker were recorded. It is important to emphasize here that the signal-to-noise level difference is defined in terms of values inherent to the speech- and noise-source outputs, independent of their acoustical environments, and not, as it is common, in terms of the values at a receiver position which are, of course, dependent on the acoustical environment.

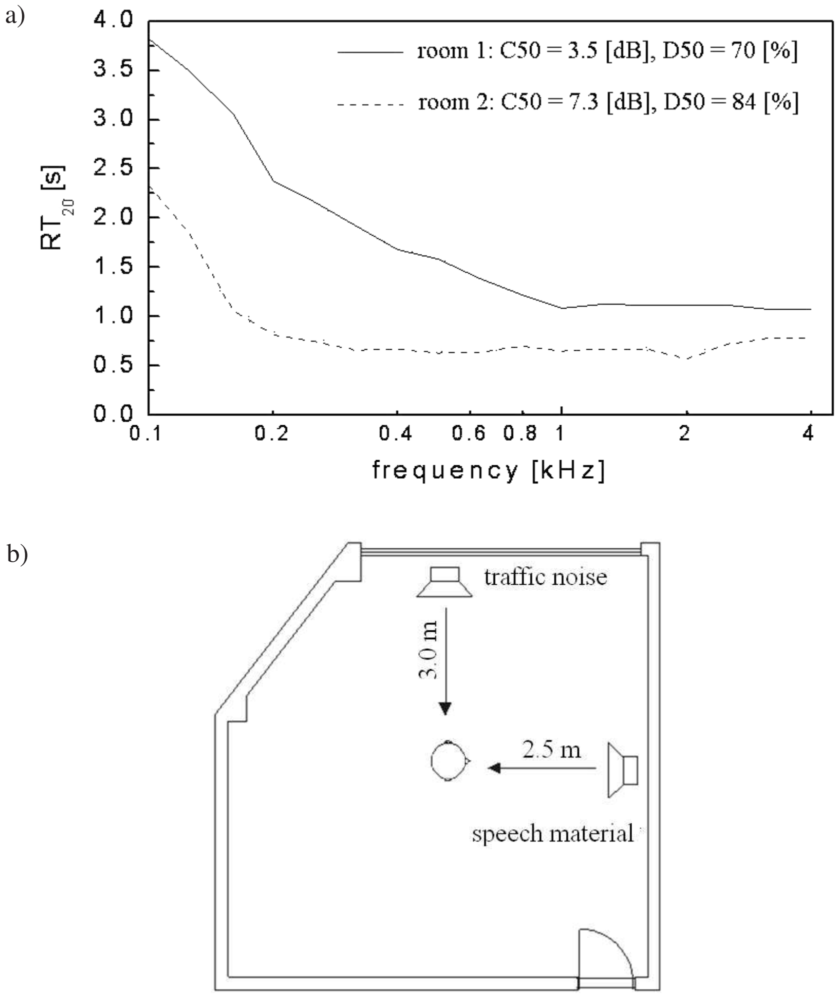

Fig. 2. Experimental scenery: a) room characteristics (RT_{20}, C50, D50), and b) signal sources and receiver (dummy head) positions.

Twelve phonetically balanced 100-elements lists and six vowels were presented via headphones. Each subject listened to a total of 1200 non-words and 864 vowels (6 vowels × 4 SNS × 3 enclosures × 12 repetitions). All speech signals were presented in random order. The experimental conditions were preceded by instructions and examples of the required tasks. Listeners were tested individually in three sessions for about 1.5 hour. During the session the subject could take a brake whenever he felt tired or could not concentrate what is very important in reaction time experiments. The average output level of masking noise was equal to 65 dBA and did not change during the experiment.

During the experimental exposure to the sounds, the listeners were seated in a sound-attenuated booth and all the equipment was located outside. Each subject had to respond to the stimuli by identifying the perceived vowel or non-word (pronouncing it to a microphone) and writing it down on the keypad connected with a PC. Before the experiment, all the subjects have been trained with stimuli similar to but not the same as those used in the experiment.

