Car navigation system via the telephone. Part 1: Entering alphanumeric data

Institute for Perception Research
PO Box 513, 5600 MB Eindhoven

18.10.1994

Rapport no. 1011

Car navigation system
via the telephone

Part 1

A. Garlich and A. Tiritoglu

Dit rapport kwam tot stand in het kader van het afstudeerproject van A. Garlich. Het project gold als afsluiting van de opleiding Techniek en Maatschappij aan de Technische Universiteit Eindhoven, in de specialisatie Informatietechnologie en Gedrag. Drs. R. Gobits van de faculteit Wijsbegeerte en Maatschappijwetenschappen trad op als eerste begeleider. Tweede begeleider was Prof.dr.ir. F.L. van Nes, groepsleider Informatie-ergonomie op het IPO.
De tweede auteur, A. Tiritoglu, was gedurende het project op het IPO werkzaam als buitengewoon medewerker. Afkomstig van de Universiteit van Kansas City volbracht hij zijn fellowship alhier.

Car Navigation System via the Telephone

Part 1: Entering Alphanumeric Data

A. Garlich and A. Tiritoglu

Abstract

In this report, the user interface study for a Car Navigation System via the telephone is laid down. The experiment's set-up is explained and the results of the first experiment are presented. In this experiment, subjects had to enter city and street names to the car navigation system using a standard desk telephone. Three different input techniques were tested: keypad entry, speech recognition entry and entry using a human operator. In the first two systems, the names had to be typed and spelled letter-by-letter. With the human operator system full names could be entered. Subjects, half of them older than 65 and half of them younger than 65, were interviewed before and after the experiment and gave their opinion about the three systems on several attributes. The human operator system was liked the most, followed by the keypad system. The speech recognition system was liked the least. No difference was found between the older and younger subjects. Suggestions for improvement of the user interface are given.

Introduction

In today's complex city environment, city navigation is becoming more important. There are several solutions that would help drivers in finding the way around the cities. With the current technology it is now possible to navigate in metropolitan areas and prevent the loss of time and energy due to complex road systems. Another advantage of car navigation systems is the possibility of avoiding traffic jams and road blocks. Navigation aids can also have a positive influence on safety.

Most of the car navigation systems currently available make use of complex hardware and software which are installed into the vehicle. The Philips CARIN Navigation System using the CD-Rom is an example. A completely different approach would be a telephone service that would guide the driver using a standard or a mobile telephone. This technology is currently available and it is in running condition at Philips Research Labs in Briarcliff Manor, New York. However, the system has not been tested properly or used as an application for various reasons. In addition the system is set up for the U.S. road and traffic system.

It is clear that the system is not comparable to in-car navigation systems such as CARIN in many respects. However, such a telephone service has many advantages. First, the system is simple and it is based on a service rather than being an actual product that the users would buy and install into their vehicles. Second, it is available to everyone who has access to a regular or mobile telephone. Third, the system can be used interactively as well as in batch-instructions mode. Interactive use is intended for drivers with a car telephone who can use the system online. Batch mode use is for people who call the service via house telephones. Those people can tape the guidance messages (on a tape recorder or dictaphone), get a transcription on fax, or write the messages down. The last option is also possible when using a public telephone. The system is meant for in-city navigation but it could also be implemented for highway systems.

A telephone service for navigation may have several pitfalls. First, the system operates without a vehicle location device. Therefore the driver has to pay more attention to the given instructions and the environment to perform and confirm his/her actions. Second, errors made during navigating are not taken into account and no alternative routes are being calculated. Third, there are no visual cues to help the driver during the navigation.

Such a system requires a user interface study to find out if it would work for people. In addition, navigational issues of such a service have not been fully understood e.g., what instructions to give, preventing ambiguous instructions, intelligibility of auditory instructions. Any extra attentional demands on the driver must be considered with great care. This holds especially for elderly drivers, whose attentional resources are known to be limited. Therefore this study covers elderly people as well as younger people. The gained knowledge from this study may be incorporated with gerontechnology studies that have been started at the Eindhoven University of Technology.

Methods

The user interface study is divided into two separate experiments. The tested entities in both experiments are part of the proposed Car Navigation System, but the results of the experiments can also be used outside this context. In the next figure the experiments set-up can be seen.

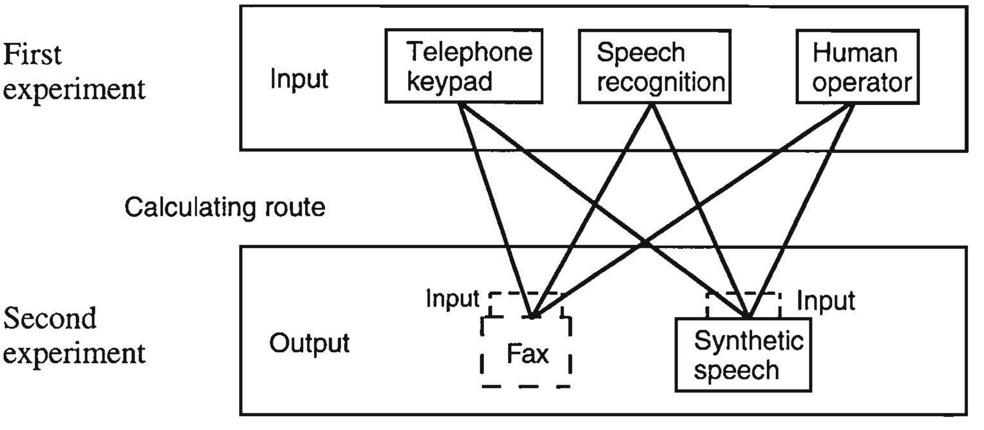


Figure 1: Experiments set-up.

The first experiment, that is covered in this report, only dealt with the input part of the system, i.e. giving the names of the current location city and street, and the destination city and street. To do this, three input techniques were tested. With the human operator input technique, that was added for comparison, subjects had to say full street and city names. With the other two (automatic) systems, subjects had to enter the names letter-by-letter; i.e. for telephone keypad entry typing the names and for speech recognition entry spelling them. This letter-by-letter entry was considered usable because the necessary number of letters to enter before the street name is recognized, is ≤6 for 91% of all the street names in the Eindhoven region. Recognizing means that the number of possible names with the letters entered up to then, has dropped to 5 or less. These names are then read to the user who can pick the right one from the list. When the names are entered with the most distinguishable part first, the percentage of street names that can be recognized within the first 6 letters, increases to 97% (See Appendix A).

The speech recognition system was added to this experiment as an extra comparison to keypad entry. However, the speech recognition system was not designed the optimum way: it did not have the dialogue structure of a system that was designed particularly for speech recognition. Instead, the dialogue structure was more or less the same as that of the keypad entry system. Therefore, the results of this experiment, specifically the differences between the two automatic systems, have to be considered with care.

Implementation

For each input technique used in the first experiment, a software implemented interface was made. Programming was done in Hypertalk on an Apple MacIntosh, using HyperCard 2.1. The systems were implemented according to the general design guidelines for telephone systems as proposed by Gould and Boies (1987), Halstead-Nussloch (1989), and Schumacher (1992).

The layout of the letters that was used for the telephone keypad, was the ITU-T (CCITT) Recommendation E.161 option A, layed down in CCITT (1993). This layout was chosen as the result of an inquiry to what telephone keypad letter-layout is the most suitable in the Netherlands. Appendix B reports on this inquiry. The chosen layout can be seen in figure 2.

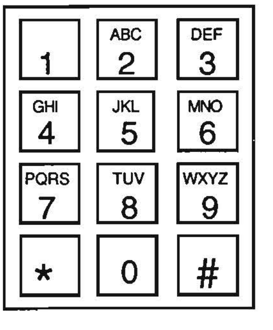


Figure 2: Telephone keypad letter layout.

Letters were typed on the keypad using the single keying method because it was found in the literature (Garlich, 1994) that this technique is superior to other keying techniques and can be used in this limited-database situation.

