Multimedia Systems
DOI 10.1007/s00530-011-0240-2

REGULAR PAPER

User-centred process for the definition of free-hand gestures applied to controlling music playback

Andreas Löcken · Tobias Hesselmann · Martin Pielot · Niels Henze · Susanne Boll

Abstract Music is a fundamental part of most cultures. Controlling music playback has commonly been used to demonstrate new interaction techniques and algorithms. In particular, controlling music playback has been used to demonstrate and evaluate gesture recognition algorithms. Previous work, however, used gestures that have been defined based on intuition, the developers' preferences, and the respective algorithm's capabilities. In this paper we propose a refined process for deriving gestures from constant user feedback. Using this process every result and design decision is validated in the subsequent step of the process. Therefore, comprehensive feedback can be collected from each of the conducted user studies. Along the process we develop a set of free-hand gestures for controlling music playback. The situational context is analysed to shape the usage scenario and derive an initial set of necessary functions. In a successive user study the set of functions is validated and proposals for gestures are collected from participants for each function. Two gesture sets containing static and dynamic gestures are derived and analysed in a comparative evaluation. The comparative evaluation shows the suitability of the identified gestures and allows further refinement. Our results indicate that the proposed process, that includes validation of each design decision, improves the final results. By using the process to identify gestures for controlling music playback we not only show that the refined process can successfully be applied, but we also provide a consistent gesture set that can serve as a realistic benchmark for gesture recognition algorithms.

Keywords Music · Gestures · Camera · Gesture recognition · CD · Process · User centred

A. Löcken (✉) · N. Henze · S. Boll
University of Oldenburg, Oldenburg, Germany
e-mail: Andreas.Loecken@uni-oldenburg.de

N. Henze
e-mail: Niels.Henze@uni-oldenburg.de

S. Boll
e-mail: Susanne.Boll@uni-oldenburg.de

T. Hesselmann · M. Pielot
OFFIS, Institute for Information Technology, Oldenburg, Germany
e-mail: Tobias.Hesselmann@offis.de

M. Pielot
e-mail: Martin.Pielot@offis.de

1 Introduction

Gestural interfaces have been actively explored since the work of Bolt [4]. Those interfaces are often used in movies such as Minority Report and show up routinely in popular TV series like CSI: Miami to demonstrate a futuristic interaction between users and computers. With the introduction of the Wiimote [27] gestural interfaces became available to a large audience. Computer science literature often motivates gesture-based interaction with sentences such as "Gestures are a natural form of communication and are easy to learn" [2]. However, even using gestures to communicate with other humans is a learned communication technique. Furthermore, there is nothing natural about using gestures to control a computer per se. Gestures used for communication between humans evolved over centuries. Because of this, it cannot be assumed that gestures can be chosen by simple intuition. In fact, they must be carefully designed to be usable.

Controlling music playback (e.g. play, stop, pause, and next) is often used to demonstrate new interfaces and interaction techniques (e.g. [13, 20]). Using a set of functions to control music playback has also been used to demonstrate and evaluate gesture recognition algorithms (e.g. [11, 17]). In order to derive meaningful conclusions from an evaluation of a gesture recognition algorithm it is, however, necessary to use a gesture set which is not purely based on the designer's intuition, the algorithms capabilities, or chance. Most work in the area of gestural interaction focused on algorithms and robust recognition of gestures (e.g. [7, 8, 19, 22, 27]). However, gestural interfaces must fulfil the same requirements as any other interaction technique. In particular, it is important to define usable gestures for the functionalities that the particular application offers. In order to deduce usable gestures a process that ensures valid results must be employed.

Our main contribution is to refine a user-centred process defined by Nielsen et al. [24] for designing a set of free-hand gestures for a specific system by adding more validation to the steps involved. We implement this process on the example of a gesture-controlled music player. As result of this applied process shown in Fig. 1 we are able to present a validated set of free-hand gestures.

After discussing the related work in Sect. 2, we propose a refined process for deriving free-hand gestures from constant user feedback in Sect. 3 . Along this process, a set of free-hand gestures for controlling music playback is developed. In Sect. 4, the situational context is analysed to shape the usage scenario and derive an initial set of necessary functions. In the successive user study described in Sect. 5 the set of functions is validated. Furthermore, proposals for gestures are collected from the participants for each function. Two gesture sets containing static and dynamic gestures are derived in Sect. 6. In Sect. 7, the gesture sets are analysed in a comparative evaluation. Based on the results the gestures are refined to form a consistent set of gestures for music playback. We close the paper with a conclusion and outlook to future work in Sect. 8.

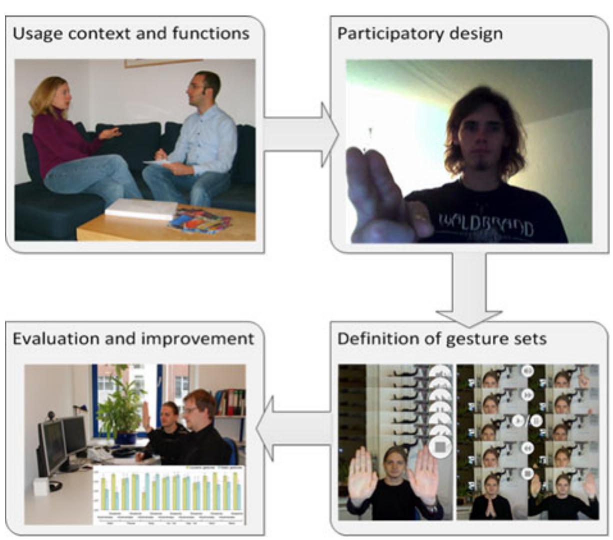

Fig. 1 Steps of the proposed process applied to controlling music playback

2 Related work

This article focuses on the definition of a refined user-centred process for the definition of free-hand gestures, which we validate using a case study to derive gestures for a music playback system. In the following, we discuss related work to these fields.

