I can't Hear You? Drivers Interacting with Male or Female Voices in Native or Non-Native Language

Ing-Marie Jonsson and Nils Dahlbäck,

Department of Computer and Information Science, Linköping University, SE-581 83
Linköping, Sweden
ingmarie@ansima.com, nda@ida.lu.se

Abstract. Many vehicles today are equipped with navigation systems, and all of these systems use speech or a combination of speech and graphics to provide drivers with directions to their destinations. This study investigates the effect of gender of voice when providing driving instructions in English to drivers that are non-native speakers of English. In a 2(native/non-native) by 2(gender of voice) between participant study, 40 participants in age group 18-25 drove in a driving simulator for 25 minutes with navigation information system that gave drivers directions to a set destination. Results show that gender of voice did not affect native English speaking drivers. For non-native speakers, however, a female voice worked better for both female and male drivers. Non-native speakers consistently missed to act on navigational information give by the male voice. Design implications for voice systems are discussed.

Keywords: In-vehicle Information System, Navigation systems, Voices, Gender, Non-native speakers, Driving Performance.

1 Introduction

Navigation systems are becoming more and more common in vehicles, rising from 5.5 million installed units in 2004 to an estimated 16.2 million units in 2009 [1]. Users of navigation systems doubled from 2006 to around 50% of drivers in 2009 [2]. Whether it is a portable unit or built into an automobile's dashboard, navigation systems that use radio signals from global positioning satellites to pinpoint an exact location, or plot travel directions are becoming popular items for drivers. They come as standard in high-end vehicles from manufacturers like BMW, or Mercedes, and can be bought as optional extras in many other makes of vehicles. Navigation systems have still not seen the same market share in the US as it has in Japan or Europe. This could partially be explained by how the layout of cities and roads follow uncomplicated grid patterns in the US but not in Europe and Japan. It is however worth to note that navigation systems are most often offered and in demand by rental car companies and trucking fleets. It makes sense to offer navigation systems since car renters are often not local to or familiar with the area where they rent the car. Navigation systems in trucks potentially save both time and money when drivers don't get lost. Drivers using navigation systems are most often provided driving direction to a destination by speech or a combination of speech and graphics. Some

solutions also enable you to select and change the voice used by your navigation system.

When renting a car in a foreign country, drivers often have to use a navigation system communicating in their non-native language. In the vast majority of cases this happens to be English for non-native English speakers. In previous studies it has been shown that first and second language users of English voice systems have different voice preferences [3, 4]. A limitation of previous work is that it has been solely concerned with attitudes towards the information presented. But in car navigation systems it is even more important that the users are able to take in and act on the information presented.

To select female voices for navigation systems is based on old studies from the aviation industry where it was found that the higher pitch of a female voice was easier to understand in a noisy environment [5, 6]. However, recently it has been argued that this is not is not a general feature of female voices, but a consequence of specific situational factors.

"It has nothing to with acoustics or taste," said Judy Edworthy, a professor of applied psychologist at the University of Plymouth in England who specializes in "alarms, auditory warnings, beeps and buzzer"[7]. "They used female voices because they were different from most voices heard in the then male dominated aviation sector," said Dr. Edworthy, "and the men were more likely to pay attention to them, particularly in combat situations."[7]

Professor Clifford Nass, Stanford University offered a similar explanation. "The main reason you have female voices in cars is not the technical qualifications like hearability," said Dr. Nass. "It's that finding a female voice that is pleasing to almost everyone is infinitely easier than finding a male voice."[7]

The present study investigates these assertions in a single experimental study where either a male or a female voice was used by a navigation system driven by native and non-native English speakers. The study also addressed other factors such as the spatial abilities of participants. In this paper, however, we only report on those aspects of the study concerned with male and female voices for native and non-native speakers.

2 Method

2.1 Experimental Apparatus

To investigate issues of using male or female in navigation systems we used a driving simulator - STiSIM - from Systems Technology Inc. Drivers were seated in a real car seat and 'drove' using a Microsoft Sidewinder forced feedback steering wheel and pedals consisting of accelerator and brake. The simulated journey was running on a laptop and was projected onto a six-foot projection screen. The same simulator set-up was used for both the training course and the main driving course.

Two tests from Woodcock Johnson's battery of cognitive abilities tests (WJ-III) [8] were used for auditory processing. The two tests on Auditory Processing were the following: First, test 4 on Sound Blending that is a test of synthesizing language

sounds where subjects are asked to listen to syllables and phonemes and then to construct words from them; and Second, test 8 on Incomplete Words that is a test of auditory processing where subjects are asked to listen to words where phonemes are missing and then re-create the correct word. The cognitive abilities tests were administered according to WJ-III instructions using a stopwatch, earphones and WJ-III test booklets [8].

