White Paper

Immersive Audio HMI to Improve Situational Awareness

Visteon®

Immersive Audio HMI to Improve Situational Awareness

Dr. Alexander van Laack[1], Axel Torschmied[2], Gert-Dieter Tuzar[3]

Leader Advanced HMI Europe, Visteon Innovation & Technology GmbH[1]
Lead Engineer Audio Infotainment, Visteon Innovation & Technology GmbH[2]
Principal Designer HMI, Visteon Electronic Germany GmbH[3]

Summary

Humans' are multimodal beings that perceive their environment with all their senses, which allows them to create situational awareness (van Laack, 2014). For the driver vehicle interaction the idea of multimodality has been neglected in the past and displays became the most common information output in a cockpit. To handle the significant increase of traffic complexity and information availability inside the cockpit a multimodal combination of immersive audio and visual representation can lead to a better user experience and a more intuitive HMI.

1 Introduction

The abbreviation HMI is a buzzword, which in itself can mean different things. While the H stands for "Human" and the M stands for "Machine", the I is either understood as "Interface" or as "Interaction". In fact the meanings of the words interface and interaction focus on very different elements. Interaction aims at how the human being, the operator, interacts with a machine. To do so, the operator will always need an interface. This could suggest that the "I" can only stand for "Interface", but an interface has no value without any kind of interaction.

With this adapted definition HMI can be understood as an ecosystem of human beings and machines in which information is exchanged through different interfaces or modalities which may address in theory any or all of the human senses.

This paper takes a closer look at the auditory channel and how immersive audio can be used to enhance the HMI and therefore benefit the user experience and situational awareness within the vehicle.

2 General

Advanced Driver Assistant Systems (ADAS) support today's drivers in critical traffic situations. To inform the driver when a system is active or to warn him about a particular danger, ADAS HMI uses visual, haptic and also acoustic outputs (Muños-Benavent, 2012) . A vibrating steering wheel for example can indicate that the driver is about to switch lanes without using the turning signal. A forward collision warning is usually emphasized through a visual and an acoustic warning. Acoustic warnings offer the unique benefit that they can be always perceived, without the need to look in a certain direction or touch a specific device. The acoustic channel is always accessible, even when sleeping and can be used to wake a person up. Additionally psychoacoustics allows the human hearing to determine not only the distance but also the direction of a sound with a resolution of up to 1° (Görne, 2008). Although acoustic cues are already used today, they are very primitive and are mostly generated by the instrument cluster speaker or a simple front-rear scheme is used as in parking aid.

To use the full capacity of human hearing, a system has to be developed that is able to use advanced vehicle sensor input data and process it into a directional sound output covering all x,y,z areas around the vehicle and which is able to reproduce virtual sound sources independent of the physical speaker locations in the vehicle.

3 Sound Output

The most commonly used sound systems are in stereo and surround sound. Both technologies generate phantom sound sources which are created by the superposition of the output of discrete audio channels. During production the scene is created by positioning sound sources or applying effects then mixed to the relevant number of discrete audio channels for predefined speaker positions. Post-mixing or modification of the audio scene and the location of its objects on the reproduction side is not intended. Sequentially the correct reproduction of the spatial information depends on the speaker setup matching the format of the production.

The stereo or surround sound signals contain phase and level information for each audio channel which creates a spatial effect of the recorded scene in a predefined location relative to the speakers – this is called "sweet spot". Outside the sweet spot the spatial sound impression is limited or inexistent. (Rumsey, 2001)

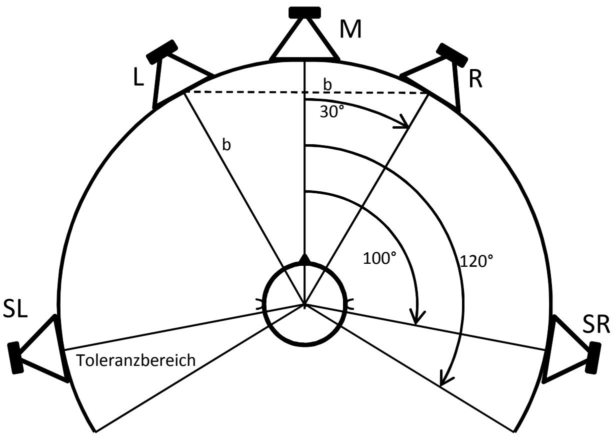

Figure 1: Surround System Setup. Based on (Rumsey, 2001)

Figure 1 illustrates a surround sound setup with 5 channels. The speakers are placed in a circular setup around the listening position. Similarly speakers are placed circular in 7.1 systems or systems with higher number of channels.