2.1 Detecting in-air hand gestures

Technically, different methods can be used for detecting and identifying in-air hand gestures. Kela et al. [16] classify these into two categories: movement-sensor-based approaches and camera-based approaches. Movement-sensor-based approaches rely on specific hardware for the recognition of in-air gestures, such as accelerometers, gyroscopes and similar sensor systems, which are typically built into third-party devices, such as smartphones. In contrast, the latter approach entirely relies on a combination of digital cameras and computer vision algorithms to detect gestures without further tools.

2.1.1 Movement-sensor-based approaches

One strand of research focuses on using custom-made hardware devices such as data glove as an input device. Hofmann et al. [15] used the sensor data from data gloves for gesture recognition. They proposed to divide the recognition low-level processing using a vector codebook and a high-level stage that uses Hidden Markov models (HMMs). Another example is the work by Baudel et al. [2] that used a data glove for controlling remote objects.

Several works used existing devices to detect gestures conducted in-air. With the introduction of Nintendo Wii controller Wiimote this device has been widely adopted by researchers. Schlömer et al. [27] for example, use the accelerometers of a Nintendo Wii controller to detect three-dimensional gestures in open space. Similarly, Wu et al. [32] developed an expensive gesture recognition algorithm based on multi-class Support Vector Machines that is evaluated using the Wiimote. Extending the work by Wobbrock et al. [31], Kratz et al. developed a lightweight gesture recognizer for 3D acceleration sensors that can be implemented easily while still achieving a recognition rate of 80% using only little training data.

Another strand of research exploit the sensors integrated in recent smartphone for gesture recognition. Pylvänäinen [25] implemented and evaluated a accelerometer-based gesture recognition using continuous HMMs for mobile devices. Similarly, Choi [5] developed a recognition algorithm for mobile phones and used it for performing music. Mäntyjärvi et al. developed an algorithm for mobile devices and showed that their recognition accuracy is 97% for exemplary gestures to control DVD playback.

2.1.2 *Camera-based approaches*

In contrast, several approaches, including the one presented in this paper, are using digital cameras to track and identify gestures using computer vision algorithms. Gavrila [9] provides an overview about early work visual analysis of human movement while Wu and Huang [33] specifically reviewed work on vision-based gesture recognition.

Recent approaches include the popular Sixth Sense system of Mistry et al. [21], in which a wearable camera is used to detect gestural interactions with everyday objects. Further, many in-air gesture-based applications have been built based on the Microsoft Kinect in the recent time. The Kinect is a low-cost in-air gesture detection device originally bundled with the Microsoft X-Box, which is based on an infrared projector and a camera system [6]. While originally used for controlling games, the Kinect is more and more used for many 'homebrewn' and research projects [10].

2.2 Gesture-based applications for music playback

On the application side, several works can be identified which are controlled using in-air gestures tracked by visual sensors. A product from the industry with similarities to our work is Symbian Moove, a gesture-based music player for Symbian S60 based mobile phonesz [29]. By performing hand gestures in-air, users are able to control a music player application that comes with their smartphone. Moove works by sampling and interpreting pictures from the built-in camera of the mobile device. In contrast to the approach presented in this paper, the system uses a set of four predefined hand gestures: left-to-right and right-to-left motions, covering the camera and tapping the camera. Nevertheless, similar gestures for next track and previous track are used in our approach and Moove. Stenger et al. [28] present a remote control based on a single camera mounted on a public display, which is controlled by hand gestures. The system is aimed for public settings in particular and can be used to control several applications, including a system to browse video collections as well as viewing a gallery of 3D objects. In contrast to the kind of gestures found in our work, Strenger et al. use in-air gestures as replacement for a mouse, meaning that the cursor is moved into the same direction as the hand and a special gesture is equivalent to a mouse click. Bergman et al. [3] present a gesture-based music player based on the orientation of a mobile device. While the application itself is related to our work, it makes use of a mobile device for detecting gestures, while we focus on in-air gestures for controlling music playback.

2.3 User-centred definition of gesture sets

In the area of identifying appropriate gestures for common and special tasks, less research can be seen. Most gesture sets available for different domains have been predefined by system designers (e.g. [7, 8, 19, 22, 27]) out of technical constraints (e.g. to ensure a reliable recognition of the used detection sensors), or they were based on best practices, the experience of HCI researchers, or industry guidelines. In contrast to this approach, user-centred approaches for gesture definition have been proposed by research. Nielsen et al. [24] claim that, despite skilful design, technology-based approaches lead to an awkward gesture vocabulary without intuitive mapping towards functionality and a system which works under strictly pre-defined conditions. Kray et al. [18] asked participants to spontaneously produce gestures with their phone to trigger a set of different activities. Their results suggest that phone gestures have the potential to be easily understood by end users and that certain device configurations and activities may be well suited for gesture control. The findings of Morris et al. [23] indicate that users prefer gestures authored by larger groups of people, such as those created by end-user elicitation methodologies or those proposed by more than one researcher.

Our approach explicitly builds on the work of Nielsen et al. [24] who proposed a procedure, consisting of three user studies, to derive a usable set of free-hand gestures for a given task. They demonstrated the procedure by deriving a gesture set for a simple architectural design application. In our work, we refine the proposed process, including an aspect left for future work by the authors: Nielsen et al. used only a single gesture vocabulary, which was assessed by a follow-up benchmarking step. In our work, we propose the definition of multiple gesture sets, which can then be compared with each other to derive the best possible gesture candidates for each function.

