amazon alexa

Alexa Auto: Satisfying Drivers' Needs with IVISs

Daniela Barron, Cora Rankin, & Daxton Mitchell for Kerstan Cole

Top Insights:

1. Unimodal visual information presented on the IVIS during driving can be distracting.
2. Bimodal presentation of information allows drivers to select the stimulus they want to pay attention to.
3. Time between notifications determines whether two notifications are perceived as one or two events.

Top Recommendations:

Safety

1. Present information bimodally with visual and auditory stimuli to minimize attentional strain and maintain availability of information.
2. Present information for deaf and hearing impaired audiences bimodally with haptic and visual feedback.
3. Test and implement methods of reducing urgency and temptation of notifications.

Efficiency & Enjoyment

4. Time notifications within 0-150 ms if audio precedes the visual stimuli and 0-215 ms if the visual stimuli precedes the audio. Notifications to be read as separate events should be timed outside these windows.
5. Alexa should understand users will use average max pauses of 1 sec ± 100 ms.

Purpose:

The purpose of this document is to review research regarding the use of multiple sensory modalities in In-Vehicle Information Systems (IVISs) to increase usability and safety of Alexa Auto. The audience for this document is Alexa Automotive UX and other internal teams who build in-vehicle, consumer products.

Background:

IVISs with voice assistants, such as Alexa Auto, are becoming more common in vehicles and consumers expect interactions with IVISs to be safe, enjoyable, and efficient. IVISs should follow the guidelines we have presented here that prioritize the needs of the driver to make sure users are satisfied with their IVIS interactions.

Contact:

Please contact us through email if you have questions or concerns:
Daniela.Baron@ttu.edu Cora.Rankin@ttu.edu Daxton.Mitchell@ttu.edu

Safety: Preventing IVISs from Distracting Drivers

Display Modality and Distractions

Drivers' safety can be impacted by IVISs when they act as a distractor to the driver. The main source of distraction present in IVISs occurs during driver-system interactions that move attention away from the road. Driving is a mostly visual task and using visual displays leads to a competition for attention between driving and interacting with the IVIS. Drivers indeed report higher mental workload when using visual displays compared to auditory displays, which suggests that auditory displays pose less risk to safety and should be prioritized for communicating information to drivers[7]. Alexa Auto already provides simultaneous auditory and visual display of information, which allows the driver to select the stimulus they will pay attention to depending on the driving conditions.

While auditory displays may pose less risk than visual displays, using only auditory product feedback makes it more difficult for users to recall information, unless the information is repeated, or presented in a simultaneous, supplementary sensory feedback mode[1] as mentioned above. By presenting information both through visual and auditory channels together, drivers only need to glance at visual feedback to help with information recall, while the auditory format helps with initial information processing[5]. When visual and auditory stimuli are combined, drivers report more satisfaction and decreased stress with ADAS$_3$[1], because of mitigated driver distraction.

Despite reports of driver satisfaction with combined auditory and visual product feedback, using auditory feedback does not embrace universal design, and presents an obstacle for deaf and hearing impaired users. In order to ensure that deaf and hearing impaired audiences will have the same accessibility to bisensory product feedback, haptic feedback presented alongside a more minimal visual signal (i.e., a flash of a certain color) might be considered. For instance, if the user requests something from Alexa, the system can respond by sending different vibration signals through the steering wheel alongside different visual signals to signify that Alexa understands, does not understand, or needs more information to complete a command. By using minimal visual signal plus haptic signal, deaf and hearing impaired audiences are still able to experience bimodal sensory feedback without having to use all of their visual attention to process Alexa's messages.

Reducing the Temptation of Notifications

A more recent issue involving notifications has developed from IVISs becoming more connected to wireless devices as well. Occasionally, notifications will appear on the

IVIS and since an IVIS is fairly close to the driver, this interface is more easily interacted with and could lead to checking of notifications during unsafe conditions. Notifications on devices that are more easily accessed than a cell phone during the driving task, such as a smart watch, lead to longer brake response times and this is likely true for IVISs as well[3]. A possible way to make notifications less intrusive is to allow the IVIS to replay the last several notifications upon the driver's request or to visually display the notifications until the driver is available to attend to them. This function could reduce the urgency of listening to a notification if the road conditions do not permit the driver to attend to the notification immediately, but this method needs to be tested with users to ensure that it is effective.

Efficiency & Enjoyment: Ways to Reduce Distractions and Miscommunications

Additionally, drivers often utilize IVISs to engage in entertainment and to maintain comfort. Enjoyment encompasses these tasks as well as the perceived usability of the IVIS[7]. Because users have complained about misunderstanding notifications with multiple stimuli, Alexa Auto can improve both the efficiency and enjoyment of IVISs by modifying the timing of the visual and auditory stimuli associated with using the voice assistant to fit within the recommended temporal binding window. The research our recommendations are based on was conducted with English speaking audiences in mind, so the timings, bimodal cues, and safety concerns discussed here may not apply internationally.

