Audio-based mobile applications for Android using multisensory feedback in assistive technology

Hunor Nagy, Ádám Csapó, György Wersényi
[1] Széchenyi István University, H-9026, Győr, E-Mail:{nagy.hunor, csapo.adam, wersenyi}@sze.hu

Summary

Mobile devices offer a wide range of applications for assistive technology. Computational capacity, built-in sensors and easy-to-use interfaces allow for improved accessibility to information. Smartphones, tablets and other devices running extensible platforms, such as Android or iOS, opened the possibility for developers and users to interact with an unlimited variety of applications. In the area of assistive technologies, target groups include the visually impaired, as well as users in need of movement rehabilitation. In special cases, audio and/or tactile feedback is provided by the same device via loudspeakers/stereo output, using spatialized audio and vibration motors. This paper briefly introduces some background and three Android-based audio-only applications which were recently developed by the authors and aim to support the visually impaired e.g. to avoid veering, test reaction times to audio stimuli or improve their auditory memory.

Introduction

The term ‘assistive technology’ refers to developments aimed at people with disabilities, generally supporting users who have difficulty in seeing, hearing or performing body motions needed for everyday life. In many cases, primary sensory modalities are supported by replacing their role with information presented through other modalities [1]. Auditory and/or tactile stimuli, for example, are commonly used to provide information substituting the visual, proprioceptive or vestibular modalities.

Today’s mobile devices offer the needed computational capacity and sensors/feedback parts at relatively low costs in a handheld size. Nevertheless, issues concerning latencies, memory usage, real-time rendering, spatial resolution, wireless communication speed, response times, and battery life introduce new problems and solutions. Using existing channels and hardware, spatial information together with sonification and/or haptification open new possibilities for providing (concurrent) information [2].

This paper briefly summarizes the most important aspects of multisensory feedback in mobile devices, focusing on sonification methods, spatial sound and haptics. Following the summary, the paper shortly introduces three recent applications targeting the visually impaired, which were recently developed by the authors for the Android platform. They are accessible via the Google App Store for free.

Methods used in multisensory feedback

Sonification

In the past few decades, the use of sound besides visual feedback has become increasingly useful in a wide variety of computing applications. As researchers aimed to arrive at a principled approach towards auditory interface design, several new concepts emerged, mostly as an analogy to 2D metaphors used in graphical interfaces.

For example, the concept of auditory icons as “caricatures of everyday sounds” was proposed by Gaver in 1986, based on earlier work by Canfield-Smith on (visual) icons in 2D graphical interfaces [3-5]. Later, based on the idea of creating analogies of more serial, primarily character-based communication, Blattner, Sumikawa and Greenberg, among others, developed the theory behind earcons [6] (although initially earcons were equally used to refer to both ‘representational’ and ‘abstract’ messages, today they refer exclusively to abstract content, while the term ‘auditory icon’ is reserved for representational connotations).

In parallel to the development of these two key elements of auditory interface design, complementary theories of sonification began to emerge. If auditory icons and earcons represent analogies with representation types in 2D graphical interfaces, the notion of sonification can be seen as corresponding to that of visualization, and was first defined by Kramer as “the use of nonspeech audio to convey information or perceptualize data” [7]. Later, Hermann and his colleagues extended this definition to encompass all data-driven sound production models that are systematic, reproducible and give rise to perceptual differences with the sole purpose of reflecting objective differences in the underlying data [2].

Since the original formulation of these concepts, several newer kinds of auditory representations have emerged. For an overview of novel representations – including representations with speech-like and/or emotional characteristics (such as spearcons, spindexes, auditory emoticons and spemoticons), as well as representations used specifically for navigation and alerting information (such as musicons and morphocons) – the reader is referred to the overview provided in [8]. As detailed in that paper, much work has also been carried out on creating frameworks allowing for the combined use of various auditory representation types – both in task-oriented and task-agnostic (exploratory) settings. Further, frameworks incorporating analogous representation types in a modality-agnostic way, such as the framework of CogInfoCom channels have also appeared [9].

Spatial Sound

In the case of auditory feedback, another important question is how to achieve spatialization. Everyday mobile devices usually offer low quality loudspeakers. Smaller handheld devices, such as smartphones equipped with one loudspeaker are not able to provide directional information. Larger devices, such as tablets may have two dedicated speakers separated in space to a large enough degree for the creation of spatial effects during playback. Spatial sound can basically be created in two ways. One is simple panning or panoramic stereo between the left and right channels [10, 11]. This can feature classical stereo field in a 60-degree triangle, where left and right speakers appear at 30 degrees left and right. This is similar to how music is played back in classical stereophonic recordings. On the other hand, left-right separation can be made 90 degrees left and right as transducers are placed left and right of the head. It allows for panning in a 180-degree hemisphere without any problems as “left” means only the left speaker is active, while “front” corresponds to 50-50% of the channels. This method does not require a high computational capacity or time consuming post-processing.