The Speech recognition input technique was implemented by means of a Wizard-of-Oz simulation. In such an experiment, subjects interact with a machine without knowing that (part of) the machine is simulated by a so-called wizard. Everything that was said by the subjects and that was valid at a certain point in the dialogue was entered to the system by the wizard. The wizard operated as an ideal speech recognizer: all letters were recognized the right way. This was done because no literature could be found that dealt with confusion matrices for letters with current speech recognition technology. Nusbaum and Pisoni (1987) only measured the total error rate of six commercially available speech recognizers. Also, those speech recognizers were all speaker-dependent; i.e. only recognized speech from the person with which they were trained. The only confusion matrices for individual letters that are known, are those from human-to-human communication investigated by Pols (1977). However, only consonant-consonant and vowel-vowel confusion was measured. Also, the letters were part of three letter words.

Subjects interacted with the systems through a normal desk telephone (type T65) of which the keypad was connected to the MacIntosh by flat cable. The letter layout of figure 2 was brought on the keypad by using stickers. For the Human Operator system, a microphone was connected via an amplifier to the desk telephone's loudspeaker. Also for this system and the Speech Recognition system, the desk telephone's microphone was connected to an amplifier and made audible to the experimenter who operated as the Wizard-of-Oz speech recognizer.

A sampled male voice was used for the big part of all the messages that were used in the systems. Only for reading names a synthetic voice was used. This was done because having to sample all the street names and city names seems impractible. The synthetic voice was generated by the IPO speech board which was connected to the serial port of the MacIntosh. The input of the speech board was the ASCII representation of the names, including the accent indications that were brought in by hand. A typical example of a complete message is: "*Als u op* **Twentelaan** *bent, druk dan een 1*", in which the italic part was read by the sampled human voice and the bold part by the synthetic voice.

Subjects

The subjects used in this experiment, were 24 people from outside IPO. Half of them were younger than 65 (average age 37.6), half of them were older (average age 69.3). In both age groups half of the subjects were male and half female. Subjects were selected from the IPO subjects list and were paid Dfl 8,- an hour. The whole experiment lasted for about 1.5 hours.

Procedure

All subjects had to do three tasks with all three input techniques, keypad entry (K), speech recognition (S), and human operator (H). However, the order in which they did was changed. Each order was followed by four subjects: 2 older than 65 and 2 younger than 65. Also, 2 of them were male and 2 of them were female. Table 1 gives an overview of this classification.

Table 1: Classification of subjects.

	Order of input technique					
Task order	KSH	KHS	SKH	SHK	HKS	HSK
123	F >65	F >65	F <65	F >65	M >65	F >65
	M <65	M <65	M >65	M <65	F <65	M <65
321	M >65	M >65	M >65	M >65	F >65	M <65
	F <65	F <65	F <65	F <65	M <65	F >65

Total: 24 subjects; 12 older than 65 (6F, 6M) and 12 younger than 65 (6F, 6M)

As can be seen in the table, within each group there was another grouping of subjects. Some performed the tasks in the order 123, others in 321. This was done to see the effect of an 'easy' or 'difficult' task as being the first one. An easy task is a task in which subjects who enter (part of) a name to the system do not have to press the same key twice (for two successive letters) and only get back one suggestion for a real street or city name. They can then simply confirm or deny the proposed name. A difficult task is a task in which subjects sometimes have to press the same key twice and get back a list of possible street or city names after they have entered part of the name. They must then pick the right name from the list.

Subjects had to do the same three tasks with all three input techniques. The tasks were handed to them on paper (see appendix C).

Measurements

Both objective and subjective measurements were performed. For the objective measurements, logging facilities were included in the software of the telephone systems used in the experiment.

Subjective measurements were collected by interviewing the subjects before the experiment and after the experiment. The first interview was done to get people's general opinion about automated telephone services, getting to know people's attitude in finding their way around cities, and getting to know people's reaction to the description of the proposed car navigation system via the telephone. Appendix D contains a copy of the questions used in this interview. The second interview was meant for evaluation of the three different input techniques and getting to know if people would actually use the system if it was available in real life. Appendix E has a copy of this interview.

Besides the interviews, subjective measurements were done by semantic differential: subjects were given a form after doing all three tasks with a system and were asked to give a rating for each system on several attributes; e.g. 'bad - good', 'slow - fast'. These attributes were partly taken from Dintruff et al. (1985) and from Roelofs (1987). Subjects could mark their opinion regarding these attributes on a 7-point scale; See appendix F for a copy of the semantic differential form.

Results

From the interview before the experiment it turns out that only 2 older subjects had used automatic telephone services before (Telebingo and ordering of clothes). Four of the older subjects also had never heard of these services. Of the younger subjects only one had never heard of such services. Seven of these younger subjects had used automatic telephone services before.

Before the experiment it was asked what people liked the most and the least about automatic telephone services. There was no agreement on this. Some subjects liked the speed the most (fast), others liked the speed the least (slow). Things liked the most by more than two subjects were the absence of waiting queues, and not being dependent of the time the service is called. Things liked the least by more than two subjects were the impersonality of the systems, the absence of the possibility to ask questions, and the negative consequences for the employment rate.

Most subjects (19) were positive in their reaction to the description of the new telephone service. The other five subjects said the service wasn't necessary because maps and other means would do. Two subjects who gave a positive reaction said it shouldn't be allowed to call the service while driving a car.

Subjects' general impression after working with all three systems ranged from good, nice, handy and simple to not necessary and very slow. In total there were 21 positive reactions and 3 negative reactions. Subjects complained about the messages being too long (especially the introduction message), the systems being too slow, and the need for paying close attention when interacting with the systems.

Of all subjects, 15 said they would use the car navigation system via the telephone. The other 9 said they wouldn't and argued they could do with maps and prefered to find the information themselves. The fee that the subjects were willing to pay, varied extremely. Some were not willing to pay at all ("It should be a service"), some would pay up to Dfl 5,- per call. Subjects who liked to pay per minute were willing to pay 40 cpm in average. Subjects who liked to pay per call, i.e. getting the information they want, were willing to pay Dfl 2,70 in average. Subjects who would use the system, would do this 9 times a year on average, provided the charge they had suggested was implemented.

Preferred input technique

When asked what input technique subjects liked the most, all but one subject said they liked the human operator the most. Reasons for this were that that system was more personal, easier, and faster. The keypad system was ranked in second place, leaving the speech recognition system as being liked the least. Reasons for the dislike of the last system were that it was chaotic, slow, less understandable, and the most difficult system.

The same ranking can be seen when looking at the outcome of the semantic differential. In the figures 3 to 5 the results of the semantic differential for all three systems are layed down. In these graphs the two age groups are compared for their scores on the scales.

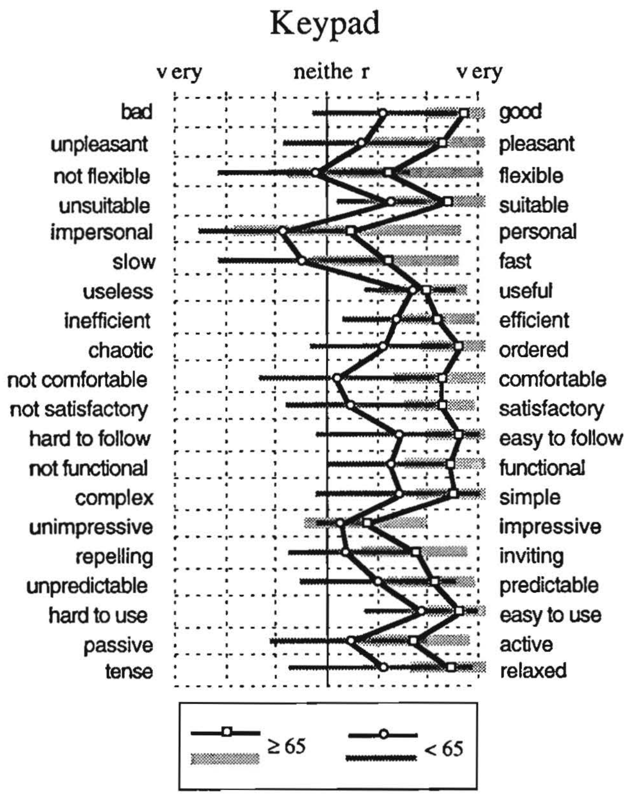


Figure 3: Results for the keypad entry system with comparison of the two age groups.

