Speech Understanding while Driving: Sentence and word perception using a hand-held mobile phone *versus* a hands-free head set

Jukka Selenius†, Reijo Aulanko†, Markus Vaalgamaa‡, Jukka Harjula* and Heikki Summala*

† Department of Phonetics, University of Helsinki, Finland

‡ Acoustic Technology Team, Nokia Mobile Phones, Helsinki. Finland

*Traffic Research Unit, Department of Psychology, University of Helsinki, Finland

E-mail: jukkaselenius@hotmail.com, reijo.aulanko@helsinki.fi, Markus.Vaalgamaa@nokia.com, heikki.summala@helsinki.fi

ABSTRACT

Sixteen licensed drivers repeated short sentences and nonsense words which they heard either through a hand-held mobile phone or a hands-free head set while driving in normal city traffic in central Helsinki. The error rates showed that perceptual accuracy decreased by the use of telephone connection, compared with a 'simulated passenger speech' where the stimuli were presented from a loudspeaker in the car. Somewhat surprisingly, using the hands-free head set produced more errors than the hand-held mobile phone. This may in part be due to the dark sound colour of the head set used. The results also suggest that the cognitive load due to the traffic situation (e.g. difficult crossings) had a negative effect on the performance in the verbal task.

1. INTRODUCTION

Using a mobile phone while driving can have a disturbing effect on the driver's performance and produce dangerous situations in traffic. This is partly due to the difficulties caused by the increasing cognitive workload when the driver is trying to understand the telephone conversation [1] and dividing her/his attention between the traffic situation and the verbal message [2]. In Finland, the proportion of fatal road accidents which were affected by the use of mobile phone increased from 0.1% in 1991–92 to about 1.5% in 1995–98 [3] and further to about 3.3% in 2001 [4]. The use of a mobile phone (whether hand-held or hands-free) while driving may be safe in quiet traffic conditions, but it is, nevertheless, an attention-demanding factor which, in a critical situation, may have fatal consequences [5], [6], [7].

In this study [8] we investigated the intelligibility of meaningful Finnish sentences and nonsense words delivered through a mobile phone (hand-held or hands-free reception) while the listener is driving in normal city traffic among other cars, trams and pedestrians in an environment with numerous crossings, hills, narrowing lanes and different kinds of pavement. This kind of natural situation, while admittedly impossible to control to the last detail, provides more relevant data on the drivers' behaviour in real traffic situations than do laboratory-based driving-simulator studies [9].

The aim of the study was twofold: we wanted to find out how much the use of a mobile phone and simultaneous driving affect the ability to perceive and understand spoken messages, and how this verbal task affects driving performance. Thus, the study combines the interests of phonetics, traffic psychology and environmental acoustics. In the following, we report preliminary results of the phonetic aspects of the study.

2. METHODS

Participants. In a field test 16 Finnish male participants (aged 24–54, mean 36.2, SD 7.8) drove an instrumented car in normal city traffic in central Helsinki listening to and repeating the speech material they heard. The participants (all unpaid volunteers) were experienced drivers, eight of them also licenced race drivers. They were all used to driving in city traffic and also in the specific area of Helsinki where the field test was performed. All participants were tested to have normal hearing and normal vision with regard to acuity and visual field.

Driving Task. The car used in the field test was a Mitsubishi passenger car equipped with special measurement instrumentation and double controls. During the test each participant drove along a predetermined 3.3-km route, each round taking about 8–9 minutes. Before the test session proper, each driver drove three practice rounds getting acquainted with the route. On average, the total duration of the driving session for each participant was about one hour. The whole test route in central Helsinki was supposed to be well covered by the mobile phone field, but during the first day there were still some interruptions in the transmission. Luckily the phone problems disappeared after the first day. The test route was also driven through with a test dummy placed on the driver's seat recording all the acoustic environmental information in all three listening conditions (see below).

ISBN 1-876346-48-5

(The results of these acoustic control measurements are not dealt with in this paper.)

Stimulus Material. The spoken material consisted of 160 normal, meaningful Finnish sentences (of 5–8 words) and 42 two-syllable VCV words (mostly nonsense, but not violating the phonotactics of Finnish). The original material was spoken in a quiet room both by a young male and by a young female speaker with clear laboratory-speech style and an average articulation rate of 5.9 and 5.8 syllables/second, respectively. In one half of the VCV words the initial vowel was /e/ and in the other half /i/. The consonants used in the middle position were /d h j k l m n ŋ p r s t v/. The vowel in the final position was one of the eight vowels of Finnish, /a e i o u y æ ø/.