Further work has employed similar processes to derive sets of gestures for 3D selection of neural pathways [1] and basic gestures for surface computing applications using participatory approaches [30]. Hesselmann et al. [14] have extended this approach and proposed a user-centred process for the definition of applications for surface computers, including a method for the definition of gestures for such systems similar to [30].

3 Procedure

As described in the previous section, there has been much work done on how gestures can be recognized. Also there is research on how gestures can be defined in general. In contrast to this work, we aim to show a concrete process on how to define free-hand gestures involving users as much as possible. People who want to control their applications using free-hand gestures can apply the proposed process to derive appropriate sets of in-air gestures for their systems. In this article, we exemplify the proposed approach by applying our process to derive gestures for a music-playing application. Nevertheless, it should be noted that the process is designed to work independent from a concrete domain. Thus, it can be employed for a broad range of gesture controlled applications.

As basis for this process we assume that the procedure defined by Nielsen et al. [24] can be refined by collecting more information from each of the conducted user studies. In particular, we propose to validate the outcome of each user study in the subsequent study. The procedure employed in this paper consists of four steps shown in Fig. 2 and described and justified in the following. The general approach here is to let the users choose functions and gestures as freely as possible and let them concrete the gestures for the functions in each step.

3.1 Usage context and functions

In the first step the of the process, the usage context of the intended interface is analysed and an initial set of functional requirements is defined. Designers and developers usually start with a vague idea that a particular application or user interface can be improved by a gestural interface. Thus, understanding the usage context is important not only to derive adequate functionalities, but also it can be used to validate the developers' initial idea to design a gestural interface for a particular use case. The designers must concretise the usage scenario in order to design the interface for situations where it makes sense to use gestural input. In particular, location and audience had a significant impact on a user's willingness to perform gestures [26]. Furthermore, a set of functional requirements (referred to as 'functions' in the following) should be collected, expressing what functions should be featured by the application. In order to support non tech-savvy participants, demonstrating a simple prototype can help them to imagine the final system. As no concrete usage scenario is defined beforehand, the initial function set is not necessarily conclusive.

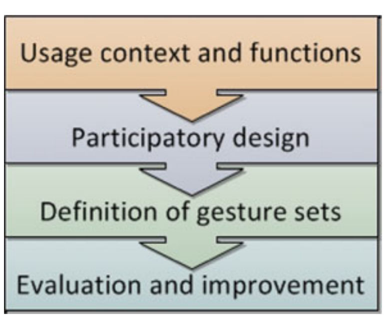

Fig. 2 The four steps of the procedure to identify usable gestures

3.2 Participatory design

In the second step the initial set of functionalities is validated and multiple gesture sets are derived from participatory design techniques. As the lack of important functionalities can render a useless interface, the initial set of functionalities must be validated for the concretised usage scenario. Furthermore, potential gestures for each of the functionalities are collected by conducting a user study and ask the participants to perform gestures for the respective functionalities. These gestures should be recorded on video to analyse the results afterwards. Nielsen et al. [24] highlight that "scenarios take the testees away from technical thinking, especially when conducting the tests on technically minded people". Likewise we would like to highlight that non tech-savvy participants need a concrete idea about the behaviour of the system. They must be enabled to imagine the situation in which they would use the application in order to avoid socially unacceptable gestures [26]. In order to optimize the results of this step the group of participants should be distinct from the group that took part in the first step. As a result, a conclusive set of functionalities are created. Furthermore, there exist a set of ideas of what gestures could be used for these functionalities.

3.3 Definition of gesture sets

The next challenge is to develop usable gestures. Therefore, the proposals for gestures are formalized to define a consistent set of gestures. In this context, consistent means that a user conceives a gesture within a set of gestures to fit to the other gestures within the set. In contrast to Nielsen et al. [24] we propose to derive multiple gesture sets. By not limiting the outcome to a single set of gestures the risk to reject promising candidates is reduced. Nonetheless, every gesture must be part of a consistent gesture set to ensure that a gesture can be combined with other gestures in a reasonable way. After this step there are some sets defined that can be tested against each other.

3.4 Evaluation and improvement

After determining concrete sets of gestures, it is necessary to evaluate and eventually refine the found gesture sets to confirm their quality. Nielsen et al. conduct a final user study to benchmark the derived gesture set. However, conducting a comparative evaluation of multiple gesture sets enables to compare the respective sets (similar to [26]) as well as the individual gestures. Thereby the designers are not only able to select the gesture set with the higher performance but also to mix gestures from multiple sets if necessary. Furthermore, the evaluation can be used to refine the gestures by collecting qualitative advices from participants. As qualitative feedback it should at least be asked whether all expected functionalities were covered and if there was any gesture that should be used instead of the proposed ones. Again, a group of participants distinct from the groups of the other steps should take part in this evaluation. This way the quality of the results is improved.

4 Usage context and functions

Following the process described above the first step is to investigate the usage context of the intended gesture-based application and define the initial set of functions. As our case study involves designing a gesture-controlled music player, the goal of this study was to understand the situational context, in which people listen to music. We aimed at specifying in which situations a gestural interface to control music playback should be applied. In addition, we tried to identify the most important functions such an application has to provide.

4.1 Participants

Nine participants from different educational and social backgrounds were interviewed. Their age was between 23 and 32 with a mean of 27.5 (SD 3.4). Seven of them were males, two females. The individual sessions took about 20–30 min each and were conducted in informal settings such as the participants' living rooms. An exemplary setting is shown in Fig. 3.