Although panning methods basically suffer from well-known playback errors introduced by the headphone (and the lack of head tracking) such as in-the-head localization or front-back errors, subjects can overcome this by “knowing” what they have to hear [12, 13]. As panning methods theoretically work in the frontal hemisphere, subjects usually do not have problems reporting sound source directions even if they have a reversal of perception or a lack of externalization simply by knowing that the sources are in the front. These phenomena however become a serious problem when the space is extended to the rear hemisphere, and sources are also allowed in the back. Panning cannot deliver this kind of information, so the Head-Related Transfer Functions (HRTFs), as a set of digital filters are often used to provide the necessary information [14, 15]. Real-time filtering demands high computational capacity (or pre-recorded sound files), and spatial resolution is typically low, as the use of interpolation and headphone equalization is almost always skipped. It can be assumed – and recently it is being tested – that the required accuracy of spatial simulation can in such cases be delivered through stereo panning by restricting the auditory panning to the horizontal plane and the frontal hemisphere. Additionally, other modalities such as timings, changing loudness or additional filtering can be used to also bias localization judgements also vertically.

Haptics

Haptics, as a form of tactile feedback is commonly used in mobile devices. Vibrations alongside or replacing sound can provide additional information to the user. Although Android allows users to set vibrations parameters, the resulting perceptual channel is more limited than for auditory feedback. Sensitivity to vibrations, thus amplitude level and pattern recall is restricted [16].

Sensitivity to pressure is not uniformly distributed and equal on the body (skin). A high threshold means that the part of the body is less sensitive. Tactile pressure sensitivity is usually highest on the face, the trunk, arms and fingers [17]. What is even more important is to be able to sense changes in pressure over time – i.e. vibration. It is frequency dependent: at higher vibrating frequencies, the absolute threshold decreases. In the case of pressure detection, the usual method to test sensitivity to spatial resolution on the skin is the ”two-point touch threshold”. This is the smallest separation at which one can tell that there were two points of contact, and not just one. Highest acuity is at the fingertips, face and toes of about 5-15 mm, resulting in a better accuracy than auditory (but worse than visual) modality. Still, the use of vibrators is different, as the fingers/palms offer the best body parts for such feedback. Sensitivity to temporal changes means to decide whether two tactile pulses are simultaneous or successive. Subjects can resolve temporal differences of only 5 ms [18]. Here, touch is better than vision (25ms) but worse than audition (0,01 ms) [19].

The use of mobile devices with vibration introduced new phenomena in perception, the phantom vibration (or phantom ringing) as a symptom of overuse of technology [20-22]. Although everyday mobile phone users do not find it bothersome, in cases where important information is provided using vibration, fake alarms can be as problematic as missed ones.

Applications

Within the frame of a current EU funded project, called „Sound of Vision”, a new assistive device is being developed for the visually impaired [23]. This system is not a simple navigation device, but provides information about the environment in a „natural way”. This means 3D stereographic image processing, spatial sounds via bone conduction and/or multi-speaker headphones, haptic feedback using vibrating bracelets and vests and a complete training environment including virtual scenes, serious gaming scenarios and BCI monitoring.

With the aim of assisting this project, some recent developments were introduced to the blind community for pre-testing. All are based on the Android platform.

The „walking straight” app

This application was developed to help blind pedestrians to avoid veering [24-26]. Veering from the straight path is one of the key issues during walking. The application uses the magnetic sensor (compass) of the smartphone to detect the current direction the user faces. If the subject veers from the initialized direction by 5-7 degrees, the device gives haptic and auditory feedback in the form of different vibration patterns and speech. Users can set the accuracy, sensitivity, sounds and vibration patterns etc. Sound is not spatialized, as it would only make sense via headphones which is not optimal for blind users. The application works well both indoors and outdoors, independent of other sensors (e.g. GPS, accelerometer or gyrosensor).