The reaction time was measured between the offset of the generated signal and the onset of the repeated answer.

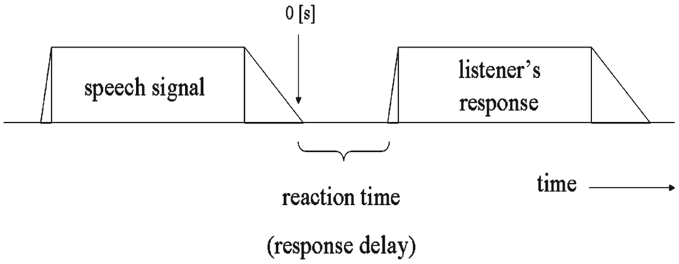

Fig. 3. Response delay measurements.

The CRT was retained for analysis only in the case when the answer was correct. Response times above 4 s or below 200 ms were excluded from analysis, which led to rejection of less than 1% of the data.

Seven young normal–hearing listeners (YNH) with the threshold <20 dB HL at octave intervals from 125–8000 Hz participated in the experiment. YNH were native speakers of Polish and had normal time of simple reaction. All the listeners were very carefully selected. The reflex and attention ability was evaluated using the Bourdon's test based on the psychometric method and standardized for Polish population [4]. The test permitted the elimination of persons with concentration disturbances. The task the candidates were asked was to cross out the letters W, R and N hidden in 28 verses of a text in the time up to 10 minutes. To pass, the test should be performed at the same pace that is measured by the percent of the test performed per minute.

A PC-compatible computer with a signal processor (TDT System 3) generated stimuli through a 16-bit D/A converter (TDT HB7) at the 24.414 kHz rate and recorded the listener's responses. Senheisser THD 47 earphones were used.

3. Results

As the reaction time was measured only for correct responses, differences in their values were assumed to be a result of cognitive processes. For this reason, it was expected to find a delay in getting the responses with deteriorating acoustic conditions. Table 1 presents the percentage of incorrect answers for vowels and non-words in all the considered conditions.

However, the main question was what differences in reaction times, if any, would be found using two different types of speech materials and which of these materials would be more suitable for such experiments. The results were subject to the ANOVA statistical analysis whose results are given in Table 2.

Table 1. Mean incorrect responses for vowels and non-words (in percent).

	SNS [dB]	Incorrect responses [%] (room 1)	Incorrect responses [%] (room 2)	Incorrect responses [%] (anechoic chamber)
vowels	−3	13	13	8
	0	11	10	7
	3	6	8	7
	without noise	6	7	6
non-words	−3	62	76	44
	0	40	72	39
	3	42	64	37
	without noise	27	16	15

Table 2. Results of ANOVA statistical analysis.

	F	p-value
speech material	$F(1, 7633) = 603.25$	$p < 0.05$
SNS	$F(3, 7633) = 31.29$	$p < 0.05$
enclosure	$F(2, 7633) = 66.77$	$p < 0.05$
enclosure * SNS * speech material	$F(6, 7633) = 0.32$	$p > 0.05$

* (interaction)

The ANOVA analysis has shown that significant differences exist due to factors such as: speech material, SNS and enclosure effects what confirms the expectations. As can be seen in Fig. 4a, the response time values in room 1 increased by about 80 ms for vowels and 160 ms for non-words relative to those obtained in the anechoic chamber. This means that the dynamics of changes in this parameter was about twice greater for non-words, although qualitatively the results were almost the same. Moreover, there was no interaction effect: enclosure* SNS* speech material which means that the mean reaction time differences between vowels and non-words were similar in all the considered acoustic conditions.