For vehicle HMI usage current surround technology has several drawbacks, some of which may partly be overcome by modern DSP processing capabilities:

One, if the listening room does not allow a symmetric deployment of the speakers around the listening position (like in a vehicle) it is required to perform time correction between the channels. Due to the nature of the system it will only reveal its full acoustical capability in a single sweet spot. Since the driver is the most relevant user and his position is fixed it could be argued that the sweet spot is negligible.

Second, the most common surround sound formats derive from home cinema applications thus assume the majority of the acoustical information to come from the front (screen or stage). Center, left and right channels allow relatively high spatial precision, rear channels usually fill the scene with ambient information. The spatial playback performance of the system is not unidirectional so it would not necessarily address the HMI requirements of a vehicle which may intend to highlight events from rear or alongside the car.

Third, creating dynamic acoustical events on a channel based audio formats requires thorough consideration of the speaker setup. The ability to place sounds not only where speakers are but also in 'between' them is not straight forward and will either result in complex matrixes of pre-rendered sounds or in a significant amount of algorithm variants to optimize for multiple speaker setups, vehicle types and events.

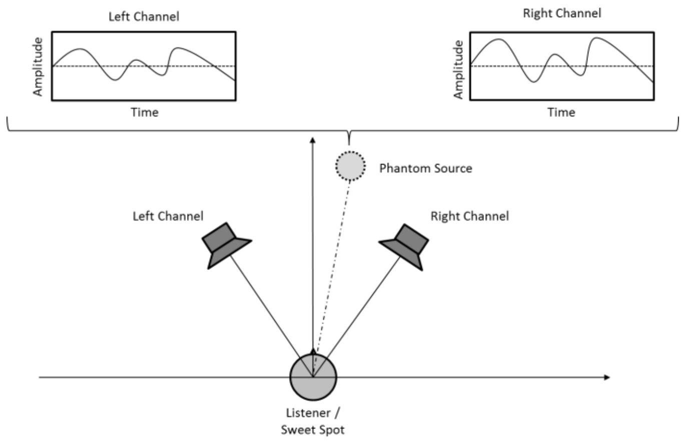

Figure 2: Phantom sound source (based on Sladeczek 2014)

A proposed alternative is the object based audio technology. It assumes transmission of multiple mono sound objects and a separation of their related object meta data. On the reproduction side the meta data is still available to allow user or application controlled modification. A spatial rendering algorithm calculates the channel output at runtime considering the metadata and other rendering information such as information about the speaker locations or other physical or acoustical parameters.

The benefit of this technology for sound HMI is the availability of separated sounds, meta data as well as rendering parameters on reproduction side which allows rendering of virtual sound objects at runtime and serves as the foundation for more complex rendering algorithms. These allow significantly more precise reproduction of virtual sound sources in space while being virtually agnostic to the physical speaker setup. (Engdegard, 2008)

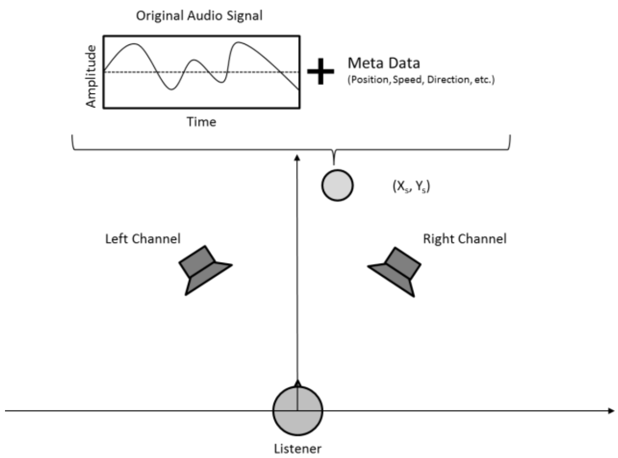

Figure 3: Virtual sound source (based on *Sladeczek 2014)*

4 Customer Clinic

The previous chapter was focused on the system view and has introduced different technologies that have a distinctive relevance for immersive audio HMI. To understand the benefit for the human being this chapter presents customer clinic research that was conducted to answer the following research questions:

To which extend can people differentiate between different sound locations played through an object based sound setup?

Is there difference in perception accuracy between static and dynamic sounds?

Which sound characteristics determine how a sound is perceived?