Listening Conditions. There were three listening conditions in the car: the driver received the prerecorded stimuli either (1) through a hands-free head set connected to a Nokia 6150 mobile phone, (2) through the same hand-held phone without the head set, or (3) from a CD player through a loudspeaker situated near the front-seat passenger's head ('simulated passenger speech'). In a control condition (4), the material was presented in a quiet room from a CD player through a loudspeaker at a distance of about two meters. In each of the four conditions each participant heard a different list of 40 sentences followed by a list of 26 VCV words. In each stimulus list there was 1-second 500-Hz beep as a warning signal about 6 seconds before the first stimulus. Between the sentences there was always a pause of about 6 seconds in the recording giving the driver time to repeat the stimulus. Between the VCV words this pause was about 4 seconds. The duration of each prerecorded stimulus list was a little under 8 minutes.

Verbal Task. In conditions 1 and 2 the driver received a phone call from the sending wire telephone about half a minute after the start. When the driver answered he was instructed to say aloud his name for the recroding before the stimulus list was played to him. In condition 3 the stimulus tape was started by an experimenter in the back seat of the car.

Each participant was presented with the total of 160 sentences (each sentence in the original material occurred once) and 104 VCV words (items in the original list of 42 syllables in different combinations). The material was balanced so that no-one heard the same list more than once. For each participant, the session consisted of the three listening conditions in the car in a pseudo-random order. The control condition in the quiet room was always arranged later.

Responses. The subjects' vocal reponses were recorded on minidisk in the car and also at the sending telephone. All situations in the car were also videotaped. The video recording (see Figure 1) was composed of four simultaneous screens (the driver's mouth and face and two different views of the traffic). A computer in the car continuously recorded the noise levels, car movements and the use of the steering wheel and other controls.

Figure 1: Example screen of the video recording made in the instrumented car, consisting of two views of the driver, a view ahead and to the left. In this test condition the stimuli were delivered through the hand-held mobile phone. The numerical data include the date, car speed, steering wheel angle, use of acceleration pedal, brake pedal etc.

Altogether 2560 sentences and 1664 VCV words were obtained for analysis.

Scoring. In this preliminary analysis of the data, error rates were calculated in all conditions. An 'error' in the repeating of a sentence was defined as any kind of deviation from the original stimulus. First one of the authors (JS) scored the drivers' responses on the basis of the minidisk recordings made at the sending (without the video recordings). After that, the responses were scored independently by two other persons using the video recordings made in the car. These two scorers were phonetically trained but were not aware of the purpose of the study. The results presented below are based on the scoring made by the two independent scorers.

In addition to just counting the error rates, the errors were also classified according to their degree of 'severity', ranging from simple sound substitutions to more complex deviations, on a scale from 1 to 3.

3. RESULTS

Sentences. On average, there were errors in 36.6% of the sentences in the hands-free condition, 28.5% in the hand-held phone condition, 7.5% in the simulated passenger condition (CD in the car), and 4.2% in the control condition in a quiet room (see Figure 2).

When the errors for the sentence material were scored according to their degree of severity, the results showed the same overall pattern as the simple error rates.

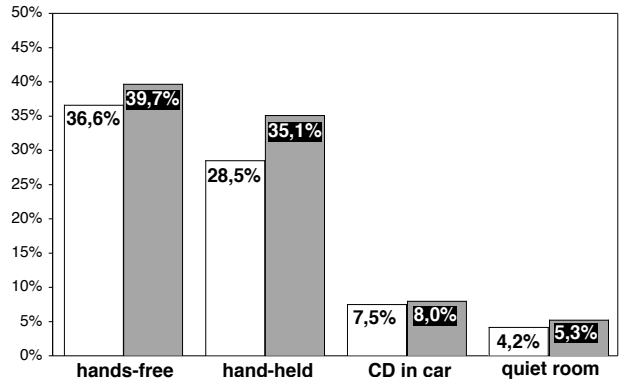

Figure 2: Mean error rates in the sentence and word repetition tasks (white and gray columns, repectively) in the three experimental conditions (hands-free head set, hand-held mobile phone, CD in the car) and the control condition in the quiet room.

Eight of the participants were licenced race drivers and eight were normal experienced drivers. Of all the errors made in the sentence test, 53.2% were made by the normal drivers, and the remaining 46.8% by the race drivers.

VCV words. For the VCV word stimuli, the corresponding error rates were slightly higher than for the sentence stimuli: 39.7% for the hands-free condition, 35.1% for the hand-held phone, 8.0% for the simulated passenger condition, and 5.3% for the control condition in a quiet room (gray columns in Figure 2).

There was no strong tendency in the errors to change from a nonsense word to a meaningful word: 76.3% of the errors in repeating a nonsense VCV word resulted in another nonsense word, the remaining 23.7% resulting in a meaningful Finnish word.

There was a difference in performance related to whether the stimuli were spoken by a man or a woman. For the VCV word material the average error rate was 54.9% for the male speaker and 45.1% for the female speaker.