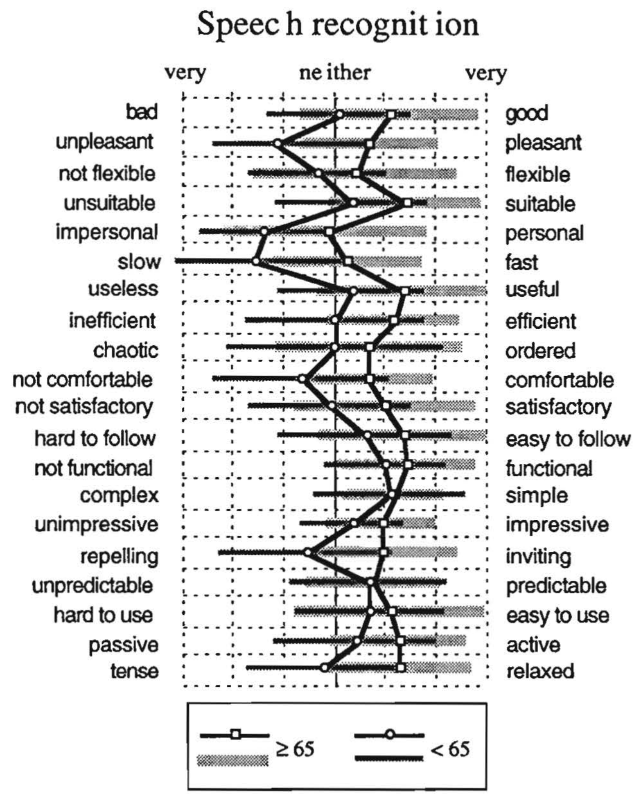


Figure 4: Results for the speech recognition system with comparison of the two age groups.

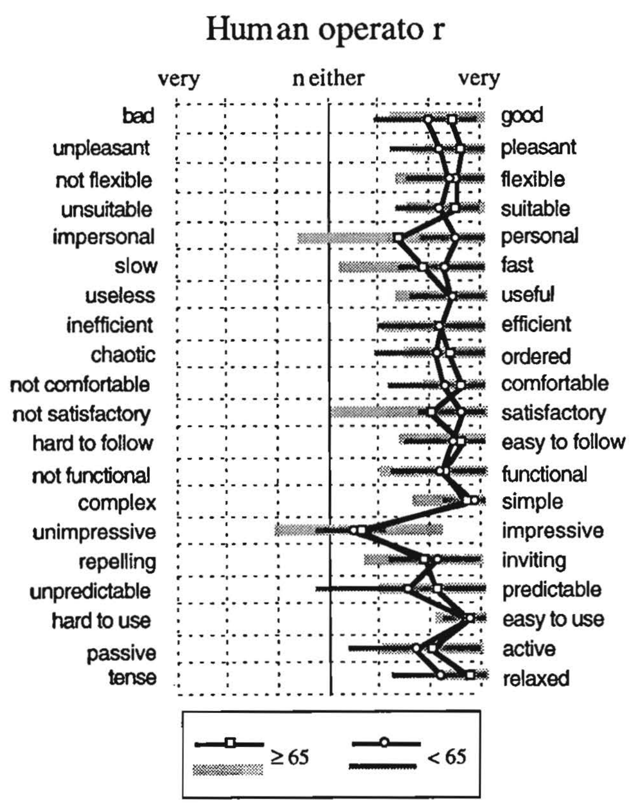


Figure 5: Results for the human operator system with comparison of the two age groups.

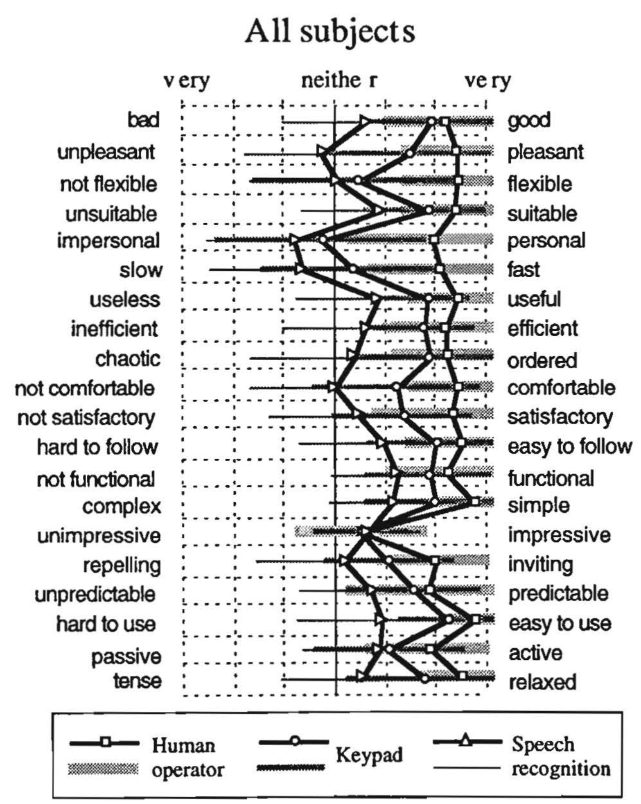


Figure 6: Rating scores for the three tested systems by all 24 subjects.

In the charts, the circles and the squares represent the mean scores for the attributes. The shaded bars represent one standard deviation on either side of the mean.

For the two automatic systems (figures 3 and 4) older subjects tend to give more positive ratings than younger subjects. However, the difference is not significant; the Mann-Whitney U test only shows 12 significantly differing objective scores ($p<0.05$) out of 20 between the two age groups for the keypad entry system, and only 2 significantly differing objective scores for the speech recognition system. In rating the human operator system, there was no difference between younger and older subjects (figure 5).

The low score by the older subjects on impersonal/personal for the human operator system can be explained from the fact that 2 of the older subjects ranked this scale 'very impersonal'. These subjects (a married couple) said the reason for this was, that they did not know the human operator *personally*. Without these two subjects, the mean score on this objective for the older subjects is 2.2.

Figure 6 shows the rating scores by all subjects for the three tested systems. According to these results, the human operator system is liked the most, followed by the keypad entry system and the speech recognition system. The differences between the three systems are significant (Kruskal-Wallis test, $p<0.05$), except for the objective 'unimpressive/impressive' ($p=0.88$).

The results of this semantic differential should be considered with some care. Only the relative scores between the three systems are reliable. The absolute scores for each system are probably too high. This assumption is based on research done in the past. Nielsen and Levy (1994) for instance, found that the median system satisfaction was much higher than the numeric middle of the rating scale. They argue that test users tend to be polite and that subjects give fairly high ratings unless they are really upset with an interface. Nielsen and Levy also found that test users tend to blame themselves rather than the system when having trouble using the interface. From their meta-analyses, Nielsen and Levy conclude that, when using 7 point scales, it is reasonable to take 5.6 as the planned value for subjective satisfaction in the goal-setting phase of a usability-engineering process. So it is not enough to just aim at getting above the numeric midpoint of the rating scale.

Intelligibility

In the interview after the experiment, subjects were asked several questions about the intelligibility of the voices used in the three systems. All 24 subjects said the male voice reading the general messages was quite intelligible. When names were read as part of a message by the synthetic voice, all younger subjects heard a different voice. Four of these subjects mentioned that voice being an artificial one. Of the older subjects 4 did not hear a different voice when the names were read to them. None of the older subjects mentioned that voice being artificial.