2.2 Design and Participants

The experiment was a 2 (English speaker: native or non-native) by 2 (gender of car voice: Male or Female) between-participants design, with random assignment to condition. Gender was balanced across conditions.

There were a total of 40 gender-balanced participants in the age group 18 to 25. 26 of the participants were native English speakers (12 Male and 14 Female), and 14 were non-native English speakers (8 Male and 6 Female). All participants were students at Oxford University, and as such, non-native English speakers are required to present a certificate with a high level of English Language skills before being accepted [9]. All participants had a driver's license and between one and five years of driving experience. Participants gave informed consent and were debriefed after the experiment.

2.3 Procedure

Each participant started the experimental session by signing the consent form. They were then seated at the driving simulator and drove the training course with verbal guidance from the experimenter. The training driving course was 5 000 feet and took approximately three minutes to finish. The purpose of the short introductory driving session was to familiarize participants with the control and feedback from the driving simulator. This course also provides screening for participants that suffer from simulator sickness [10].

None of the 40 participants felt nauseous or discomfort during or after the training course, and continued to fill in the first questionnaire with general information such as gender and age in addition to driving experience. After the questionnaire, the Auditory processing tests were administered before the participants sat back down into the driving seat again.

All participants used the same simulator configuration and the same driving course in the experiment. The course was 52 000 feet long and took on average 22 minutes to complete. All drivers completed the same driving scenario since a driving scenario in STISim is static and predetermined; it has a specific length and will take all drivers along the exact same road regardless of left and right turns. Based on this feature of STIsim all participants are guaranteed to drive the exact same route.

The driving scenario was scripted to take the driver along a road with a navigation system that gave directions to five destinations. The content of the driving directions were designed based on a storyboard description. A simulator road scenario was

designed to include the scripted destinations and the simulator was instrumented to generate reports on when drivers followed instructions at choice points.

The road scenario was divided into 38 driving sections, where directions or suggestions prompted 32 sections, and 6 sections were prompted by facts about the immediate surroundings. Directions and suggestions were designed to guide the drivers to 5 pre-programmed destinations. The facts were added to investigate how much attention the drivers were paying to the system. The 38 speech prompts were recorded in a male voice and a female voice. The Participants were randomly assigned to one of two conditions, the navigation system with the female voice or the navigation system with the male voice.

After completing the course, participants filled out a post-test questionnaire assessing their own driving and the navigation system.

2.4 Measures

Auditory Processing. This measures the average of the Woodcock Johnson's cognitive abilities tests on sound blending and incomplete words.

Driving Performance. This is a collection of measures that consists of accidents and adherence to traffic regulations. The measure for *Accidents* is comprised of collisions and off-road accidents. *Swerving* is defined as drivers crossing centre-line or entering shoulder. *Adherence to traffic regulations* is comprised of speeding violations, running red lights and stop signs.

Navigation System. This is a collection of measures that consists of instructions followed, time to destination, and facts remembered. The measure *Instructions followed* simply counts how many of the driving instructions drivers followed. *Time to destination* measures drivers' time to complete the driving scenario to the last destination. *Fact remembered* is a measure of how many of the driving scenario facts that drivers remembered after the driving session ended.

3 Results

The effects of the male and female English voices, when used in a navigation system by native and non-native English speaking drivers, were measured by a two (Gender of Navigation System voice) by two (Native and Non-native English speaker) between-participants ANOVA.

3.1 Auditory Processing

The results from the WJ-III tests on Auditory processing show a significant difference between native and non-native English speakers, $F(1,38)=12.5$, $p < .001$. There were no gender effects in auditory processing.

Table 1. Auditory processing for native and non-native speakers of English

	N	Auditory processing	
		Mean	SD
Native speakers	26	49.26	7.32
Non-native Speakers	14	39.86	9.23

3.2 Prior Driving Experience

To ensure that there were no initial differences between drivers in experimental conditions, drivers' prior driving experience was used. Data from the 2 most recent years of driving was collected, such as the number of accidents, and the number of tickets. This data was averaged for each group of drivers, and no significant differences were found across conditions.

3.3 Driving Performance Measures

Adherence to Traffic Regulations. There were no main effects of Native and non-Native English speakers or gender of voice used by navigation systems for speeding, and stopping at stop signs and traffic lights. This confirms the results from the data on prior driving experience. See Table 2.

