APPLICATIONS FOR THE HEARING-IMPAIRED: COMPREHENSION OF FINNISH TEXT WITH PHONEME ERRORS

Nina Alarotu[1], Mietta Lennes[1], Toomas Altosaar[2], Anja Malm[3], and Matti Karjalainen[2]

[1]Department of Phonetics
University of Helsinki
PO Box 3, FIN-00014 University of Helsinki, Finland
E-mail: nina.alarotu@helsinki.fi

[2]Acoustics Laboratory
Helsinki University of Technology
PO Box 2200, FIN-02015 HUT
Finland
E-mail: toomas.altosaar@hut.fi

[3]Finnish Association of the Deaf
PO Box 57, FIN-00401 Helsinki
Finland
Tel. +358-9-580 3460

ABSTRACT

This study simulates the phoneme errors made by speech recognizers and determines the phoneme error level at which a reasonable comprehension of text can still be achieved. Finnish is written almost phonemically and Finnish-speakers have no trouble in comprehending phonemic text. Phonemically corrupted text was presented to normal-hearing, hearing-impaired as well as deaf subjects and their comprehension levels were measured. According to this study, current speech recognition methods allow for limited applications in this field.

1. INTRODUCTION

One of the current challenges of speech technology is to improve speaker-independent automatic speech recognition to such a level so that applications based on this technology can be developed for the hearing-impaired and deaf. There are two phases in the speech recognition process:

1. Recognition of speech sounds (phones/phonemes).
2. Conversion of these sounds into a readable form.

Since Finnish is written almost phonemically, the second phase of the recognition process is practically non-existent. Therefore, a Finnish-speaker can easily read phonemic text, i.e., the output of a phoneme-based speechrecognizer.

The purpose of the present study was to determine at which phoneme error level written Finnish in varying contextual environments can no longer be comprehended by different target groups. The relationship between phoneme error level and comprehension rate along with response time was also studied. This study provides valuable information on the performance levels required for phoneme-based speech recognizers for different hearing-impaired and deaf target groups.

2. METHODS

2.1. Subjects

Four different groups of subjects participated in the experiment:

- normal-hearing subjects, *n=45*, 13 males, 32 females, age 21-43 years
- "experienced", normal-hearing subjects that were accustomed to reading text with spelling errors written by hearing-impaired individuals, *n=10*, 2 males, 8 females, age 29-53
- deaf subjects, *n=19*, 6 males, 13 females, age 25-71 years
- hearing-impaired subjects, *n=4*, all males, age 22-60 years

As for the group of hearing-impaired subjects, it should be noted that the findings based on the behavior of this group are only indicative because of the small size of this sample.

2.2. The phoneme error model

Nine different phonetically-based rules were used to generate phoneme errors and were induced at certain levels into test words and sentences by either insertions, deletions, or replacements of phonemes. Two different replacement rules existed: in-class (e.g., alveolar stop to bilabial stop /t/ => /p/), and out-of-class (e.g., fricative to tremulant /s/ => /r/). The model was tuned to generate similar types of errors as would a phoneme-based speech recognizer[1].

2.3. Experiments

All subjects were presented with three different types of test material:

a) Isolated words were used to measure comprehension of context-free text with phoneme errors. The first 100 words of the Finnish frequency dictionary were chosen. The phoneme error level of these isolated words was varied between 5 and 30 per cent, their order randomized once, and then presented to each subject in the same order.