As can be seen in Fig. 4, there were significant differences between the mean reaction times determined for the speech materials of the two types and between the curves presenting the reaction time versus speech and the noise-source output difference level separately for vowels and non-words. The longest mean reaction time for both speech materials was found in room 1 ($\mathrm{RT}_{500} = 1.6$ s, $\mathrm{RT}_{2000} = 1.11$ s) : 0.53 s for vowels and 0.8 s for non-words. This result was consistent with the expectation that the longest duration of mental process of signal recognition is needed in enclosures with highest reverberation. It should be noticed that the mean difference between the reaction times (measured without noise for both speech materials) in room 2 and in the anechoic chamber was about 70 ms (Fig. 4b), however, there was only a 1% difference in speech intelligibility (Table 1). It suggests that response delay may be a more sensitive measure than simple intelligibility scoring and it may be important in specifying the real "listening difficulty" in reverberation conditions.

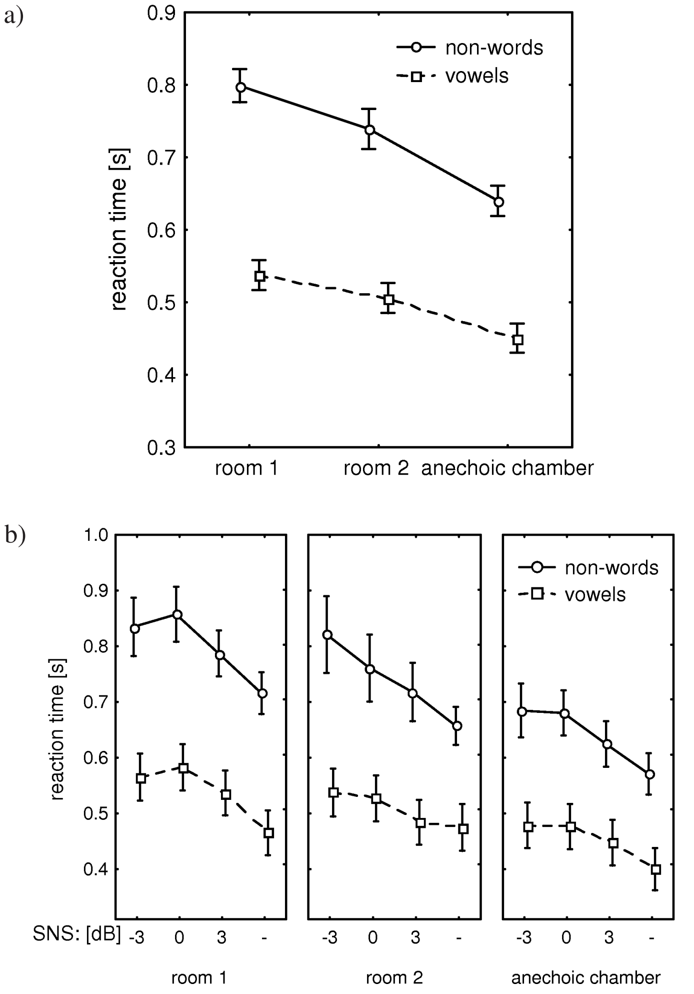

Fig. 4. Mean reaction time values: a) for two types of speech material and b) presented as a function of the speech and noise-source output difference level in three different enclosures.

It is obvious that a minimum reaction time must have a certain duration greater than zero. The stimulation of a receptor, transmission of nerve pulses from the receptor to the nervous centre and from the latter to the effector and the muscle contraction need some time. As the time needed for the response does differ between subjects, it seems reasonable to analyse relative increments in the response times, which should permit elimination of differences in the simple reaction times of the listeners. Regarding the above, the delays in the reaction times were presented as relative increments with respect to the simplest situation from the point of view of perception, i.e. with respect to the response time to signals without noise recorded in the anechoic chamber, see Fig. 5. The results obtained show that the growths of the reaction time for both types of speech material are similar. The greatest relative increase in the reaction time was noted for the signals from room 1 (of the highest reverberation), whereas the smallest one for the signals from the anechoic chamber.