The Importance of Timing

Communication issues can arise when timing is misinterpreted either when a driver is listening to their Alexa Auto or when the Alexa is listening to the driver. The amount of time between two stimuli, like two notification sounds, can determine whether or not someone will believe that those two stimuli are paired or separate. This amount of time is often called the temporal binding window (TBW)[10]. Note, this TBW can also apply to other types of stimuli pairs. Audiovisual pairs, like the ones Alexa uses to let a user know the system is listening, need specific timings so that a user can interpret them as paired or separate.

These types of miscommunications can make the driver miss important information and cause frustration with the IVIS. However, they can be avoided if drivers are given cues respecting the TBW. These issues also apply to Alexa when Alexa is listening to a user. If Alexa does not understand that two sounds made by the user are meant to be paired or

Timing misconception types:

- Separate notifications happen too quickly
- Alexa seemingly interrupts notifications
- Paired notifications separated by inappropriately long pauses

separate commands Alexa can misinterpret their voice inputs and cause frustration for the user. For this reason, it is important for Alexa to understand the pause lengths and TBWs of human speech. Giving Alexa an understanding of these timings will allow the system to avoid interrupting the user or pausing too long, making it seem like Alexa did not hear the user. Alexa Auto users have complained about this issue[2]. Alexa "being dumb" when it comes to listening to users can cause Alexa to not hear commands or to never deliver important user requested notifications.

Timing Specifics

For reference, when having an in-person conversation there is a physical difference of 100 ms between the visual and auditory stimuli reaching the eyes and ears[11]. Humans do not notice this difference because our auditory system works faster than our visual system, but the physical 100 ms limitation is still a factor.

Short pauses between paired notifications & long pauses between separate notifications

1. Minimum of 100 ms between paired notifications due to physical difference of light and sound.
2. Maximum of 1 sec ± 100 ms between paired notifications.
3. Maximum of 150 ms between paired, audio-first notifications.
4. Maximum of 215 ms between paired, visual-first notifications.

Pairing two notifications requires that the visual stimulus occurs at most 150 ms after the auditory stimulus. When the visual stimulus precedes the auditory stimulus, 215 ms is the maximum amount of time that can be between the two stimuli [6, 11, 12]. If the goal is for two notifications to seem separate, the time between the two should be greater than the times above and less than 1 sec ± 100 ms[8]. This upper limit is determined by the average maximum pause length in human speech.

Influencing factors:

- Large TBW variability among individuals
- Focusing on the road increases the TBW
- Focusing on Alexa decreases the TBW
- Simple sounds like beeps need smaller TBWs than speech or music

These times are for ideal situations, however certain situations have different effects on the TBW ranges. The research does not currently specify how much the ranges are affected, but it does cover when and how the ranges will change. To begin with, there is a large amount of variability between individuals' personal TBW. This variability means that individual users may have a TBW that is significantly above or below our recommended time ranges. Attention level is also a factor that affects the TBW[14]. If a driver is mainly focused on the road their TBW will increase, but if they are focused on interacting with Alexa it

will decrease. Additionally, the TBW becomes larger as the stimuli become more complex. So, Alexa would need the least amount of time between non-speech notifications like beeps and tones. Two notifications that Alexa will speak will need a smaller TBW than notifications Alexa might present with music, but will need a larger TBW than beeps and tones.

Additional Insights: User Input and Product Feedback

User-Input Preferences for Alexa

Preferences for user stimuli input (voice-activated or a tactile button press) for activation of the Alexa Auto may differ depending on the user's age. Senior citizens often experience a "deterioration in tactile sensory feedback,"[9] which means the voice-activated option is often the most accessible for them.

Further Examination of Auditory and Haptic Product Feedback Options

1. To avoid potential driver annoyance with specific sound notifications, further research should be conducted about the specific sound signals.
    - Alternatively, if drivers switch off the sound, the default option should switch on the haptic feedback option to maintain safe, bimodal product feedback presentation.
2. To promote the driver's easy understanding of the UI of the IVIS, further research should examine intuitive, non-distracting haptic product feedback options for a variety of messages.

Beyond TBW: Spatial Proximity and Stimuli Characteristics

There are additional cues that should be considered for bimodal communication, so that a driver does not become distracted or confused. Spatial correspondence, volume/intensity correspondence, and frequency correspondence all contribute to this goal[9, 13].