Auretigaa! – Application for testing reaction times

As the walking straight app was developed to assist blind users, the application called „Auretigaa" is a serious game-based solution for testing auditory response times [27, 28]. The goal of this development was to compare results obtained on touch screens of mobile devices with a previous experiment conducted on dedicated hardware, through the use of the same signals and scenarios. In addition, different sound signals such as white noise, pink noise, click-trains, 1 kHz sine and speech samples can be tested and compared. Sound source directions are simulated with simple stereo panning (3 or 5 directions) in the horizontal plane over headphones. Users can set sound types, number of directions, number of repetitions and delays etc. Results are being collected on a server allowing for the joint evaluation of controlled laboratory experiments and crowdsourced data [29]. Users can compete with each other based on scores calculated by number of errors and reaction times.

Auditory Memory Game

This application runs in beta (developer) mode only. It also allows both an experimenter to fulfill laboratory testing and everyday users to enjoy a simple gaming application. The game itself is the auditory equivalent of the well-known visual pairing games, where users have to find pairs by flipping two of a given pair, with cards usually arranged in a matrix. In this case, sound events have to be paired (Fig.1.). Users can set different 2D resolutions and play in single player or multiplayer mode. E.g. by selecting a 5x2 resolution, 5 pairs of sounds (white noise, pink noise, speech, 1 kHz sine, click-train) have to be found. Sound events are allocated randomly in 2 rows and 5 columns. Sounds are spatialized using stereo panning between the two sides, however, the game can be played with and without headphones as well. Number of errors and completion time are recorded, allowing for the further evaluation of results.

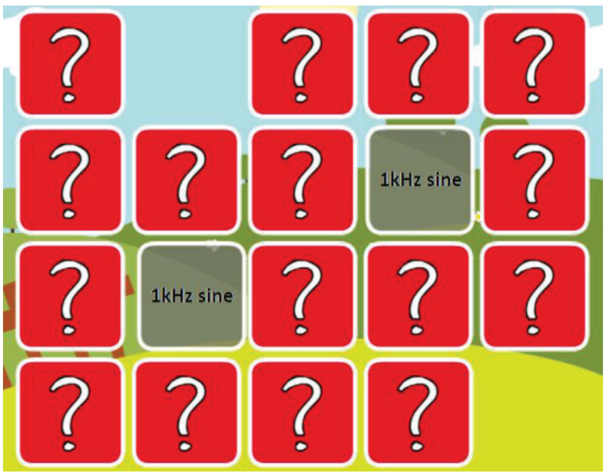

Figure 1: Audio-based memory game for finding pairs based on iconic sound samples such as noise, sine, earcons, auditory icons etc. instead of visual information.

Discussion

Results from testing environments designed for dedicated target groups in assistive and rehabilitation engineering are capable of highlighting new directions for development.

First, based on psychological experiments in tactile and auditory perception, both auditory stimuli and haptic feedback can be used to provide information.

Sound, as the most important substitutional sense after vision can be applied with and without spatial information. Directional information of sound sources can be transmitted only in a limited way in the case of handheld devices. Loudspeakers are small and often too near to provide the necessary time difference between the channels. In some cases (blind users, sports) the possibility to use headphones is also limited, especially if wireless transmission is recommended.

For vibration, a direct contact with the skin is required. If the device is in the pocket or connected to the body through clothing, sensation of vibration can be disturbed. Furthermore, the most sensitive body parts such as the hands, fingers, face and skull can introduce problems, such as not having "free hands" or "looking strange". One of the solutions can be vibrators hidden under the clothing but having a tight contact with the body (bracelets on the arms and ankles, gaming vests). In such cases, multi-channel information can be provided, limited only by the user's ability to discriminate between the vibrators and the wiring. Spatial distribution and channel-separation can be maintained by the advantageous placement of the individual vibrators.

An important question in future work is to test discrimination and memory capabilities of the users in the case of simultaneous signals. Concurrent sound sources along with multi-channel vibrations can lead to sensory overload, confusion, and inconsistent information processing.

Summary

Mobile devices, such as smartphones and tablets incorporate built-in sensors, sufficient computational capacity and multisensory feedback elements to combine auditory and haptic feedback in assistive technology and/or gaming. Spatialized sound using simple stereo panning methods in the horizontal plane and frontal hemisphere is suggested, especially if headphone usage is not restricted. Vibration can be used as an extended information channel due to lower resolution and accuracy. Dedicated hardware and software developments may contribute to special needs, but common "smart" devices using panned stereo rendering and built-in vibrators can be used for everyday applications.

Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 643636 "Sound of Vision".