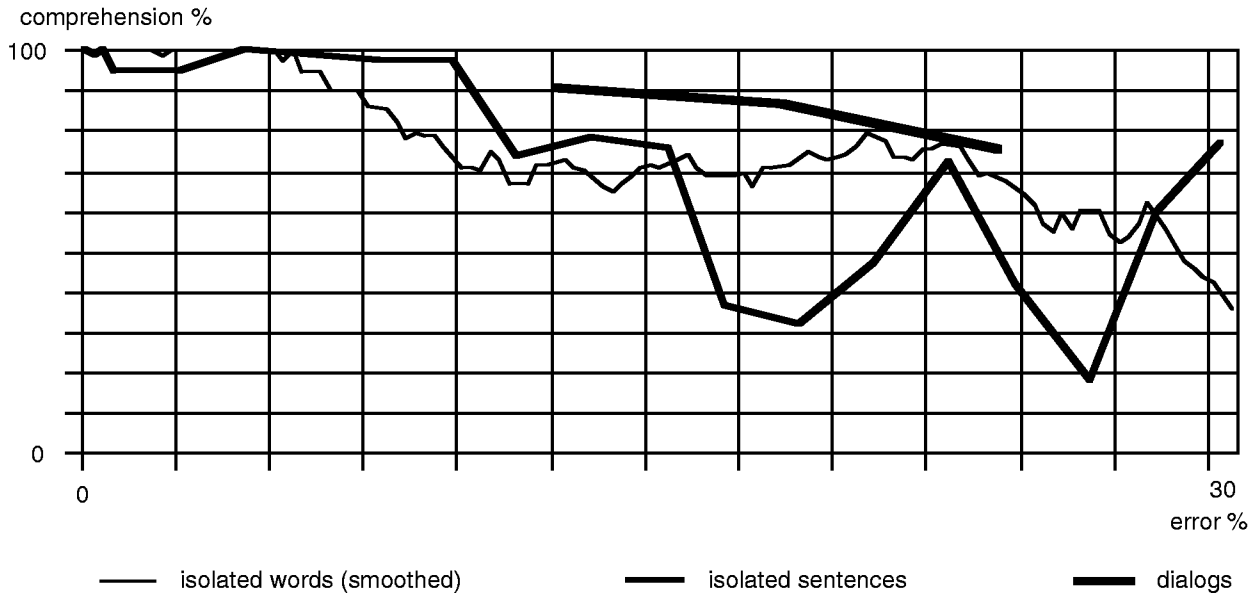

Fig. 1. Comprehension level vs. phoneme error level of isolated words, isolated sentences, and the three dialogs combined (all subjects, n=78). When a subject is given more contextual information, comprehension becomes noticeably easier.

b) Twenty isolated sentences were used as test material to measure the effect of increasing context. The phoneme error level was varied from 0 to 30 per cent, and the order of the sentences randomized three times. The subjects within a target group were split up into three sub-groups with each sub-group tested on one of the randomized orderings. This was done to minimize the effect of presentationorder.

c) Three fictitious telephone conversations of different complexity were presented to reveal the probable effects of a larger contextual framework. The phoneme error level of each conversation text was either 15 %, 20 %, or 25 %. Three different versions were generated as with the isolated sentences not only to remove the effect of presentation order but also to remove the effect of dialog complexity (i.e., easy, moderate, and difficult).

The experiments were performed in a quiet room using a computer-based prompt and measurement system. The instructions were first given orally or in Finnish sign language, and the subject was also able to read short instructions on the computer screen before and during the test. A practice sequence of five isolated words and two sentences were presented at the beginning of each experiment. The subject started the timer (by pressing the space bar) which caused a corrupted test word or sentence to appear in a window. When the subject had made a decision regarding comprehension and pressed the space bar to stop the timer, the response time was recorded and the correct text shown in another window. He or she then indicated whether the text had been comprehended correctly. In the case of the dialogs, the subject was then shown his or her own imaginary response (without errors) at this point to give continuity to the dialog.

3. RESULTS

3.1. General

The comprehension level probably depends on the familiarity of the stimulus but also on the types of the phoneme errors induced into it.

The variance of the comprehension level was very high for isolated words (see *Fig.2*). This is due to the fact that isolated words have no specific context. Even an error in a single phoneme of an isolated word can be misleading to the subject since it may change the whole meaning of the word. This will cause very different responses even for words with the same error level.

Isolatedwords were comprehended almost perfectly below a phoneme error level of 10 %. Generally, for *isolated sentences* (see *Fig.1*), the comprehension level was satisfactory (over 70 %) even if the phoneme error level was as high as 17 %. For *dialogs*, a 20 % error level still allowed for 85 % comprehension, and a reasonable comprehension level of 75 % was achieved even at a 25 % error level.