Most subjects said these alternating voices, i.e. sampled and synthetic voice, were not disturbing. Two subjects said it was irritating and confusing. On the other hand, two subjects said it was nice and pleasant because it made the systems more attractive. 18 subjects said the names were intelligible. The other 6 said they were not all intelligible, but the names they had to pick from the lists were.

Lists of names

When entering the letters with the two automatic systems (i.e. keypad and speech recognition), the entered letters were compared to the database of all the possible names. A user would continue to type letters of the name until the number of valid matches had dropped to 5 or less. If only one name was valid, this name was read to the user who could then confirm or deny the name. If more than one name was valid, a list containing these names was being read to the user which could then pick the right one from the list. For keypad entry the user had to press the corresponding number of the right name. For speech recognition entry the user had to say that number.

Half of the subjects said they did not like these name lists. These subjects prefered to type or spell the whole name, or at least until only one name was left.

Picking a name from a list yielded several confusions. 7 subjects, 5 older than 65 and 2 younger than 65, said they had not heard how they were supposed to pick the right name from the list when interacting with the speech recognition system. These subjects said they were so focused on hearing the right name, they did not hear anything else and therefore missed the instruction on how to choose. All of these 7 subjects thought they had to say 'yes' or 'correct' when the right name was read to them. Some of them also said 'no' when the read name was not the correct one.

This behaviour can be understood when one looks at the intended dialogue designed for the speech recognition system. Throughout the system all a user has to say is yes or no. Then, when the dialogue comes to the point where something has to be picked from the list, the modality changes from confirming/denying to choosing. Apparently, some subjects did not make the necessary modality switch.

The results of the subjective measurements do not show gender effects, task order effects, or effects of order of input techniques.

Performance measures

Besides the subjective measurements, objective data were also collected for evaluation. Before looking at the outcome of these objective measurements, one must realize that the three tested systems are not comparable in many respects. Because the individual systems have a completely different structure, the performance of the subjects with these systems can not simply be compared one-to-one!

For both the keypad entry system and the speech recognition system, there is a learning effect for the time it takes the subjects to finish a task. In figures 7 and 8 these significant learning effects (Kruskal-Wallis test, $p<0.05$) can be seen for the two age groups. The differences between the older and younger subjects are not significant (t-test, $p>0.05$). Each column of the charts in figure 7 and 8 is made up of two parts: one representing the time that is needed for correct entry of the four street and city names, and on top of that a part that represents the time a subject listens to messages.

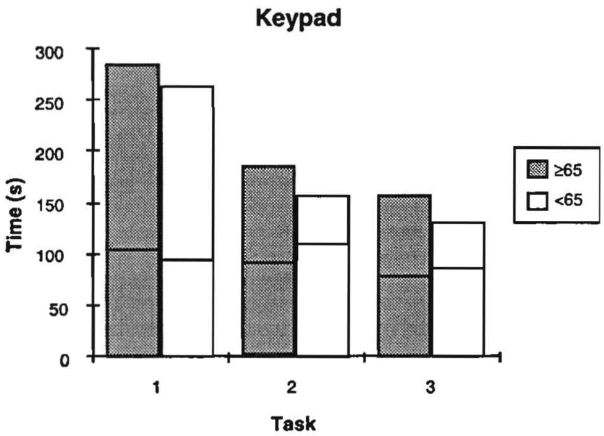


Figure 7: Time per task for keypad entry system; comparison between age groups.

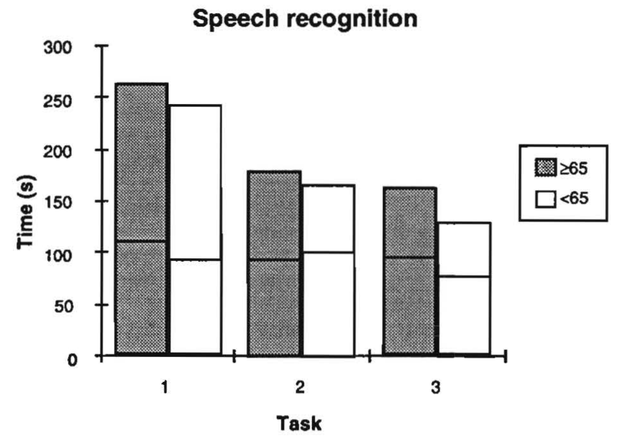


Figure 8: Time per task for speech recognition system; comparison between age groups.

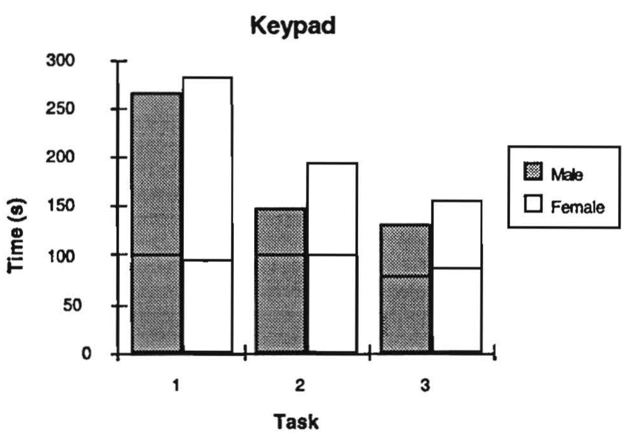


Figure 9: Time per task for keypad entry system; comparison between gender.

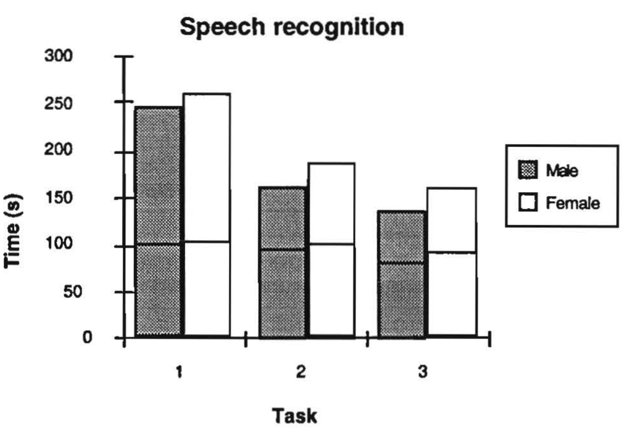


Figure 10: Time per task for speech recognition system; comparison between gender.

The time for performing the first task with the two systems, contains the time for listening to the instructions on how to interact with the system. For both systems, the instructions take about 2 minutes. Most subjects did not ask for instructions in the second and third task. Some subjects by mistake asked for the instructions a second time. They did not really listen to the instructions when they had made such a choice. After the experiment two subjects told they wanted to have a special key in the keypad system, e.g. the '0', that would let them escape from the instructions. However, subjects who asked for instructions by mistake in the speech recognition system did not use the 'skip' command that was built in for that purpose. None of the special commands in the speech recognition system ('hold', 'go on', 'repeat', 'skip') was ever used, although they were being explained in the instructions.

Something like this was seen with the keypad system. In that system it was possible to interrupt the messages if one already knew what key to press. Although a lot of subjects had their finger already at the right key and were anxious to press that key, only two subjects actually did it before a message was ended. Because this interrupting of messages can save a lot of time, especially when listening to a name list, the existence of this feature should be incorporated in the instructions.

Figures 9 and 10 show the task times for the keypad entry system and the speech recognition system respectively, with a comparison between male and female subjects. Again, the lower part of the columns represents the entry times for the tasks. It seems like male subjects need less time for doing the tasks than female subjects. However, the difference is not significant (t-test, $p>0.05$).

The average time for doing a task with the keypad system is 196.3 seconds and with the speech recognition system 190.7 seconds. As was mentioned before, these times can not be compared to each other one-to-one. However, these times give an indication how fast a caller of this telephone service can be handled, especially when the average time for the human operator system is considered. It was found that the average time for completing a task with the human operator system was 60.6 seconds. So, in the same time one of the two tested automatic systems can handle 1 caller, the human operator can handle approximately 3.2 callers.