4.1 Clinic Setup

The customer clinic with 33 participants was conducted at the Visteon facility in Kerpen, Germany. A 10.1 speaker setup was used with an object based sound renderer to create virtual sound objects. Participants were blindfolded to minimize the visual effect of the 10 speakers surrounding them.

The customer clinic was split into three parts. During the first one, people had to determine the correct distance and location of a short static sound. Cardinal points such as north, north-east, were used to express the direction, while distance was captured in close, medium and far. To give participants a distance reference, a sound in all three distances was played back at the beginning of the study.

During the second part of the clinic, participants were asked to determine the movement of a sound. The space around them was divided into four sectors as illustrated in Figure 4. Sounds were either moved through two or three sectors and participants had to state starting sector, ending sector and movement direction.

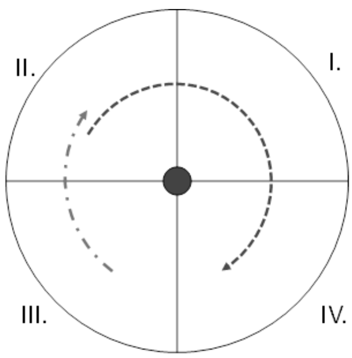

Figure 4: Dynamic sound evaluation

The third part of the clinic was focused on the perception of various sounds. People were asked to classify the sounds as alert, warning or notification and evaluate the sound characteristics using pre-defined semantic differentials.

4.2 Results

The customer clinic results showed almost no significant influence of the distance in determining the correct location of the static sound. However, for static sounds a clear difference between sounds that were played back in front of the participants and sounds that were played back behind the participants is noticeable.

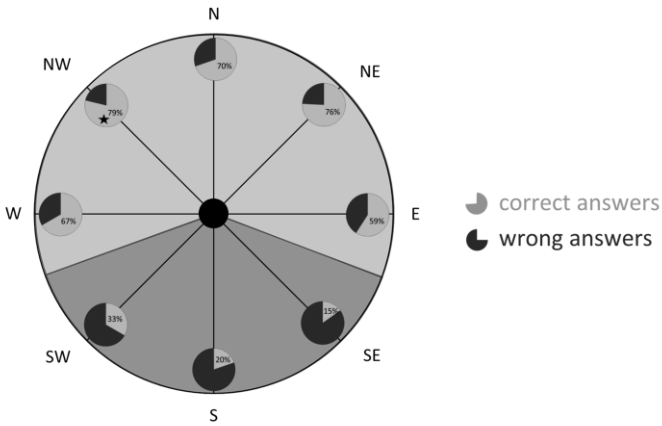

Figure 5: Direction results of static sound evaluation

Figure 5 shows that less than 50% of participants were able to determine the correct location of static sounds, if the sounds were played behind their head.

For the dynamic sound evaluation it was noticeable that participants were able to identify the correct movement direction and movement path easier than the location of a static sound. Especially dynamic sounds moving on the side of the user were identified very well. Difficulties appeared with sounds that were right behind the head for the longest time. Those dynamics are still perceived better than the static ones, but worse than the ones on the side.

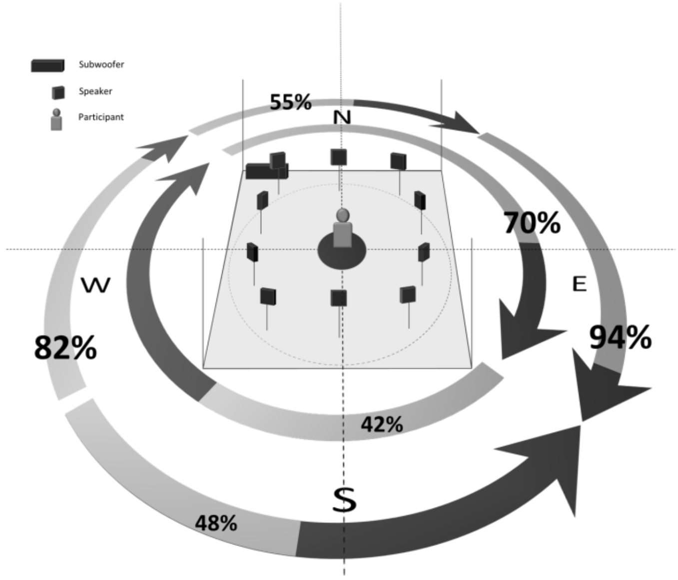

Figure 6: Dynamic sound results

During the third part of the clinic, participants were asked to evaluate different sounds on a semantic scale. They were first asked in which category (notification, warning and alert) a sound falls into and then they were asked rate them using semantic words with antonyms, such as like and dislike. The results in Figure 7 show the average characteristics for notification and warnings & alerts. Sounds that were categorized as notification were also described as playful friendly and pleasant. Notification as well as warnings and alerts were both described as informative. In contrast warnings and alerts are found to be more critical less playful and less friendly.