In general, comprehension declines with increasing phoneme error levels.

At low phoneme error levels it is obvious that increasing context makes comprehension easier.

3.2. Differences between subject groups

There was no significant difference in comprehension levels between the groups regardless of the stimulus type. Nevertheless, the response time was slightly longer

for deaf and hearing-impaired than normal-hearing subjects (see *Fig.4*). It is possible that their reading speed and accuracy are negatively affected by their inability to hear speech, and/or, that they are using a foreign language, i.e., their mother tongue is Finnish sign language.

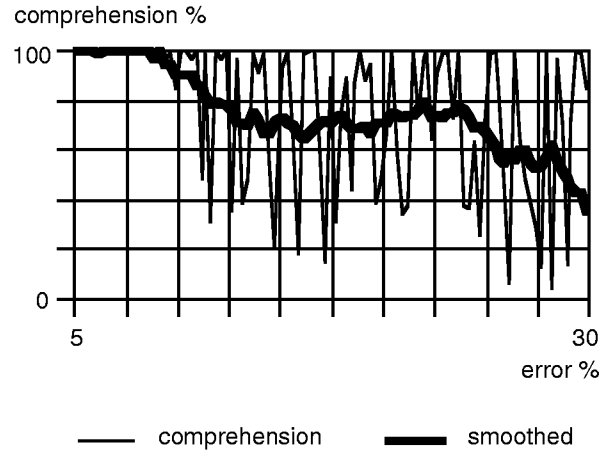

Fig. 2. Comprehension level vs. phoneme error level of 100 isolated words and a smoothed curve showing the general trend. Note the large variance in the comprehension level of individual stimuli. (all subjects, n=78)

The order of average response time was among the different groups as follows: 1. experienced, 2. normal, 3. hearing-impaired, and 4. deaf subjects.

However, the difference between the average response times of the two normal-hearing groups and, on the other hand, between the two other groups was negligible. As shown in *Fig.4*, the deaf and the hearing-impaired subjects seem to encounter more problems than the normal-hearing subjects with high phoneme error levels.

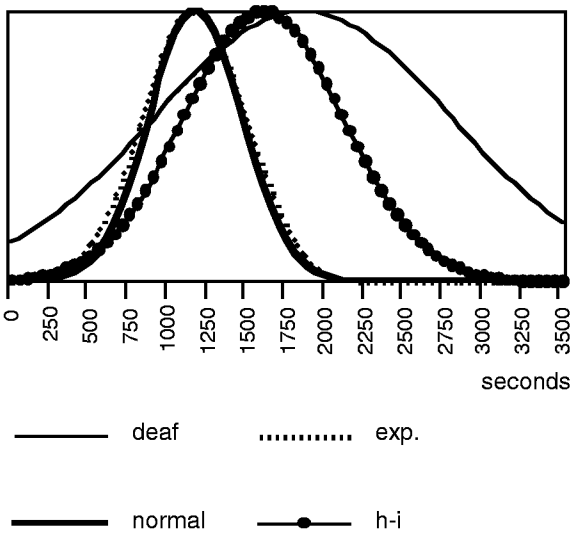

Fig. 3. The curves represent the total duration of the experiment for each of the four subject groups. The Gaussian curves were calculated from the mean and standard deviation of each group.