4.2 Apparatus

We used a simple prototype to provide a hands-on experience to the participants. The hardware setup of the prototype consists of a notebook with an integrated webcam and a collection of CDs. On the notebook runs an application that constantly analyses the image stream recorded by a webcam. Using the algorithm described in [12] the prototype recognizes selected album covers. On recognising an album cover the application would cue the CDs content into the playlist of iTunes. This, of course, only worked when the required content of the CD was available in a digital audio format on the computer. The popular music player iTunes by Apple was used to do the actual playback of the song. Besides starting songs the prototype did not offer any further interaction techniques. As shown in Fig. 4 users can hold an album cover, e.g. a CD's jewel case, in front of the webcam.

Fig. 3 Interviewer and participant during the interview in the participant's apartment

4.3 Procedure

The studies were conducted by different interviewers and one participant at a time. During the interview the interviewer took notes using pen and paper and wrote a report about the interview afterwards.

Each session was split into two parts. In the first part, we assessed the reasons and the purpose participants listen to music in different situations through semi-structured

Fig. 4 A CD in front of the computer's webcam is recognized and the associated music is started

interviews. In the second part, we presented them a simple vertical prototype that recognizes CDs using a webcam and plays its musical content. The goal was to get initial feedback about desired functionality and behaviour of the system. So as not to limit the participants' creativity we did not reveal our intention to design a gestural interface to the participants.

We opened the interview by asking the participants in what situations listening to music plays an important role for the interviewed person. From each participant, we collected a number of situations, and then decided for one, which was discussed in-depth in the following. Our aim here was not only to cover a wide range of different situations but also go into sufficient detail to understand the situational context. In the following part of the interview, we asked the participants to reflect about what purpose or role music plays in the respective situation. Finally, we investigated how the participants choose and select the music in the given situation. We asked what steps the users usually perform while listening to music, what types of music players were used, and how satisfied the participants were with these solutions. Here we aimed at understanding which functionality of music players would be most important in this situation.

In the second part of the study we presented the prototype described in Sect. 4.2 to the participants. The participants were then encouraged to test the prototype and select a few songs. Afterwards, we asked them to name the functionality the prototype should provide to be most useful to them. We also asked the participants about their opinion (good and bad aspects) of such as gesture-controlled music player.

4.4 Results

In the following we present and discuss the results we obtained from the interviews.

4.4.1 Music listening habits

4.4.1.1 Importance of music Analysing the different situations in which people listen to music we learned that there is a strong difference in the importance the music plays in the situation. We distinguished between situations were music is the key aspect, music play a major role, and music is secondary. However, the borders between the three classes are blurred and the meaning of the music in each situation can vary constantly, depending on the current context.

4.4.1.2 Listening situations Music is considered the key aspect when the participants reported to listen consciously to it while doing nothing else as a primary task. These situations typically take place at home, e.g. while relaxing, and usually occur about once a week and if the situation occurs it does not last very long. The second class of situations are those, where music is an integral part of the situation, such as at parties, at work, or in some cases while doing sports. These situations typically occur not daily but last for a longer time than the previous class of situations. The third class of situations are those, where music is listened by the way. The participants are typically engaged in another primary tasks. These include house work, surfing the internet, playing video games, and car driving. In general, this kind of situation occurs often and last long.

4.4.1.3 The role of music The role of the music strongly depends on the respective situation. The participants' answers were mostly related to emotions. The most important effect of music, which was expressed by eight participants, is keeping, changing, or amplifying emotions. For example, in cases of parties (5×), music plays an important role in supporting good vibrations. When participants listen to music while relaxing, the music should calm them down (4×). Another role that was named three times is helping the participants in concentration while they were busy with another task, such as working or playing a video game.

4.4.1.4 Used music players Participants reported the use of a wide range of music players that fall into the classes of computer media player, portable mp3-player, CD-players, and radio. Six participants use pre-installed or old versions of computer media players. Reasons for choosing certain music players are often associated with convenience aspects, such as the desire for a simple user interface, familiarity, or immediately available music. Three participants used more complex tools, such as iTunes.

4.4.1.5 Music selection patterns We identified certain patterns for using music players. One pattern is that the participants do not really care about what music would be played. four participants said that at least sometimes they just select a radio station or a readily available playlist in order to listen to any music, immediately. Seven participants (also) often pick a specific playlist, radio channel, or CD. Similarly often (7×), the participants generate playlists for certain cases, e.g. creating a playlist for chilling or a party. In these cases, the creating of the playlist is a central, artistic, and sometimes intimate process.

4.4.2 Comments on the prototype

Participants had the opportunities to express their desire for functionalities that they would like to see in a future version of the prototype. In addition, we asked them to imagine potential advantages and disadvantages of a system similar to the presented prototype.

4.4.2.1 Desired functionality Unsurprisingly, the participants demanded the usual basic functionalities that are necessary to control music playback shown in Fig. 5: play, stop, pause, selecting the next or previous track, and change the volume. This function set is consistent with the function set reported by Kranz et al. [17] and Henze et al. [13].

Three participants desired to be able to speak the name of artist or song/album title to start or cue a track. Two participants wanted to be able to lay the CD on a table in order to support situational or congenital physical impairments. Two participants suggested extending the functionality to play other types of media, such as videos. Searching for related music was requested twice. Several other types of functionality were named by a single participants, such as karaoke, displaying the title of the current song, using a mobile head-mounted web cam instead of a stationary web cam, show meta data about the current song, store the current playlist, toggle shuffle mode, alter the playlist, and having a timer.

4.4.2.2 Advantages and disadvantages Beside the required functionality, we asked the participants to name the advantages and disadvantages they thought the prototype has compared with other solutions. The participants appreciated the very simple and visual interaction. They also appreciated the aesthetic aspect of holding the music in their hands and expected it to be good for exploring their friends' music collections. They guess that they would more likely listen to music they would not have listened to.