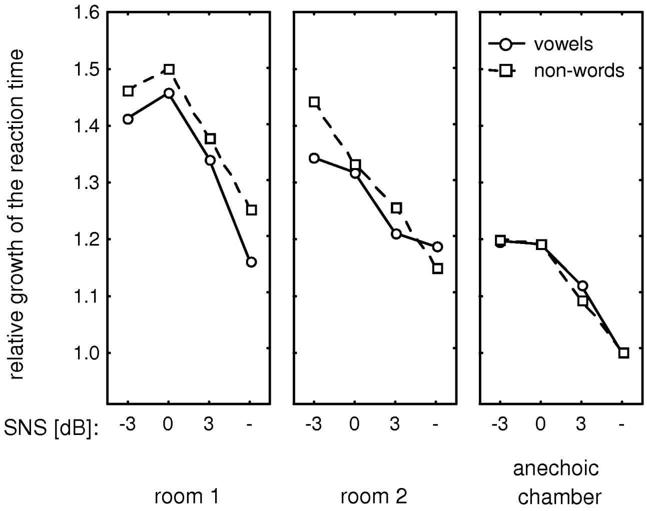

Fig. 5. Relative growth of the reaction time as a function of the speech and noise-source output difference level presented for three different enclosures.

It should be noticed that for both the signals in room 1, the reaction time was somewhat shorter for SNS = -3 than for SNS = 0. It seems to be related to the fact that with decreasing speech and noise-source output difference levels the subjects gave fewer correct answers. The number of reaction time values taken into the calculations decreased, so the more difficult are the acoustic conditions the greater is the error of the results. Moreover, for lower values of SNS, the information taken from the reaction time measurements can be misleading because the listeners ability to understand decreases and they begin to "guess". For this reason a few percent of correct responses could be made just as fast "good guesses" and the reaction time obtained can be shorter. The problem is that it is impossible to discern between guessed and actually heard responses, so the problem indicates the SNS limit of getting reliable reaction time values.

4. Conclusions

The brain processes taking place between the stimulus and the reaction need some time. This time can be divided into a constant minimum (the simple reaction time) and the remaining time dependent on the mental processes, known as the composed reaction time. This composed reaction time is needed for the brain to process the information received by the listeners in various acoustic conditions and its length is a measure of the "listening difficulty".

The results of the experiment performed have shown that in response to simple stimuli (vowels) the reaction time in different acoustic conditions is small, while for more complex stimuli (non-words) it is greater. The non-word is more complex than a vowel, so its mental perception needs more time. Moreover, the longer the reverberation time and the lower the speech and noise-source output difference level (SNS) the greater the reaction time. There were also differences in the relative growth of the reaction time values but they were relatively small. It should be noted that there is no reaction time value that can be said to be good enough or that corresponds to no listening difficulty, and such a value cannot be found because of inter-subject differences in the simple reaction time [6].

It seems possible and important to determine the curves of the reaction time increase in a wider range of acoustic conditions (with different maskers, reverberation, spatial separation of noise and signal) and to compare the results obtained for isolated vowels and non-words.

References

[1] BŁASZAK M., RUTKOWSKI L., *Logatom's intelligibility in a room with respect to traffic noise transferred through an open window*, 53 Otwarte Seminarium z Akustyki, Zakopane 2006.

[2] DELOGU C., CONTE S., SEMENTINA C., *Cognitive factors in the evaluation of synthetic speech*, Speech Communication, **24**, 153–168 (1998).

[3] MORIMOTO M., SATO H., KOBAYASHI M., *Listening difficulty as a subjective measure for evaluation of speech transmission performance in public spaces*, J. Acoust. Soc. Am., **116**, 3, 1607–1613 (2004).

[4] RUBINSZTEJN S.J., *Methods of experimental pathopsychology* [in Polish]: *Metody patopsychologii eksperymentalnej*, PZWL, Warszawa 1967.

[5] SATO H., BRADLEY J.S., MORIMOTO M., *Using listening difficulty ratings of conditions for speech communication in rooms*, J. Acoust. Soc. Am., **117**, 3, 1157–1167 (2004).

[6] WAGNER E., FLORENTINE M., BUUS S., MCCORMACK J., *Spectral loudness summation and simple reaction time*, J. Acoust. Soc. Am., **116**, 3, 1681–1686 (2004).