For time per task with the keypad and speech recognition system, a task order effect was found. Subjects who started with the difficult task (i.e. task order 321) took more time for their first task but less time for the second and third task than subjects who started with the easy task (i.e. task order 123). There was no effect of order of input techniques.

For the keypad system, 8 subjects older than 65 and 5 subjects younger than 65 did not make any errors during the three tasks in entering the four names. The other 11 subjects made 24 errors in total in entering their 132 names. One a third of these errors was due to typing too many letters of the names. In such a situation the subject keeps typing letters although the system comes with a list of possible names. If a certain key that the subject uses for typing a next *letter* is a valid *number* for choosing from the list, there is a big chance the unwillingly chosen name is not the correct one. Seven of the errors were made because subjects started with typing a street name when actually a city name was necessary. Doing the wrong task number yielded 3 errors, picking a wrong name from the list 2, and using a wrong key for a letter only 1. Finally, one younger subject made 3 errors on purpose; she wanted to test the system for robustness.

One younger subject made the same error 3 times and was transfered to an operator who helped the subject for completing the task. This feature was built into the system.

For the speech recognition system the error rate was far higher. 7 subjects (4 older and 3 younger than 65) even had to be helped by the wizard, otherwise they would be completely stuck in the name lists. This helping was done by generating a wrong letter when the subject already had typed part of a name and waited for things to come. This wrong letter then resulted in a name that was not in the database. As a consequence the subject was asked to pick the right name from the list again. In extreme situations subjects were helped by picking the right name from the list.

Of the total of 31 errors, 27 had to do with subjects not knowing how to pick a name from the name list. 1 error was made because a subject repeated the last entered letter, 7 because subjects restarted the spelling of a name from the beginning, and 4 because subjects were just waiting what would happen and therefore did nothing. 15 errors were made because subjects were using another way to pick the right name from the list. They did not use the appropriate number for the name, but tried to confirm the name after they had heard it by saying the full name aloud (5 times) or using 'yes'/'correct' and 'no' (10 times).

Six older subjects and 7 younger subjects did not make any errors during their interaction with the speech recognition system. For error rate in using the tested systems, there were no gender effects, task order effects, or effects of order of input technique.

From the observation of the subjects it can be concluded that at least some subjects perceived the speech recognition system as being human. They did not start spelling right away, but first said e.g. "Ok. The Achterstraat. A... C... etc.". Also, one subject consistently said "The A... The C... etc." when spelling. After the experiment one subject mentioned she rather would spell the letters by using the letter alphabet, .i.e. "Arnold... Chantal... etc.". The problem with this is that there are so many letter alphabets and most people come up with their own name for a particular letter.

Discussion and conclusions

What user interface to use for the proposed car navigation system via the telephone remains unknown for the most part. Due to some implementation problems, the results of this experiment for the three tested systems can not be compared on a low level. To do this, the systems should be more comparable to each other.

What can be compared is the perception of the three systems by the subjects. When doing this, the human operator interface turns out to be the most liked system, followed by the keypad entry system. The speech recognition system is liked the least by the subjects. This ranking of the tested systems must be considered a ranking of the systems as they are implemented at this moment; it is not a ranking of the techniques used in those systems!

The performance results are also influenced by the current implementation of the tested systems. When designed in an optimum way, the speech recognition system could have yielded better results. It then would also not be necessary for the wizard to help some of the subjects in choosing names from the lists, which influenced the data.

The results of this experiment show no age effects or gender effects. Subjects older than 65 feel the same about the tested systems and have the same performance as subjects younger than 65. The same holds true for male and female subjects.

A lot of subjects did not like the name lists that were used to let subjects pick the right name after having only entered the first few letters of that name. Probably, this is a result of the confusion that was raised in the speech recognition system on how to pick a name from the list. If the way of doing this in the speech recognition system is changed, a lot of resistance to the name lists will probably be gone because such lists have a big advantage: only a few letters have to be typed or spelled before a name is recognized.

The necessary change in the speech recognition system for these name lists, is keeping the same modality throughout the whole dialogue; i.e. the confirming/denying modality ('yes'/'no') should also be applied to the name lists. This can be implemented in several ways. The most promising is giving people the opportunity to say 'yes' while the list is being read to them. A technically easier alternative is reading the name lists in parts. People then have to say after each name 'yes' or 'no' to pick the right one from the list. It was found by Virzi et al. (1992) that these so-called skip and scan menus yield good results when used in telephone information systems.

Another thing that can be changed to the speech recognition system is leaving out the special commands. They were not used by the subjects, and leaving them out will shorten the instructions. This speeds up the system and is a solution to the biggest complaint by the subjects: the messages being too long. For the same reason, the introduction message should be shortened.

For the keypad system two things should be changed. First, the possibility to interrupt messages should be mentioned in the instructions. Second, the system should account for people who keep typing letters while the name list is already read to them. This can be implemented by continued checking of the extra typed letters to the name-database. If these extra entries turn out to be valid, the name list should immediately be stopped and a new list should be generated. In most cases however, only one name will then have to be read to the user.

One result from this experiment is astonishing: some subjects did not hear the difference between the sampled human voice and the synthetic voice. Considering the quality of the synthetic voice, this is remarkable. Also, synthetic spoken names in a list seem to be intelligible enough. Not all of the subjects could understand all of the names being read to them, but all could pick out the right name from the lists. This can be understood when one realises that picking the right name from the list is basically a recognition task, in which people perform better even if the noise (the other names) reaches a high level.

When using the results of this experiment, a few considerations must be made. The experimental situation is different from that of the situation in which people would use the proposed car navigation system via the telephone. As subjects mentioned, there is a need for paying close attention when interacting with the systems. In 'real life' this could raise some troubles. The same goes for typing and spelling names letter-by-letter. Sometimes you know the name of the street, but the spelling can be a problem e.g. with French names. Also, the name signs of streets are sometimes absent or hard to read.

Using a hand-held telephone may yield other problems. With the desk telephone used in this experiment, it is easy to listen and type on the keypad at the same time. With a hand-held telephone however, this is far more difficult. Also, it can be very dangerous to interact with the keypad system, or better with any telephone system, when at the same moment driving a car. Therefore, designing a user interface for the proposed telephone system should be done with great caution.

References

CCITT (1993). Arrangement of digits, letters and symbols on telephones and other devices that can be used for gaining access to a telephone network. CCITT.

Dintruff, D. L. et al. (1985). User Acceptance of speech technologies. *Speech Technology* (Feb./Mar.): 16-21.

Garlich, A. (1994). A literature study on alphabetic input using telephone keypads. Report no. 967. Institute for Perception Research (IPO).

Gould, J. D. and S. J. Boies (1987). Human factors challenges in creating a principal support office system - the speech filing system approach. *Readings in human-computer interaction: a multidisciplinary approach.* San Mateo, California, Morgan Kaufmann Publishers, Inc. 25-37.

Halstead-Nussloch, R. (1989). *The design of phone-based interfaces for consumers.* Proceedings of CHI'89 Human factors in computing systems, Austin, Texas.

Nielsen, J. and J. Levy (1994). Measuring usability; preference vs. performance. *Communications of the ACM* 37(4): 66-75.

Nusbaum, H. C. and D. B. Pisoni (1987). Automatic measurement of speech recognition performance: a comparison of six speaker-dependent recognition devices. *Computer Speech and Language* (2): 87-108.

Pols, L. C. W. (1977). *Spectral analysis and identification of Dutch vowels in monosyllabic words.* Amsterdam, Academische Pers B.V.

Roelofs, J. C. (1987). Synthetic speech in practice: acceptance and efficiency. *Behaviour & Information Technology* 6(4): 404-410.

Schumacher, R. M. J. (1992). *Phone-Based Interfaces: Research and Guidelines.* Proceedings of the Human Factors Society 36th Annual Meeting, Atlanta, Georgia, The Human Factors Society.