4. DISCUSSION

Both the sentence material and the VCV word material indicated that the use of a phone resulted in a decrease of repetition performance compared with the simulated passenger condition where the stimuli were heard from a loudspeaker in the car. As might be expected, nonsense word material proved slightly more difficult than meaningful sentence material in all listening conditions.

Surprisingly, speech (both sentence material and – less clearly – VCV word material) was perceived better when heard through a hand-held mobile phone than through a hands-free head set. This may be partly due to the dark sound colour of the headset used. It is also possible that in general the participants found it easier to control e.g. the sound volume with the hand-held phone than with the head-set plug inserted in the ear canal. While the use of a hands-free device (instead of a hand-held mobile phone) probably increases safety by allowing the driver to use both hands to control the car, the present results suggest that there may also be more cognitive workload involved if the driver finds it more difficult to understand the message delivered via a hands-free device.

On the basis of the VCV word results, it seems that the female voice is slightly easier to understand in a mobile phone. This may be partly due to the (small) difference in articulation rate between the two speakers (see Methods).

Both listening and talking to a mobile phone and driving a car demand a lot of cognitive capacity. The more often one drives a car, the more automatic the driving becomes. Highway driving in general is not nearly so demanding as driving in heavy city traffic. The difference in performance between race drivers and normal experienced drivers observed in this study may reflect the fact that race drivers are more used to handling the car in difficult situations, and are able to allocate more cognitive capacity to the verbal task.

According to the subjective reports of the drivers the cognitive and physical workload of the traffic had a strong impact on their ability to repeat what they heard from the mobile phone. This demonstrates that in heavy city traffic it is difficult to perceive non-redundant speech from a mobile phone.

The methods used in this preliminary analysis proved problematic in certain respects. Since it is obviously impossible for the driver to give his/her verbal reponses in writing one has to rely on his/her spoken responses. Then there is always the danger of perception error on the part of the scorer. Indeed, the scorers found their task quite demanding; they had difficulties with the quality of the soundtrack of the video recordings made in the car. In order to avoid further perceptual errors (made by the scorers) the best possible sound (and video) recording should of course be used for the scoring. For the final scoring and analysis of the responses in the present material we shall use a combination of the video recording made in the car and the minidisk recording made at the sending telephone.

Despite the difficulties in the methodology of this kind of study, the advantages of a field study are obvious. A field test involved with driving and using a mobile phone gives more realistic results than do studies using driving simulators. The test drivers driving in normal city traffic perform under constant threat of real hazard.

In further studies we will also make a detailed analysis of the phonetic errors (phoneme confusions and effects of phonetic context). The errors made by the drivers in the verbal task will also be correlated with their performance in the driving task to see whether it is primarily certain critical, demanding situations in the traffic (crossings etc.) that cause the impaired performance in the verbal task.

From the viewpoint of traffic psychology, the effects of

ISBN 1-876346-48-5 © 2003 UAB

simultaneous phone conversation on driving performance will be dealt with in future reports.

REFERENCES

[1] O. Tsimhoni, P. Green and J. Lai, "Listening to natural and synthesized speech while driving: Effects on user performance." *International Journal of Speech Technology*, vol. 4, pp. 155–169, 2001.

[2] P. J. Cooper, Y. Zheng, C. Richard, J. Vavrik, B. Heinrichs and G. Siegmund, "The impact of hands-free message reception/response on driving task performance." *Accident Analysis and Prevention*, vol. 35, pp. 23–35, 2003.

[3] A. Holopainen, "Matkapuhelin kuolemaan johtaneissa liikenneonnettomuuksissa. [Mobile phone in fatal traffic accidents.]" Helsinki: Liikennevakuutuskeskus, Vakuutusyhtiöiden liikenneturvallisuustoimikunta, 2000.

[4] A. Holopainen, personal communication to J. Selenius, February 2003.

[5] D. Lamble, T. Kauranen, M. Laakso and H. Summala, "Cognitive load and detection thresholds in car following situations: safety implications for using mobile (cellular) telephones while driving." *Accident Analysis and Prevention*, vol. 31, pp. 617–623, 1999.

[6] D. E. Haigney, R. G. Taylor and S. J. Westerman, "Concurrent mobile (cellular) phone use and driving performance: task demand characteristics and compensatory processes." *Transportation Research Part F*, vol. 3, pp. 113–121, 2000.

[7] L. Nunes and M. A. Recarte, "Cognitive demands of hands-free-phone conversation while driving." *Transportation Research Part F*, vol. 5, pp. 133–144, 2002.

[8] J. Selenius, R. Aulanko, M. Vaalgamaa, J. Harjula & H. Summala, "Understanding of Mobile-phone Speech While Driving." Manuscript.

[9] D. Haigney and S. J. Westerman, "Mobile (cellular) phone use and driving: a critical review of research methodology." *Ergonomics*, vol. 44, pp. 132–143, 2002.