1. If notifications are intended to be linked, they should come from the same physical space. Audio should play through the speaker closest to the visual stimulus to pair the two stimuli.
2. Equal volume and/or intensity of two sounds or an audiovisual help cue the user to the paired relationship between the two notifications. Conversely, noticeably different volumes and intensities may cue the user to the separation between two notifications.
3. A difference in the frequency of two audible notifications will make time delays more noticeable, making different frequencies good for separate notifications. However, minimal frequency differences make time delays less noticeable, making similar frequencies good for paired notification sounds.

Glossary

1. Modality: The way sensory information is displayed or received (through sight, sound, touch, smell, or taste).

2. IVIS: In-Vehicle Information System

3. ADAS: Advanced Driver-Assistance Systems

4. Primary driving tasks: maneuvering pedals and brakes

5. Secondary driving tasks: managing safety maintenance, such as turning on windshield wipers, or turning signal

6. Tertiary driving tasks: managing entertainment, comfort, and information functions

7. UI Design: User Interface Design

8. Temporal Binding Window: The amount of time between two different sensory inputs for the inputs to be perceived as one input.

9. Spatial correspondence: stimuli which share a location

10. Volume and intensity correspondence: stimuli sharing similar volumes and/or intensities

11. Frequency correspondence: sounds sharing similar frequencies

Endnotes

1. Cao, Y., Mahr, A., Castronovo, S., Theune, M., Stahl, C., & Müller, C. A. (2010). Local danger warnings for drivers. *Proceedings of the 15th International Conference on Intelligent User Interfaces - IUI 10*, 239–248. doi: 10.1145/1719970.1720004.

2. Echo Auto review https://www.theverge.com/2019/8/6/20754404/amazon-echo-auto-review-alexa-car-price-specs-features

3. Giang, W. C. W., Shanti, I., Chen, H.-Y. W., Zhou, A., & Donmez, B. (2015). Smartwatches vs. smartphones. *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI 15*. doi: 10.1145/2799250.2799282

4. Harvey, C., Stanton, N. A., Pickering, C. A., Mcdonald, M., & Zheng, P. (2011). In-Vehicle Information Systems to Meet the Needs of Drivers. *International Journal of Human-Computer Interaction*, *27*(6), 505–522. doi: 10.1080/10447318.2011.555296

5. Hecht, D., & Reiner, M. (2008). Sensory dominance in combinations of audio, visual and haptic stimuli. *Experimental Brain Research*, *193*(2), 307–314. doi: 10.1007/s00221-008-1626-z.

6. Lee, HweeLing. & Noppeney, Uta. Long-term music training tunes how the brain temporally binds signals from multiple senses. *Proceedings of the National Academy of Sciences of the United States of America*. 108 (51).

7. Llaneras, R., Lerner, N., Dingus, T., & Moyer, J. (2000). Attention Demand of IVIS Auditory Displays: An On-Road Study Under Freeway Environments. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *44*(20). doi: 10.1177/154193120004402003

8. Cohen, M. H., Cohen, M. H., Giangola, J. P., & Balogh, J. (2004). Voice user interface design. Addison-Wesley Professional.

9. Skedung, L., Rawadi, C. E., Arvidsson, M., Farcet, C., Luengo, G. S., Breton, L., & Rutland, M. W. (2018). Mechanisms of tactile sensory deterioration amongst the elderly. *Scientific Reports*, *8*(1). doi: 10.1038/s41598-018-23688-6.

10. Spence, C., & Squire, S. (2003). Multisensory Integration: Maintaining the Perception of Synchrony. *Current Biology*, *13*(13). doi: 10.1016/s0960-9822(03)00445-7.

11. Stevenson, R. A., Zemtsov, R. K., & Wallace, M. T. (2012). Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(6), 1517–1529. doi: 10.1037/a0027339.

12. Stone, J. V., Hunkin, N. M., Porrill, J., Wood, R., Keeler, V., Beanland, M., ... Porter, N. R. (2001). When is now? Perception of simultaneity. *Department of Psychology, and Department of Clinical Neurology, University of Sheffeld, Sheffeld S10 2UR, UK*.

13. Wojtczak, M., Mehta, A. H., & Oxenham, A. J. (2017). Rhythm judgments reveal a frequency asymmetry in the perception and neural coding of sound synchrony. *Proceedings of the National Academy of Sciences*, *114*(5), 1201–1206. doi: 10.1073/pnas.1615669114.

14. Zhou H.-Y., Cheung E.F.C., Chan R.C.K. (2020). Audiovisual temporal integration: Cognitive processing, neural mechanisms, developmental trajectory and potential interventions. *Neuropsychologia*, Volume 140.