Accidents. When driving with the navigation system, non-native English speakers had significantly more accidents than native English speakers, $F(1,36)=8.5$, $p < .006$. There was no effect of gender of voice. See Table 2.

Swerving. There is a main effect where non-native English speakers swerve more than native English speakers. $F(1,36) = 8.5$, $p < .007$. There is no main effect of gender of voice.

More interesting, there is an interaction effect for swerving. Where non-native English speakers driving with the Male voice in the Navigation system swerve significantly more than Native English speaking drivers $F(1,36)=10.8$, $p < .002$ Native English and non-native English speaking drivers with the Female voice swerve equally much, see Table 2.

Table 2. Measures of Driving Performance

		Native Speakers		Non-Native Speakers	
Gender of Voice		Male	Female	Male	Female
Accidents	M	2.4	2.4	4.6	3.3
	SD	1.1	1.7	2.4	1.7
Swerving	M	18.2	30.9	51.4	28.9
	SD	3.7	26.3	13.0	11.4
Adherence to traffic regulations	M	3.9	7.2	4.8	5.2
	SD	2.4	47	3.0	4.5

3.4 Navigation System Measures

Instructions Followed. This data show how well drivers followed driving directions such as turn right, turn left or continue straight, show two main effects; Gender of Voice, with Female voice better than Male $F(1,36) = 40.3$, p < .001, and Native speakers follow direction better than Non-native speakers, $F(1, 36)=46.0$, p < .001.

Once again we have interesting interaction effect. Non-native English speakers that drove with a navigation system with a male voice did not follow directions as well as drivers in other conditions, $F(1,36) =66.0$ p < .001. There were no significant differences for drivers that heard the female voice. See Table 3.

Table 3. Measures for Navigation System

Gender of Voice		Native Speakers		Non-Native Speakers	
		Male	Female	Male	Female
Following instructions	M	18.8	17.4	6.6	18.4
	SD	1.7	2.7	2.1	3.0
Time to Destination(s)	M	1310	1304	1552	1275
	SD	198	273	59	81
Facts remembered	M	2	2.2	.8	2.3
	SD	1.3	1.7	.8	1.2

Time to Destination. Data show that it is a main effect of gender of Voice, where drivers with the Female voice reached their destinations quicker than drivers with a Male voice, $F(1, 36)=4.6$, p <.05. There were no effects of being a native or non-native English speaker. There was however an interaction effect so that non-native drivers listening to a system with a Male voice took significantly longer to reach their destinations than other drivers, $F(1,36) = 4.2$, p < .05. See Table 3.

Facts Remembered. There was also a non-significant but interesting trend in the data that show that fewer facts are remembered when uttered by a Male voice than when uttered by a Female voice, $F(1,36) = 3.2$, p < .08. There were no other effects. See Table 3.

4 Discussion and Conclusion

In the study presented in this paper this we addressed two aspects of voice interaction. First differences between native and non-native language users, and second, differences between information presented with a male or female voice.

For differences between native and non-native speakers of English, it is clear that native and non-native English speakers do not share the same preferences for gender of voice. Or rather, the data show that for a non-native English speaking driver in general a female voice works better. Whereas for native English speaking drivers, the general pattern is that gender of voice does not matter, with a small but insignificant trend towards gender bias.

It should be admitted that there are exceptions from this pattern on some variables. We suspect that this probably is due to the small number of subjects in the study, and as a consequence, lack of statistical power. But since the data show a pattern that goes against previously established beliefs, we present them here to initiate discussion and not the least to inspire further work on these issues.

These results support the results previously reported by Dahlbäck et al [3, 4]. An important difference between the results presented here and previous work, in that this is to our knowledge the first time that it is shown that the differences between native and non-native speakers also affect performance measures. This is true both for direct measures, as shown by the results on the *instructions followed* and the *facts remembered* measure (though not significant in the latter case), and for indirect ones, measuring a secondary task, as shown by the results on the *swerving* and *time to destination* measures.

The results show that there is some truth to the assumption that female voices are better than male voices in information systems. On the measures (*swerving, instructions followed,* and *time to destination* the data show that drivers that had a navigation system with a female voice performed better. But the significant interaction effects found in the data also show that the female does indeed not work best in all cases. In fact, the results clearly indicate that the advantage gained by using a female voice only holds for non-native English speakers. For native speakers, on the other hand, our data largely support the views of Nass and Edworthy [7].