Most participants, however, had concerns that storing, managing, and handling the CDs would require lots of effort. One participant asked if she had to obtain the CDs by herself. Another concern was that it would be cumbersome for people who mostly listened to single songs that had no relation to full albums.

4.5 Discussion

Systems that demand repeatedly physical interaction, such as handling a number of physical artefacts are not adequate for situations where music plays only a secondary role. In these situations participants pay only very little attention to the music and to controlling music playback. The played music must only roughly fit with the listener's taste and interaction with the playback device is often limited to turning the playback device on and off. When music plays a more important role users are willing to put more effort in selecting and the music. We assume that a system similar to the used prototype could serve the users' needs if the current handling pattern involves more complex but still simple tasks, such as selecting an album. A particularly promising application scenario for the intended system is, thus, a private party.

We observed that music listening is highly emotional and experiencing music not only with the ears but also by physically exploring and controlling the music might further support the emotional experience. If current handling patterns involve very intensive adjustment of the music playback, we assume that users need a very detailed control of the music. Limiting the functionality in any way seems to be not acceptable for the respective participants and a system using physical representatives for large collections of unstructured music no adequate solution.

The participants had the impression that the prototype is very easy to use. Nonetheless, some participants demanded complex and sophisticated functionalities. We observed a discrepancy between what the participants typically do with their music players and which functionalities they demand. If focusing on specific music listening situations, the prototype's ease of use can be retained by restricting the functionalities to only ones that are necessary in this context. It seems clear that users must be able to start, pause, and stop the music. In addition, selecting individual tracks from an album was a highly demanded feature that should be supported.

The participants' most prominent concern was the handling and management of the CDs. Since the use of CDs to select albums is not necessarily a core aspect of the interaction design this can potentially be avoided. For selecting an album it was proposed to use other artefacts that consume less space, such as the cover sheets only or plastic cards. Such cards would be easier to store and manage. For the further interaction, it was proposed to forgo the use of artefacts, as they were not needed anymore. Instead, informants suggested the use of free-hand gestures.

Fig. 5 The initial set of function that consists of play, stop, pause, next, previous, decrease volume, and increase volume

5 Participatory design

In the previous section play, stop, pause, selecting the next or previous track, and change the volume were found to be the basic functions when controlling music playback. For more complex functionality such as choosing a specific track or album, it was proposed to use a physical artefact like a CD's jewel case. The next step of the proposed process was to validate this initial set of functionalities and to derive multiple gesture sets from participatory design techniques. Accordingly, the aim of the second study was to derive free-hand gestures for the functions to control music playback found in the first study. A participatory design approach was used to collect gestures from potential users. The intended use case was to control music playback at private parties.

5.1 Participants

In order to collect gestures for the intended situational context the study was conducted in the context of a private party but with one participant at a time. Ten people from different educational and social backgrounds participated in this study. No participant has taken part in the first study. Three participants were females and seven males. Their age ranged from 17 to 25 years with a mean of 20.8 (SD 2.5). The study took about 10–20 min.

5.2 Apparatus

To provide feedback to the participants a Wizard of Oz approach was used. Participants performed the gestures in front of a monitor equipped with a webcam (see Fig. 6). A jewel case and a list of functions, that should be tried, were given to the respective person. The participants were asked to orally announce which function they wanted to perform and to use the jewel case whenever they found it appropriate. The systems functionality was simulated by the investigator according to the respective function announced by the participant.

Fig. 6 Exemplary camera image from the participatory design study

5.3 Procedure

All participants were interviewed by the same person. The study was split into two parts. During the first part the interviewer filled out the questionnaire and took additional notes, if needed. In the second part the participants were filmed by a webcam.

The goal of this part was to validate the set of functions to control music playback. In addition, we collected further information about the music listening habits of the participants. To collect this information, semi-structured interviews based on a questionnaire were used.

In the second part we aimed at collecting potential gestures. Participants were asked to perform one gesture for each of the six functions. Thereby they invented gestures to control the music player. By using this participatory design approach we tried to ensure that intuitive gestures are derived.

5.4 Results

One person did not want to be filmed. Thus, only nine persons participated in the second part of the study. No significant differences were observed between groups formed by age or gender.

The results for the set of functionalities are consistent with the first study and previous work [13, 17].

When inventing gestures for the set of functions, the jewel case was only used to switch to another album. For the other functions all participants performed free-hand gestures. For the analysis we differentiate between static gestures, dynamic gestures, and hybrid gestures. The three types are described in the following.

5.4.1 Used classification of gestures

Dynamic gestures are defined by the movement of the user's hands. The position of the hands and fingers is unimportant. An example is shown in Fig. 7

Static gestures are the opposite of dynamic gestures. The posture of the hands and fingers is important, while the

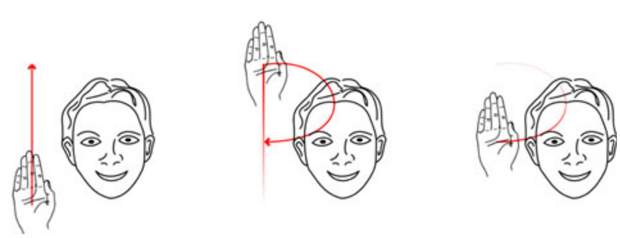

Fig. 7 Example of a dynamic gesture: the hand follows the shape of the character "P"

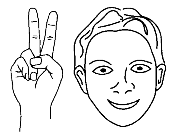

Fig. 8 Example of a static gesture: two fingers form the shape of a "V"

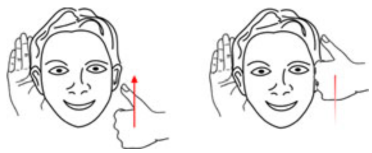

Fig. 9 Example of a hybrid gesture: one hand is held behind the ear while the other hand is moved upwards

movement can be neglected. An example for this kind of gestures is shown in Fig. 8

For hybrid gestures, both, the position and the movement of the user's hands, are important to recognize the gesture. The example in Fig. 9 shows a right hand that performs a dynamic gesture, while the left hand performs a static gesture.