References

[1] Hersh, M., Johnson, M. (Eds.), Assistive Technology for Visually Impaired and Blind People, Springer, London, pp. 51–58, 289–299, 2008

[2] Hermann, T., Hunt, A., Neuhoff, J.G. The Sonification Handbook. Logos Verlag Berlin, 2011

[3] Gaver, W.: Auditory Icons: Using Sound in Computer Interfaces. Human Computer Interaction 2(2) (1986), 167–177

[4] Gaver, W.: The SonicFinder: An Interface that Uses Auditory Icons. Human Computer Interaction 4(1) (1989), 67–94

[5] Smith, D. Pygmalion: A Computer Program to Model and Stimulate Creative Thought. PhD thesis, Stanford University, Dept. of Computer Science, 1975.

[6] Blattner, M., Sumikawa, D., Greenberg, R.: Earcons and Icons: Their Structure and Common Design Principles. Human Computer Interaction 4(1) (1989), 11–44

[7] Kramer, G. (Ed). Auditory Display: Sonification, Audification, and Auditory Interfaces. Santa Fe Institute Studies in the Sciences of Complexity, Proceedings Volume XVIII. Reading, MA: Addison-Wesley, 1994

[8] Csapó, Á., Wersényi, Gy.: Overview of auditory representations in human-machine interfaces. ACM Computing Surveys 46(2) (2014), 19 pages

[9] Baranyi, P., Csapó, Á., Sallai, G. Cognitive Infocommunications. Springer International Publishing, 2015.

[10] Vazquez-Alvarez, Y., Oakley, I., Brewster, S. A.: Auditory display design for exploration in mobile audio-augmented reality. Personal and Ubiquitous Computing 16(8) (2012), 987-999

[11] Griesinger, D.: Stereo and Surround Panning in Practice. 112th AES Convention Paper Preprint no. 5564, Munich, Germany, 2002.

[12] Begault, D. R., Wenzel, E., Anderson, M.: Direct Comparison of the Impact of Head Tracking Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source. J. Audio Eng. Soc. 49(10) (2001), 904-917

[13] Handel, S. Listening. An Introduction to the Perception of Auditory Events, Cambridge, USA, MIT Press, 1989.

[14] Blauert, J. Spatial Hearing, The MIT Press, MA, USA,1983.

[15] Cheng, C. I., Wakefield, G. H.: Introduction to Head-Related Transfer Functions (HRTFs): Representations of HRTFs in Time, Frequency, and Space. J. Audio Eng. Soc. 49(4) (2001), 231-249

[16] Wolfe, J. M.: Sensation & Perception, 4th edition, Sinauer Ass. Inc., USA, 405-413, 2013

[17] Weinstein, S.: Intensive and extensive aspects of tactile sensitivity a function of body part, sex, and laterality. in D.R. Kenshalo (Ed), The Skin Senses, Springfield Ass., USA, 195-222, 1968

[18] Gescheider, G. A.: Temporal Realations in Cutaneous Stimualtion. in Proc. of Conf. on Cutaneous Communication Systems and Devices, Oxford, 1974

[19] Sherrick, C., Cholewiak, R.: Cutaneous sensitivity. in K. R. Boff (Ed). Handbook of Perception and Human Performance 1 (1986), 12.1-12.58

[20] Deb, A.: Phantom vibration and phantom ringing among mobile phone users: A systematic review of literature. Asia Pac Psychiatry 7(3) (2015), 231-239

[21] Rosenberger, R.: An experiential account of phantom vibration syndrome. Computers in Human Behavior 52, (Nov. 2015)

[22] Kruger, D. J., Djerf, J. M.: High Ringxiety: Attachment Anxiety Predicts Experiences of Phantom Cell Phone Ringing. Cyberpsychology, Behavior, and Social Networking 16(1) (2016), 56-59

[23] URL http://www.soundofvision.net/

[24] Nagy, H., Wersényi, Gy.: Comparative Evaluation of Sighted and Visually Impaired Subjects using a Mobile Application for Reducing Veering during Blindfolded Walking. Acta Technica Jaurinensis 9(2) (2016), 140-157

[25] URL https://play.google.com/store/apps/details?id=com.johnny.straightlinewalk.app

[26] URL https://www.youtube.com/watch?v=eQKNYVUv6Hw&feature=youtu.be

[27] Wersényi, Gy., Nagy, H., Csapó, Á.: Evaluation of Reaction Times to Sound Stimuli on Mobile Devices. in Proc. of ICAD15, Graz, Austria, 5 pages, 2015

[28] URL https://play.google.com/store/apps/details?id=com.sze.crowdaudio

[29] Csapó, Á., Wersényi, Gy., Nagy, H., Stockman, T.: A survey of assistive technologies and applications for blind users on mobile platforms: a review and foundation for research. J. Multimod. User Interfaces 9(4) (2015), 275-286