Virzi, R. A. et al. (1992). *Skip and Scan Telephone Menus: User Performance as a Function of Experience.* Proceedings of the Human Factors Society 36th Annual Meeting, Atlanta, Georgia, The Human Factors Society.

Appendix A   Names to type or spell

Total number of streets in Eindhoven region database: 3912

All streets should be typed/spelled without hyphens, spaces, dashes and dots.

A program was written that calculated the necessary number of letters that has to be entered. This program simply looked at the first letter of all the street names and put all the streets starting with the same letter in the same group. The names in these groups were then scanned on the second letter. Again, names with the same letters were put in the same group. This process continued up to the point that all groups had 5 or less names in it.
For calculating the necessary number of letters for typing, the names were first converted into their numerical representation according to the chosen telephone keypad layout. In this case no letters were compared but numbers. The grouping of the 'names' up to 5 or less was then started.

Entering names with official order of name chunks

Necessary numbers for typing:	≤5	3184	81.4%
	6	380	9.7%

			91.1%
Necessary letters for spelling:	≤5	3353	85.7%
	6	226	5.8%

			91.5%

Entering names with non-official order of name chunks (most distinguishable*)

Necessary numbers for typing:	≤5	3456	88.3%
	6	349	8.9%

			97.2%
Necessary letters for spelling:	≤5	3633	92.9%
	6	181	4.6%

			97.5%

* Most distinguishable means that part of the name that is the most unique within the entire database; e.g.:

not "het Veken" but "Veken",
not "'t Bijsterveld" but "Bijsterveld",
not "1e Franklinstraat" but "Franklinstraat",
not "A.P. de Bontstraat" but "Bontstraat",
not "Aert van der Neerstraat" but "Neerstraat", and
not "Generaal Montgomerylaan" but "Montgomerylaan".

Appendix B   Letters on Dutch Telephone Keypads

A. Garlich

Abstract

This appendix is the sum-up of the inquiry into what telephone keypad letter-layout is the most suitable for the Netherlands. A lot of different letter-layouts were found in the current situation. Some of these seem very usable, others are very country-specific. No special Dutch standard for letter-layouts was found. The juridical consequences turned the scale for recommending the new European standard as being the most suitable for the Netherlands.

In IPO report no. 967 (Garlich, 1994), the results of a literature study on alphabetic input using telephone keypads were presented. In that report it was concluded that the single keying technique was superior to other keying techniques. However, the results were only valid for the situation in the United States, in which the letter-layout of figure 1 is being used; i.e. the Blauvelt-layout. In the conclusion a remark was made saying the Dutch situation was somewhat different from that in the US. The main question that remained from that report was which letter-layout would be the most suitable in the Netherlands. This report will answer that question.

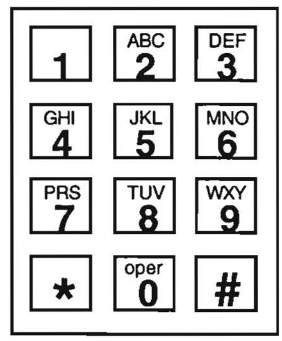


Figure 1: US letter-layout.

Dutch standard?

Several phone calls to relevant companies and governmental departments made clear, that there is no Dutch standard for the letter-layout for use on telephone keypads. PTT (now known as KPN) does not even print letters on their desk telephones. Currently however, PTT-phones come to the market that do have letters assigned to the keys. These telephones are almost all mobile or cordless telephones and are manufactured by several companies. At choosing from all the available phones, PTT does not look at the letter-layout: "As long as it looks nice, the keypads are of no concern to us" (PTT Logistic Telecommunication / Mobile). On inquiry at Dr. Neherlab, it turns out that there is no standard within PTT that deals with keypad letter-layouts. Therefore, the current telephones that are sold by PTT do have several different letter-layouts for the keypads. In the figures 1 to 3, some of these layouts can be seen.

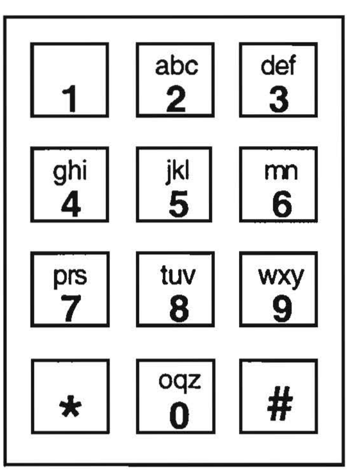


Figure 1: Letter-layout used by Nokia.

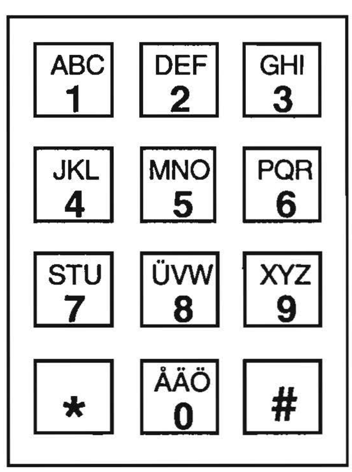


Figure 2: Letter-layout used by Ericsson.

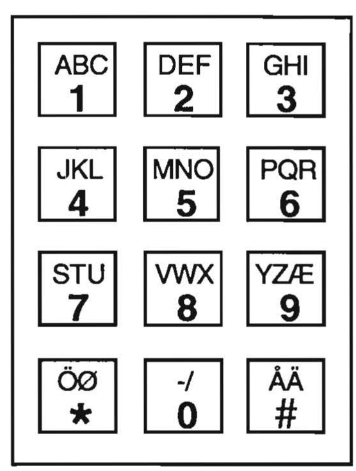


Figure 3: Letter-layout used by Motorola.

The Nokia-layout is used for the PTT Pocketline series "Einstein" and "Vespucci" and for the PTT Carvox "Europa". The Ericsson-layout can be found on the PTT Pocketline "Nobel" and the PTT Carvox "Compact". The Motorola-layout can be seen on the PTT Pocketline "Columbus plus" and the PTT Carvox "Cruiser".

The layouts used by the manufacturers do not confirm to any standard. It can also be seen that some country-specific letters are sometimes used by a certain manufacturer. Designers at Philips say, they use *the* European standard layout for their telephones. Mr. Meurquez of PCD pointed out that this means ABC on the 1-key, and so on. This layout, that can be seen in figure 4, is however never approved as the European standard.

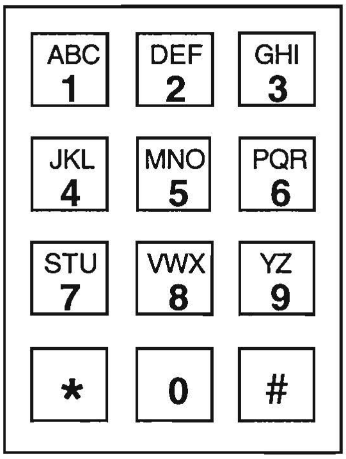


Figure 4: The European standard?

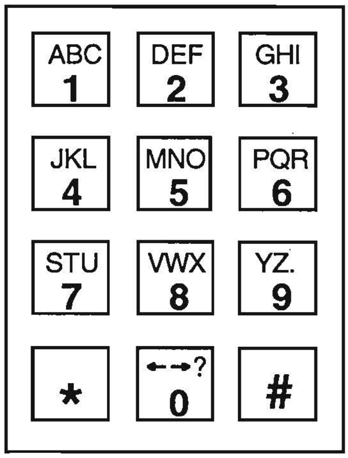


Figure 5: Letter-layout for "teksttelefoon".

Besides the layouts that are chosen by the manufacturers, PTT itself has used a letter-layout on their phones for a special purpose: the deaf-phone. It is still possible to get the letter-layout stickers from Primafoon that were used for this so-called "Tekst-telefoon". In figure 5 the used layout can be seen[1]. This layout was designed in 1984 by the Dutch PTT and the Swedish PTT (no reference available). It is currently re-introduced on GSM-phones for sending messages up to 160 characters; the so-called Short Message Service (Oorschot, 1994).