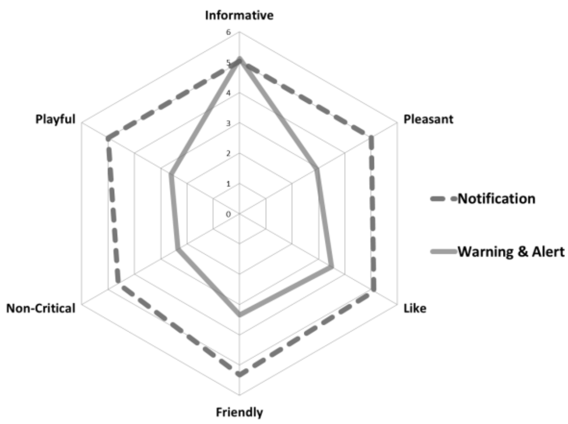

Figure 7: Semantic evaluation of sounds

4.3 Summary

The results of the customer clinic showed that with certain limitations an object based sound system can be used to create spatial sound objects that are perceived as such by human beings. It was identified that dynamic sound objects are easier to be located than static ones.

Furthermore, the research showed how people associate certain situations with specific sounds. Some sounds can clearly be used to warn the driver while others are rather used for notifying him.

5 Outlook

The research presented in this paper suggests that object based audio can enhance the vehicle HMI by offering situational context and directional elements to the driver through a modality, which is not visual. The research findings further indicate that dynamic sound objects will be very well perceived on the sides of the driver, e.g. for fast approaching cars or bicycles, while static sounds should only be used in the field of vision, e.g. to support contact analogue information in a head up display. Object based audio proposes a natural HMI using humans´ available sonic perceptual channel to guide drivers´ attention and to extend ones perception space beyond vehicle boundaries.

6 Literature

Engdegard, J; Resch, B.; Falch, C.; Hellmuth, O.; Hilpert, J.; Hoelzer, A.; Terentiev, L; Breebaart, J.; Koppens, J.; Schuijers, E.; Oomen, W. (2008). *Spatial Audio Object Coding (SAOC) – The Upcoming MPEG Standard on Parametric Object Based Audio Coding*. In: Audio Engineering Society. May 17-20, 2008. Amsterdam, The Netherlands.

Görne, T. (2008). *Tontechnik*, 2. Aufl. Leipzig: Carl Hanser Verlag.

Muñoz-Benavent, P.; Armesto, L.; Girbés, V.; Solanes, J. E.; Dols, J.; Muñoz, A.; Tornero, J. (2012). *Advanced Driving Assistance Systems for an Electric Vehicle*. In: International Journal of Automation and Smart Technology.

Rumsey, F. (2001). *Spatial Audio*. 1: Auflage. Oxford: Focal Press.

Sladeczek, C.; Gatzsche, G. (2014). *A Flexible System Architecture for Collaborative Sound Engineering in Object-Based Audio Environments*. In: Audio Engineering Society, April 2014.

van Laack, A. W. (2014). *Measurement of Sensory and Cultural Influences on Haptic Quality Perception of Vehicle Interiors. 1. Auflage*. Aachen: van Laack Buchverlag.

Kontaktinformationen

Dr. Alexander van Laack, E-Mail: avanlaac@visteon.com
Axel Torschmied, E-Mail: atorschm@visteon.com
Gert-Dieter Tuzar, E-Mail: gtuzar@visteon.com

About Visteon

Visteon is a global company that designs, engineers and manufactures innovative cockpit electronics products and connected car solutions for most of the world's major vehicle manufacturers. Visteon is a leading provider of driver information and controls, audio and infotainment, and domain controllers; its brands include Lightscape®, OpenAir® and SmartCore™. With corporate offices in Van Buren Township, Michigan, (U.S.); Shanghai, China; and Chelmsford, UK; Visteon has 50 facilities in 21 countries. Learn more at www.visteon.com.

Visteon®

Visteon Corporation
One Village Center Dr.
Van Buren Township, MI 48188
1-800-VISTEON
www.visteon.com