Tutorial: Voice and Multimodal Interaction in the Car

Garrett Weinberg
Nuance Communications, Inc.
1 Wayside Rd.
Burlington, MA 01803, U.S.A.
+1-781-565-4814
garrett.weinberg@nuance.com

ABSTRACT

We present a tutorial on voice and multimodal user interfaces in the automotive context. After briefly explaining the technology behind automatic speech recognition (ASR) and text-to-speech (TTS) systems, we will explore the benefits of such systems for reducing driver distraction. Recent trends toward natural-language and multimodal interfaces will also be discussed. Throughout the tutorial we will point out key research results and commercial deployments that merit further reading and study.

Keywords

Speech; voice; ASR; TTS; automotive; multimodal; multimodality; natural language understanding; NLU; voice user interfaces; VUI; driver distraction

1. INTRODUCTION

In the eyes- and hands-busy environment of a moving vehicle, speech input and output has been trusted for over a decade as a safe and effective alternative to manual-visual interaction. This talk examines the roots and branches of voice I/O in cars, including, importantly, how voice can most thoughtfully be combined with other input and output modalities to enhance the overall automotive user experience.

2. OVERVIEW OF THE TUTORIAL

After a brief introduction to ASR and TTS core technologies, the tutorial will cover both commercial deployments and research results in the arena of automotive voice I/O.

We will examine studies that demonstrate the benefits of voice over manual interaction, but then we will take the "devil's advocate" position that a good manual-visual interface is better than a bad voice interface. This will serve as an introduction to voice user interface (VUI) design, a little-understood discipline which is equal parts art and science.

Next we will discuss the trend towards natural language understanding (NLU) voice interfaces, those in which users can speak to the ASR system as if they were speaking to a real person, that is using flexible and verbose sentence structure rather than a stilted-sounding command jargon.

Finally we will survey the literature on multimodal interaction, in which voice, tactile and gestural inputs are combined in various ways with audible, haptic and graphical outputs.

At the close of the talk, participants will be invited to speculate on how NLU and multimodality might push automotive voice interfaces even further into the mainstream, and how the academic community might foster this process—or sound notes of alarm if such systems are being implemented in ways that cause too much driver distraction.

3. SUGGESTED READING

The mathematical underpinnings of contemporary ASR and TTS are discussed in [6] and [4]. [3] offers a more application-focused ASR tutorial. An important survey of voice interfaces' effects on drivers' performance is [1], and [5] examines the role that ASR accuracy plays. [7] covers several crucial VUI design guidelines, and [2] and [8] discuss easy-to-implement forms of multimodality that confer significant usability advantages.

4. REFERENCES

[1] A. Barón and P. Green. 2006. Safety and Usability of Speech Interfaces for In-Vehicle Tasks while Driving: A Brief Literature Review. Technical Report UMTRI-2006-5.

[2] S. Castronovo, A. Mahr, M. Pentcheva and C. Müller. 2010. Multimodal Dialog in the Car: Combining Speech and Turn-and-Push Dial to Control Comfort Functions. In Proc. of Interspeech.

[3] CMU Sphinx project. Basic concepts of speech. http://cmusphinx.sourceforge.net/wiki/tutorialconcepts. Last accessed 24 September 2012.

[4] A. J. Hunt and A. W. Black. 1996. Unit selection in a concatenative speech synthesis system using a large speech database. In Proceedings of the Acoustics, Speech, and Signal Processing (ICASSP), Vol. 1. IEEE Computer Society, Washington, DC, USA, 373-376.

[5] A. Kun, T. Paek and Z. Medenica. 2007. The Effect of Speech Interface Accuracy on Driving Performance. In Proc. of Interspeech.

[6] L. R. Rabiner. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 77, 2, 257-286.

[7] B. Schmidt-Nielsen, B. Harsham, B. Raj, and C. Forlines. 2008. Speech-Based UI Design for the Automobile. In Lumsden, J., ed., Handbook of Research on User Interface Design and Evaluation for Mobile Technology. NRC of Canada, Ottawa. 1, 15, 237-252.

[8] G. Weinberg, B. Harsham, C. Forlines, and Z. Medenica. 2010. Contextual push-to-talk: shortening voice dialogs to improve driving performance. In Proceedings of the 12th international conference on Human computer interaction with mobile devices and services (MobileHCI '10). ACM, New York, NY, USA, 113-122

AutomotiveUI'12, October 17-19, Portsmouth, NH, USA.
Copyright (c) 2012 ACM 978-1-4503-1751-1/12/10 ... $15.00