Finally, there is an interesting development within PTT with using letters on telephone keypads. For the new numbering plan (introduction in October 1995) it is considered to use letters for the so-called mnemonic dialing which is currently used in the United States. The reason for this is that the new 7-digit numbers (the "abonneenummer") are harder to memorize than the old 5-digit numbers (for small and medium cities and villages). Including a few letters in the numbers, might make this a bit easier. However, at the moment these plans are just an intention!

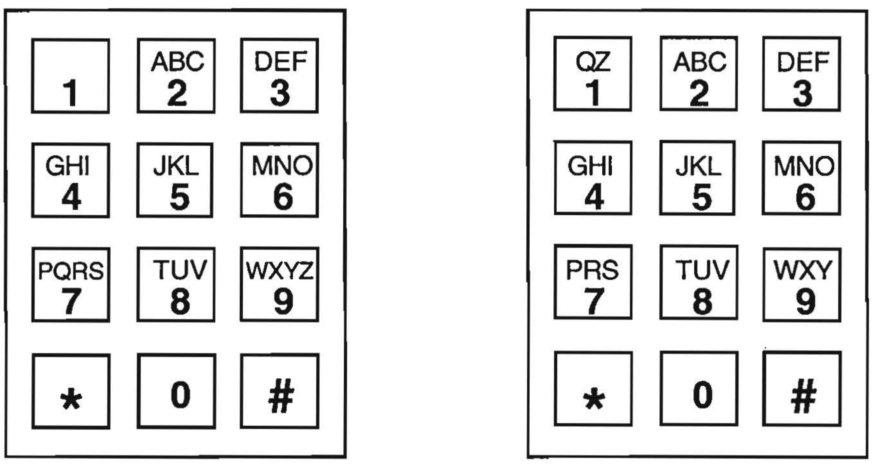
Besides PTT, there are a lot more tenders at the telephone market. Most of these are Japanese which adhere to the US-layout for their keypads. A lot of these 'cheap phones' can be found in Dutch households as a second telephone. With some of these phones it is possible to use the letters on the keys for keeping and updating a number database. This option is always implemented by means of a double-keying technique, because each letter must be specified unequivocally (see Garlich, 1994).

The governmental department of "Verkeer en Waterstaat" also has no standard for assigning letters to telephone keypads. This department, specifically the HDTP ("Hoofddirektie Telecommunicatie en Post"), is also responsible for the testing of telecommunication equipment which manufacturers want to put out on the Dutch market. However, at approval of telephones the outside of the phones is not considered. Only the electrical specifications are tested.

[1] In this situation, the deaf person has a telephone with a little display on which the letters that are typed by the other party are displayed. For picking letters, a double-keying technique is used: For picking the left letter on a key, first the *-key has to be pressed, followed by the key on which that letter is located; For the right letter, first the #-key has to be pressed; For the middle letter only the key with that letter has to be pressed. For picking numbers, one has to press the *-key then the #-key and then the key on which that number is located.

European standard?

On inquiry for a European standard, several things were found. First, there is a CCITT recommendation E.161 of 1993 (CCITT, 1993) in which two options are recommended; see figure 6. Option B is the 'extended Blauvelt' layout that is often used on Japanese phones. Second, option A is also recommended by the ISO/IEC in their Draft proposal 9995 part 8 (ISO/IEC, 1993). Third, the ETSI has chosen the same option A as being the European standard (ETSI, 1993). All other layouts that are used in Europe, including figure 6 option B, are recommended to be phased out by 1 April 1996. The members of ETSI were aware of the fact that the proposed layout was not the most ideal one, from a human factors viewpoint; e.g. some keys having three and others having four letters printed on them. But "in the light of the present ITU [the former CCITT] and ISO activities there is no point for ETSI in trying to introduce a separate European solution, even if it is obvious that the layout most often used in Europe has advantages over the proposed international standards. It is simply too late" (ETSI, 1993, p 2).



Option A Option B

Figure 6: Letter-layouts in CCITT recommendation E.161

The chosen layout is equal to the US-layout, except for the extra letters on the 7-key and the 9-key. This was necessary because the Q and the Z, in contrast with the United States, are often used in Europe. As a consequence, the 7-key an the 9-key have four letters on them. Using such a keypad in combination with a double-keying technique, will therefore yield some problems.

Also, the chosen layout has no provision for specific national characters. Especially for the Scandinavian countries, this is a pity (see figure 2 and 3). Although these countries are now part of the European Community, at the moment ETSI decided on the layout they were not. They probably would have voted against the proposed layout, but it now is a fact that the only true European standard for the telephone keypad letter-layout, is the one that is marked option A in figure 6.

Conclusion

As it turns out, there is no Dutch standard on assigning letters to telephone keypads. However, there is a European standard that will be effective on 1 April 1996. Manufacturers and governments within the European Union are legally bound to this standard. Therefore, it is recommended to use the proposed European standard (Figure 6, option A) as the only valid telephone keypad letter-layout in the Netherlands.

References

CCITT (1993). Arrangement of digits, letters and symbols on telephones and other devices that can be used for gaining access to a telephone network; CCITT Recommendation E.161, CCITT.

ETSI (1993). Assignments of alphabetic letters to digits on keypads; ETSI TC-HF, ETSI.

Garlich, A. (1994). A literature study on alphabetic input using telephone keypads. IPO report 967. Institute for Perception Research.

ISO/IEC (1993). Numeric zone of the numeric section, allocation of letters; ISO/IEC Draft proposal 9995 - 8.2, ITU.

Oorschot, W. (1994). GSM-berichtendienst verhoogt bereikbaarheid. Telecommagazine, februari 1994, p 11.

Several folders of PTT Primafoon / PTT Businesscentre.

Personal communication from several people from within PTT, HDTP, and Philips.

Appendix C Subject's tasks

For task order 123:

Bel het Autoroute-systeem en zorg dat u een routebeschrijving krijgt om van uw vertrekpunt naar uw bestemming te komen.
Als u een opdracht heeft volbracht, leg dan de hoorn weer op de haak en ga naar de volgende opdracht.

Het telefoonnummer van Autoroute is: 06 - 2886

1 Vertrekpunt: Glaslaan in Eindhoven
Bestemming: Klavierhoek in Franeker

2 Vertrekpunt: Twentelaan in Eindhoven
Bestemming: Telexweg in Franeker

3 Vertrekpunt: Gildelaan in Eindhoven
Bestemming: Achterstraat in Franeker

For task order 321:

Bel het Autoroute-systeem en zorg dat u een routebeschrijving krijgt om van uw vertrekpunt naar uw bestemming te komen.
Als u een opdracht heeft volbracht, leg dan de hoorn weer op de haak en ga naar de volgende opdracht.

Het telefoonnummer van Autoroute is: 06 - 2886

1 Vertrekpunt: Gildelaan in Eindhoven
Bestemming: Achterstraat in Franeker

2 Vertrekpunt: Twentelaan in Eindhoven
Bestemming: Telexweg in Franeker

3 Vertrekpunt: Glaslaan in Eindhoven
Bestemming: Klavierhoek in Franeker

Appendix D Interview before the experiment

In de vragen die ik u straks zal stellen, komt de term "geautomatiseerde dienst over de telefoon" veelvuldig voor. Hieronder verstaan we een systeem waarbij u een bepaalde dienst kunt bellen via een telefoon. De instructies over hoe u de dienst moet gebruiken, krijgt u via de telefoon. Omdat de dienst is geautomatiseerd, krijgt u geen telefoniste aan de lijn maar een computer. Met dergelijke diensten is het bijvoorbeeld mogelijk om het saldo op uw bankrekening te weten te komen, en telefonisch kleren te bestellen zonder tussenkomst van een telefoniste.

