

Position Paper: VoiceXML Speaker Biometrics Workshop

Background

General Motors Corporation (NYSE: GM), one of the world's largest automakers, was founded in 1908, and today manufactures cars and trucks in 34 countries. General Motors currently offers voice-based services through OnStar, a wholly-owned subsidiary of General Motors and the leading provider of in-vehicle safety, security and communication services. GM's Advanced Technology Silicon Valley Office is investigating standards-based platforms for in-vehicle multimodal applications and would like to cooperate with the W3C on developing web-based standards for these solutions. We feel that multimodal applications are a critical component in the vehicle context, both for identification and verification of drivers and for interaction with infotainment systems.

Identification and Verification

The automotive environment is a promising platform for speech-driven applications, as speech allows users to keep their eyes on the road while interacting with a secondary system. For the case of driver identification in particular, speech is reasonably lightweight and unobtrusive: the driver is not required to remember anything (e.g., a password), nor may confusion be introduced if the driver is identified by a device carried into the vehicle (i.e., I am still "me", whether I'm carrying "my" key fob or my spouse's key fob). However, speech is not the only modality available to identify drivers.

Developers may use a range of techniques to identify or authenticate drivers in addition to speech, such as fingerprint verification or weight sensors. With this multitude of options, developers can create systems that balance the confidence of the identification with the burden placed on the driver to authenticate. For example, if it is critically important that the vehicle know *exactly* who is driving, it may require users to satisfy tests from multiple techniques (e.g., both voice identification and a fingerprint match); less critical functionality, such as for comfort and convenience, may require the driver to satisfy only a subset of tests (e.g., a fingerprint match is "good enough" that the voice match may be skipped).

For these reasons, VoiceXML standards need to account for other modalities of user identification and verification besides voice. This may require optional tags for alternate modalities, or a more general tag structure wherein SIV is just one of many modalities (similar to the W3C's multimodal architecture wherein speech is just one of many multimodal components interfacing with the user).

Infotainment Applications and Development

A more general (but related) requirement for multi-modal vehicle applications is the need for a file structure that integrates VoiceXML and HTML in a standard way. This is necessary to allow developers to create applications that are more suited to the divided attention, eyes-busy context of the driver. Voice and audio will never completely replace visual information (especially for maps and other location-based applications) but these eyes-free modalities can be better integrated to allow driver interaction while in motion without sacrificing safety.

One approach to supporting the development multimodal applications is to require separate VoiceXML and HTML documents that can be served up by the application server separately on a single HTTP request. However, this does not reflect the typical application flow followed by web developers. Web developers use a model wherein the browser requests a single page from the server at a time. Writing software code that serves two pages at a time is inconsistent with this traditional methodology. Another approach is to go back to the XHTML+Voice standard. However, this standard intermingles a subset of VoiceXML tags within the HTML, precluding separation of coding duties between IVR and web developers. Perhaps a more usable file structure will combine VoiceXML and HTML into one file, yet keep them separate enough so that IVR developers and web developers can work independently. A VoiceXML standard that supports multiple biometric identification techniques may follow the same general structure and syntax of such a multimodal file format.

Disclaimer

This paper is presented for discussion purposes and does not represent the official position of General Motors Corporation.