5.4.2 Play

One person covered the webcam with his hand to start the music. Two participants used static gestures. The first showed the Victory sign as seen in Fig. 10, the second built a triangle with his fingers, similar to the play symbol on a music player. Two persons formed one pistol or two pistols with one or both hands. These pistols first pointed to the ceiling and where then moved until they pointed at the webcam. Both variants are hybrid gestures. Another person who chose a hybrid gesture rotated his hands parallel 90° to the right. A fourth participant showed his fist and opened it to start the music. Dynamic gestures were performed twice. One of this gestures was again a pistol moved towards the webcam, while the other gesture was just moving both hands towards the webcam.

5.4.3 Pause

The same person that wanted to start the music by covering the webcam with his hand performs the same gesture to pause the playback. Three different static gestures were performed. The first participant formed the Time-Out symbol used in some sport games such as basketball as seen in Fig. 11. Another person formed a pause symbol by showing both hands with their sides to the webcam. The third person formed a very simple gesture by showing his fist to the webcam. Of the remaining five persons four performed a dynamic gesture by moving one or both hands towards the webcam. The fifth person moved both hands backwards.

5.4.4 Stop

Again the webcam was covered with a hand, this time to stop the playback. Static gestures were used twice. The first person formed a rectangle which should be similar to the stop symbol of a music player as seen in Fig. 12. The second person showed his fist to the webcam. Two gestures can be seen as hybrid gestures: one person hit his left hand with his right fist above his head. Another person acted like he was threatening somebody to cut his throat. The remaining four participants performed three different dynamic gestures to stop the music. The first person moved his hand from the right to the middle. The next person moved his hands towards the webcam. The last two people moved both hands like a *V*. Therefore, they started with both hands in the middle of the body and then the right hand to the upper right and the left hand to the upper left.

5.4.5 Changing the volume

Using a static gesture, one person pointed with his index finger to the ceiling to increase the volume and pointed to

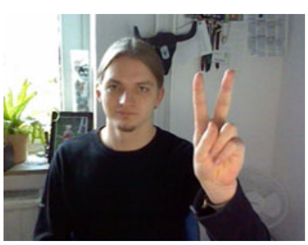

Fig. 10 *Victory* as gesture for start

Fig. 11 *Time-Out* as gesture for pause

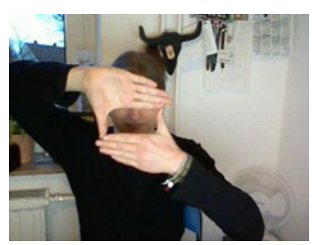

Fig. 12 *Rectangle* as gesture for stop

the floor to decrease the volume. Three people performed hybrid gestures. The first participant turned up or down a fictive volume control to change the volume. Two other participants pointed with one thumb to the ceiling or the floor and shook the hand. The other five gestures were dynamic. One person moved his hands as if he was stretching something to turn up the volume and moved converse to decrease the volume. Another person seemed as if he was lifting or pushing down something with both hands to increase or decrease the volume. The last three participants moved their index fingers up to increase the volume. Two of them moved the hand down with the index finger pointing upwards to decrease the volume, while one person pointed down.

5.4.6 *Next and previous*

A static gesture was performed by pointing to the left or to the right with both hands to move to the next or previous track. The person who rotated his hands to start the playback this time uses only one hand for the same gesture to select the next track. A rotation by 90° to the left is used to select the previous track. The other seven participants performed dynamic gestures. The first person described clockwise circles with his hand to get the next track and counter-clockwise circles to get the previous one. Another person threw a fictive object over his shoulder left or right shoulder to go to the next or previous track. The remaining five people used some movement to the right or left to get the next or previous track. One of them started from the left or right and moved to the middle, while the others started in the centre. Three, including the one who moved from outer to centre, moved their hand with the side towards the webcam, one with his palm, and one pointed with his thumb into the direction he moved to.

5.5 Discussion

In the first part of the study we successfully validated the function set defined in Sect. 4. We assume that the reason why this set cannot be improved based on the results is that the participants of the first study already had a good understanding of the use case. Furthermore, this function set is also described in previous work and we therefore assume that it is the most basic function set needed to control music playback.

The fact that the jewel case was never used for another task but switching the album let us assume that the use of jewel case and maybe other objects to control the music playback might not be the most intuitive way to quickly execute a basic function. To perform an action many different gestures were used to do the same task, which was perhaps due to the fact that the participants were absolutely free to use any gesture. Because of this widely spread set of gestures no complete set of most used gestures can be found. On the other side some coincidences can be seen: Most dynamic gestures are kept very simple as movement along one axis. Also most static gestures are chosen as known symbols.

As described in the second step of our approach we have successfully confirmed the basic functionalities and found many gestures related to this functions. In Sect. 6 we will analyse the found gestures and deduce a set of gestures.