1 Heeft u al eens gebruik gemaakt van een geautomatiseerde dienst over de telefoon?

O Ja O Nee; *ga naar vraag 5*

2 Wat was het doel van die dienst(en)?

____

____

3 Wat beviel u het meest aan die geautomatiseerde dienst(en)?

____

____

4 Wat beviel u er het minst aan?

____

____

5 Wat is uw algemene indruk van geautomatiseerde telefoondiensten?

Ze zijn:

O Erg slecht O Slecht O Geen mening O Goed O Erg goed

Waarom vindt u dat?

____

____

6 t/m 10 worden door de proefpersonen zelf ingevuld op het nu uit te reiken papier.

6 Hoe vaak reist u voor uw werk buiten de provincie?

O Nooit

O Eens per jaar
O Twee keer per jaar
O Vier keer per jaar

O Eens per maand
O Twee keer per maand

O Eens per week
O Twee keer per week
O Drie of meer keer per week

7 Hoe vaak reist u in uw vrije tijd buiten de provincie?

O Nooit

O Eens per jaar
O Twee keer per jaar
O Vier keer per jaar

O Eens per maand
O Twee keer per maand

O Eens per week
O Twee keer per week
O Drie of meer keer per week

8 Wat is het meest belangrijke aspect voor u als u auto rijdt?

Breng in onderstaande mogelijkheden een rangorde aan.

___ een zo kort mogelijke route qua afstand
___ een zo kort mogelijke route qua tijd
___ een zo mooi mogelijk uitzicht
___ een zo laag mogelijk benzineverbruik
___ een goede routebeschrijving
___ de gemakkelijkste route (minst aantal kruisingen, stoplichten, etc.)

9 Welke hulpmiddelen gebruikt u om de weg te vinden?
(bijv. kaarten, vrienden, voorbijgangers vragen, intuïtie)

____

____

10 In welk jaar bent u geboren? ________

Uitleg van de opzet van de nieuwe telefoondienst

U zult zo dadelijk kennismaken met een nieuwe telefoondienst, die u als autorijder kan helpen de weg te vinden in voor u onbekende plaatsen. Daartoe belt u het 06-nummer van deze dienst en geeft u op in welke straat van welke plaats u op dat moment bent (uw vertrekpunt) en naar welke straat van welke plaats u uiteindelijk wilt (uw bestemming).
De telefoondienst berekent daarna de beste route voor u en houdt daarbij rekening met wegversperringen en files.
De routebeschrijving wordt u daarna opgelezen, waarbij u met behulp van de druktoetsen de verschillende stappen net zo vaak kunt horen als u wilt. Als u vanuit uw auto belt kunt u dit doen terwijl u rijdt. Als u niet vanuit uw auto belt kunt u de routebeschrijving op een bandje opnemen dan wel zelf opschrijven. Het is ook mogelijk om de routebeschrijving op een Fax toegezonden te krijgen.

Vervolg interview

11 Als u naar een u onbekende plaats toe gaat, wanneer koopt u dan een plattegrond van die plaats?

O Voordat ik er naar toe ga ______________________

O Direkt als ik in die plaats aankom ______________________

O Ergens gedurende mijn verblijf daar ______________________

O Alleen als ik er verdwaal ______________________

O Voordat ik er weer vertrek ______________________

O Nooit ______________________

12 Plant u uw reizen voordat u weggaat in detail?

O Altijd O Soms O Nee; *ga naar vraag 14*

13 Als u naar u onbekende plaatsen reist, hoe plant u dan uw reis?

____

____

14 Als u in een stad de bezienswaardigheden wilt bezoeken, hoe plant u dan de volgorde waarin u dat gaat doen?

____

____

15 Als u denkt dat u verdwaald bent, wat doet u dan meestal?

(bijv. in paniek raken en niet weten wat te doen, iemand vragen, zelf proberen de weg weer te vinden, een kaart raadplegen)

____

Proefpersonen dienen nu de beschrijving van de nieuwe telefoondienst te lezen (staat op de achterkant van de zojuist door hen ingevulde vragenlijst).

16 Wat vindt u van zo'n telefoondienst?

____

____

17 Zou u deze telefoondienst gebruiken?

O Ja; *ga naar vraag 18*

O Nee; Waarom niet? ______________________________

ga naar 20

18 Hoeveel zou u voor zo'n telefoondienst willen betalen? ____________________

(minder dan 50 cent per minuut, 50 cent per minuut - 1 gulden per minuut, 1 - 2 gulden per minuut, meer dan 2 gulden per minuut)

19 Stel dit tarief wordt ingevoerd.
Hoe vaak denkt u dat u deze telefoondienst dan zult gaan gebruiken? ______________

(Eens per jaar, twee keer per jaar, vier keer per jaar, eens per maand, twee keer per maand, eens per week, twee keer per week, drie of meer keer per week)

20 Geslacht **[niet vragen]**: M / V

Opmerkingen:

____

____

____

____

____

____

____

____

____

____

____

____

Appendix E Interview after the experiment

1 Wat is uw algemene indruk van deze telefoondienst?

____

____

2 Wat beviel u het meest aan deze geautomatiseerde telefoondienst?

____

3 Wat beviel u er het minst aan?

____

4 Welke manier van het opgeven van de namen beviel u het meest; druktoetsen gebruiken, namen letter voor letter spellen, of namen voluit uitspreken?

Manier: ____________________

Waarom?

____

____

5 Welke manier van het opgeven van de namen beviel u het minst?

Manier: ____________________

Waarom?

____

____

6 Bij het intoetsen en inspreken van losse letters moest u soms twee letters indrukken of zeggen en een andere keer drie of slechts één. Vond u het vervelend of lastig dat dit aantal af en toe verschilde?

O Ja; Zou u liever altijd het zelfde aantal letters willen indrukken of zeggen?
  O Ja
  O Nee
O Nee
O Onverschillig (maakt niet uit)

7 Was de stem die gebruikt werd in deze telefoondienst goed verstaanbaar?

O Ja
O Nee

8 Hoorde u bij het oplezen van straat- en plaatsnamen een andere stem?

O Ja; Wat vond u van deze stemwisselingen?

____

____

O Nee

9 Als de namen aan u werden voorgelezen, waren ze dan verstaanbaar?

O Ja
O Nee

10 Zou u deze telefoondienst gaan gebruiken als u de letters op de druktoetsen zou moeten intoetsen?

O Ja; *ga naar vraag 11*

O Nee; Waarom niet? ___________________________________

ga naar vraag 13

11 Hoeveel zou u voor zo'n telefoondienst willen betalen? ____________________

(minder dan 50 cent per minuut, 50 cent per minuut - 1 gulden per minuut, 1 - 2 gulden per minuut, meer dan 2 gulden per minuut)

12 Stel dit tarief wordt ingevoerd.
Hoe vaak denkt u dat u deze telefoondienst dan zult gaan gebruiken? ________________

(Eens per jaar, twee keer per jaar, vier keer per jaar, eens per maand, twee keer per maand, eens per week, twee keer per week, die of meer keer per week)

13 Wat voor telefoon heeft u thuis?

O draaischijf
O druktoetsen

Opmerkingen:

____

____

____

____

Appendix F Semantic differential form

slecht								goed
niet prettig								prettig
niet flexibel								flexibel
ongeschikt								geschikt
onpersoonlijk								persoonlijk
langzaam								snel
onbruikbaar								bruikbaar
inefficiënt								efficiënt
chaotisch								geordend
onaangenaam								aangenaam
niet tevreden stellend								tevreden stellend
moeilijk te volgen								makkelijk te volgen
niet functioneel								functioneel
gecompliceerd								eenvoudig
niet indrukwekkend								indrukwekkend
afstotend								uitnodigend
onvoorspelbaar								voorspelbaar
moeilijk te gebruiken								makkelijk te gebruiken
passief								aktief
gespannen								ontspannen

Mailing list

Bouwhuis
Garlich (3 ex.)
Kemp
Van Nes
Tang
Vogten

IPO-directeur
A.C. der Kinderen, Bibl. en Inf. Nat.Lab., WY 1.36 (8 ex.)
Dr.ir. W. Strijland, Octr. en Merken, WAH

5 archief