This suggests that designers, at least when designing systems for predominantly native language users, should base selection of male of female voices more on other considerations including gender bias than on the intelligibility of female voices in noisy environments as presented in this paper. For instance, conforming to gender bias, a male voice could be used for male drivers and a female voice by female driver.

Jonsson, Nass and Harris indicate gender bias as a possible explanation for their results in a driving simulator study with native English speakers from 2008 [11]. The most striking result presented in their research is the dramatic difference between female and male drivers with respect to the effects that the in-vehicle system (that used a female voice) had on driving behaviour. Female drivers showed improved driving performance on all driving and attitudinal measures when driving with the in-vehicle system. The situation for male drivers was very different. The in-vehicle system had a positive effect on only two of six performance and attitude measures. Gender bias, related to the voice of the in-vehicle system, was offered as a plausible explanation for these differences.

Another strategy would be to divide the information into different categories, and use different voices for different categories. A male voice, could for instance, be used to convey critical security information in an in-vehicle system. A female voice, could for instance, be used to inform drivers of points of interest along the way.

It is also possible to match the voice of the in-vehicle system to states and traits of the drivers. Results from previous studies show that matching the emotional colouring of the in-vehicle voice to the emotional state of the driver has positive effects on driving performance and attitude [12]. Jonsson et al [13] also show that older drivers can benefit enormously from driving with an in-vehicle system with an appropriately selected voice, while a less appropriate voice can result in weakened driving performance.

It is however clear, from the results presented in this paper, and from previous result on voices used by in-vehicle systems [11, 12 13], that in designing voice interfaces, there is no one solution that fits all users on all occasions. One voice does not fit all!

References

1. Jones, R.: Is driving easier when Mr. T's riding shotgun? MSNBC December 28, 2005 (2005)
2. Smith, S.: NAVTEQ Reveals Data On the State of Navigation, NAVTEQ Press Center (2010)
3. Dahlbäck, N., Swamy, S., Nass, C., Andersson, F., Skågeby, J.: Spoken Interaction with Computers in a Native or Non-native Language - Same or Different? Proceedings of INTERACT 2001, pp. 294 – 301, Tokyo, Japan, July 9-13 (2001)
4. Dahlbäck, N., Wang, Q., Nass, C., Alwin, J.: Similarity is More Important than Expertise: Accent Effects in Speech Interfaces. Proceedings of CHI 2007, pp. 1553 –1556, San José, CA, USA, April 28 - May 3 (2007)
5. Nixon, C. W., Morris, L. J., McCavitt, A. R., McKinley, R. L., Anderson, T. R., McDaniel, M. P., Yeager, D. G.: Female voice communications in high levels of aircraft cockpit noises part I: spectra, levels, and microphones, Journal of Aviation, Space and Environmental Medicine, 69(7), pp. 675 – 683, (1998)
6. Nixon, C. W., Anderson, T. R., Morris, L. J., McCavitt, A. R., McKinley, R. L., Yeager, D. G., McDaniel, M. P.: Female voice communications in high levels of aircraft cockpit noises part II: vocoder and automatic speech recognition systems. Journal of Aviation, Space and Environmental Medicine, 69(11), pp. 1087 – 1094 (1998)
7. Feiler, B.: Turn Right, My Love. New York Times. New York (2010)
8. Woodcock, R., W., Mather, N., McGrew, K.: Woodcock-Johnson III Tests of Cognitive Abilities Examiner's Manual. Itasca, Riverside (2001)
9. University of Oxford, International Students – English Language Requirements, http://www.conted.ox.ac.uk/students/internationalstudents/englishlanguage.php
10. Bertin, R. J. V., Guillot, A., Collet, C., Vienne, F., Espie, S., Graf, W.: Objective measurement of simulator sickness and the role of visual-vestibular conflict situations: a study with vestibular-loss (a-reflexive) subjects. Neuroscience, San Diego, California (2004)
11. Jonsson, I., Harris, H., Nass, C. I.: How accurate must an in-car information system be?: consequences of accurate and inaccurate information in cars, Proceedings of CHI 2008, pp. 1665--1674, Florence, Italy (2008)
12. Jonsson, I., Nass, C. I., Harris, H., Takayama, L.: Matching In-Car Voice with Driver State: Impact on Attitude and Driving Performance, Proceedings of International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, Maine, USA (2005)
13. Jonsson, I., Zajicek, M., Harris, H., Nass, C. I.: Thank you I did not see that: In-car Speech-Based Information Systems for Older Adults, Proceedings of CHI 2005, pp. 1953 --1956, Portland, OR, USA (2005)