6 Definition of gesture sets

After having found gestures for all basic functions the third step of the process is to formalize the proposed gestures and to define consistent sets of gestures. We found manifold gestures in the user study. Based on these gestures we define two consistent sets. To define them the first set consists of dynamic gestures only while the second set only consists of static gestures. Most gestures were taken from the gestures proposed by the participants in the previous user study. Since we aim at defining consistent gesture sets some gestures were chosen, because they fit to the other ones within the set although the exact gesture was not proposed. In the following both sets are described.

6.1 Set of static gestures

Static gestures were not used as often as dynamic gestures. To form a consistent set, most gestures are similar to the symbols for the corresponding functions found on a music player.

6.1.1 *Play*

As shown in Fig. 13 a gesture was chosen that is similar to the symbol for play in various music players. There was no most used static gesture for this function, so this one was chosen, because it is one of the used gestures, easy to remember, and consistent with the gesture for pause.

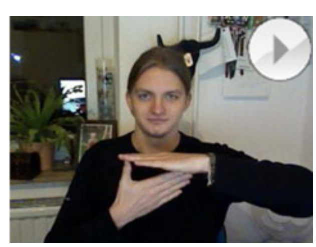

Fig. 13 *Play* of the static set

6.1.2 *Pause*

In Fig. 14 the static gesture for pause is shown. As for the start function there is no most used gesture for this function. This gesture was chosen out of the used gestures in the explorative study, because it is easy to remember, simple to perform, and consistent to the play gesture. Also this gesture is like a weakened version of the gesture for stop. An alternative gesture that is probably easy to remember and perform is the Time-Out symbol described earlier.

6.1.3 *Stop*

The static gesture for the stop function can be seen in Fig. 15. This gesture is simple and known in other contexts. For example it is common to show both palms of the hands to somebody if you want him to stop. A consistent alternative to this gesture would be to create a rectangle with the hands as described earlier and shown in Fig. 12. The rectangle is not proposed as a gesture, because it might not be as easy to perform.

6.1.4 *Changing the volume*

Most participants pointed up to increase the volume, not only when performing a clearly static gesture. For consistency the gesture for decreasing the volume is pointing down, as seen in Fig. 16. Alternative to the index finger pointing up or down the thumb could be used.

6.1.5 *Next and previous*

The only observed static gestures for these functions are shown in Fig. 17. Again this is a gesture that is similar to the corresponding symbol on a music player. The two fingers showing in a direction are representing the two arrows of the next or previous symbol on a music player. As a result this gesture should be easy to remember. Alternatively, other fingers or only one hand could be used to perform this gesture.

6.2 Set of dynamic gestures

Dynamic gestures were chosen more often in the user study. Most times the proposed gestures have one dominant direction. Although it is defined, that a dynamic function does not need a specific form how to hold the hand, one is given to ensure the comparability later in the evaluation.

6.2.1 *Play and pause*

In Fig. 18 shows how to start or pause the playback. Many participants did the same, or a similar gesture to start and to pause the playback, so these function are triggered with the same gesture. A movement towards the webcam was the most often seen gesture. We decided to use the most used form pistol.

6.2.2 *Stop*

In Fig. 19 the gesture for the stop function can be seen. Most people used both hands for their stop gesture. The

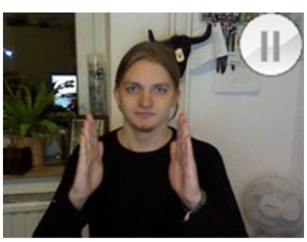

Fig. 14 *Pause* of the static set

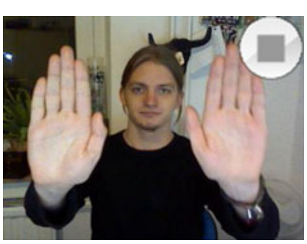

Fig. 15 *Stop* of the static set

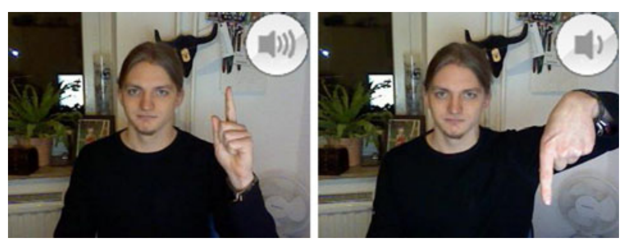

Fig. 16 Increase volume and decrease volume of the static set

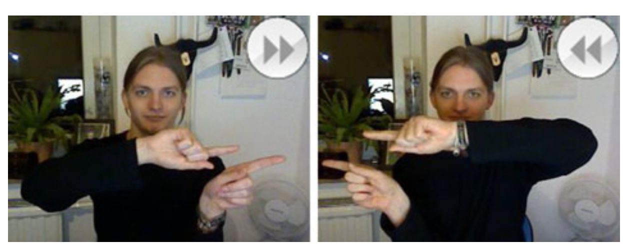

Fig. 17 *Next Title* and *Previous Title* of the static set

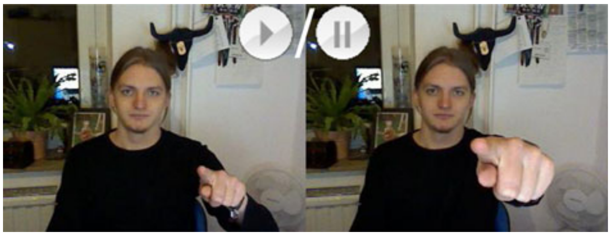

Fig. 18 *Play/Pause* of the dynamic set

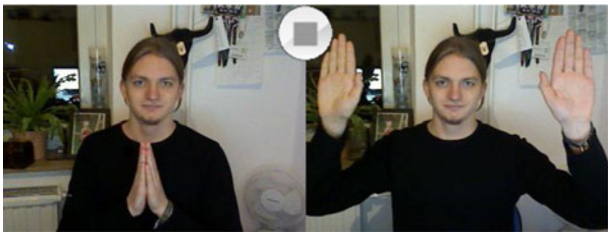

Fig. 19 *Stop* of the dynamic set

most used gesture is the *V* gesture, where both hands are moved from the centre middle to the outer left and right. A good alternative could be the static gesture combined with a movement towards the webcam.