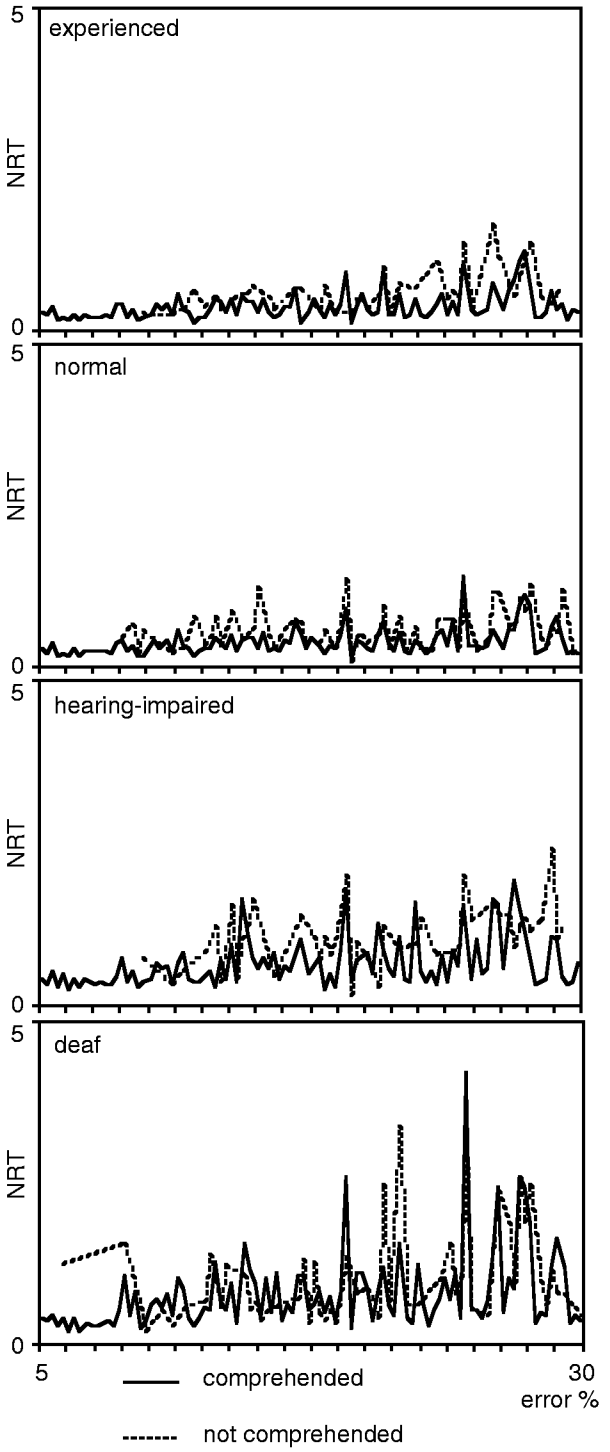

Fig. 4. Comparison of the four subject groups regarding response time of comprehended and not-comprehended stimuli in isolated words. (NRT = normalized response time = response time (sec) divided by the number of phonemes in the incorrect text)

When the stimulus was not comprehended, the average response times were systematically longer in all groups.

Differences between groups could also be observed in the total duration of the test (see *Fig.3*). The two normal-hearing groups were the quickest to complete the test. Hearing-impaired subjects took slightly more time and

deaf subjects were rather slow in comparison to others. However, one should be aware of the fact that there were only four hearing-impaired subjects.

4. DISCUSSION

Possible sources of error could be caused by the choice and structure of sentences and dialogs. The dialogs were of different levels of difficulty: dialog 1 (mother calling daughter) and dialog 2 (two friends talking) represented a very familiar subject for students while dialog 3 was quite difficult with many topic changes, very long sentences, and novel words. The sentences and dialogs represented standard spoken Finnish, which is not typically used in everyday conversation.

The subjects may also have been confused over how to respond to the question of comprehension, although they were instructed to answer "yes" whenever they had understood the general meaning of the word or sentence correctly. Phonetic knowledge may affect the subject's intuition about possible phoneme error types. The results might also have been more reliable if the errors had been generated "on-line", i.e., separately for each subject.

Speaker-dependent speech recognizers for Finnish perform at a phoneme error rate of about 5 to 15 per cent [1]. Results from these experiments indicate that this is low enough to allow for comprehension of corrupted phonemic text. The phoneme error level of isolated words could be as high as 11 % without practically any degradation in comprehension. Isolated sentences provided additional contextual information and it was possible to comprehend text even at an 18 % phoneme error level. Finally, conversational dialogues provided the highest contextual environment and a 25 % phoneme error level still did not deter comprehension substantially.

5. REFERENCES

[1] Karjalainen, M. et al. (1997) Applications for the Hearing-Impaired: Evaluation of Finnish Phoneme Recognition Methods. In This Proceedings.