6.2.3 Changing the volume

Most participants moved one or both hands upwards to increase the volume, respectively down to decrease. Additionally, most people pointed up to increase or down to decrease the volume. The gesture which combines both, as presented in Fig. 20, should therefore be very intuitive.

6.2.4 Next and previous

Almost all participants associated next with right and previous with left. Most of them moved their hand to the left or right. So this seems to be a strong connotation and should be used as dynamic gesture. Figure 21 shows that in this gesture the side of the hand is placed towards the webcam as this is the most used version of this gesture. The gesture starts at the centre of the body. Possibly it could be as good to start the gesture from the side to the centre, or hold the hand in a different way while performing this gesture.

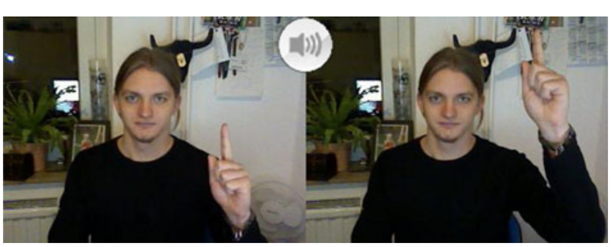

Fig. 20 *Increase Volume* of the dynamic set

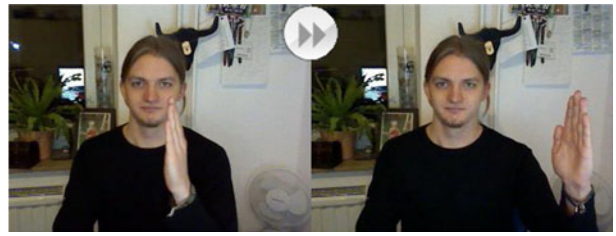

Fig. 21 *Next Title* of the dynamic set

7 Evaluation and improvement

In the final step, the defined gesture sets of Sect. 6 should be evaluated and eventually refined before implementing them. This way it can be tested if one of them is more suitable than the other one. Furthermore, it is determined if there are functions besides the basic functions, that are essential for controlling a music player.

7.1 Participants

Twelve people with different backgrounds participated in this study. Five of them were females and seven males. Their age ranged from 17 to 25 years with a mean of 23.8 (SD 3.8). The study took about 30 min. Even though most of the people of the explorative user study announced that they would like to participate in this evaluation, a divergent set of people was chosen for the evaluation.

7.2 Apparatus

The evaluation is split into three parts. The first part was a set of functions that could be rated using a Likert scale from 1 (unnecessary) to 7 (absolutely essential).

For the second part a Wizard of Oz prototype was built to simulate the functionality of both sets of gestures. This prototype was very similar to the one used in the explorative user study in Sect. 5 (see Fig. 22). The given set of functions differs slightly from the ones used at that study as the participants are asked to use some functions more than once. Another difference is that during the evaluation an extra person was the wizard so that the interviewer could concentrate on the participants. Another rather obvious difference is that the participants this time had to perform given gestures. According to principle of a Within-Subject Design, each participant evaluated both sets of gestures. To increase independency the order in which the gesture sets were evaluated was switched after every trial. The participants were asked to perform each of the six defined

Fig. 22 Setup for the evaluation of the two gesture sets

gestures with both gesture sets. During the evaluation every person was recorded on video. After each trial the participants were asked to rate the gestures' simplicity and intuitiveness. In addition, the consistency, rememberability, and delightfulness of the whole set was rated on a 7-point Likert scale (from 1 = not at all to 7 = perfect).

The third part was an interview on missing functions, where the interviewer took notes. If people had an idea for gestures that could be used to trigger a missing function the according gesture was recorded on video.

7.3 Procedure

In the first of the three parts, the participants rated a set of given functions and had the opportunity to give new functions and rate them. In the second part both gesture sets were evaluated by each user. The participants were asked for missing functionality in the third part. After that the participants had the chance to give additional comments.

7.4 Results

During the evaluation no significant differences were observed between groups formed by age, or gender. Also our assumptions about the basic functions were confirmed in the first part.

Figure 23 shows the ratings for the two gesture sets given in the second part. Simple and Intuitive refer to the means and medians of all functions of a gesture set. In the following the results are described in more detail.

7.4.1 Dynamic gestures

Most dynamic gestures were rated with a mean of 4 or above, as seen in Fig. 24. Only the intuitiveness of the gesture for stop and the simplicity of decreasing the volume were rated less. Figure 23 shows that on average the intuitiveness and simplicity of all functions were rated high. This set was also rated as consistent. How easy it is to remember gestures of this set was not rated as good, but still with a mean of 3.8 not bad. The joy of use was rated rather high again.

Four times it was announced that a good alternative for the stop function was to use a function similar to that one in the static set. Another participant mentioned that he disliked the gesture for stop. As an alternative to moving up or down to control the volume it was also suggested to spin a fictive volume regulator.

7.4.2 Static gestures

Figure 23 also shows that static gestures were rated less than dynamic gestures on average. Apart from start, all gestures were rated with a mean above 3.5 as seen in Fig. 25. The intuitivity of start was not too low, but the gesture was rated as the most difficult. A very high-rated gesture was stop. The gesture is the most intuitive of both gesture sets and the most simple one of the static gesture set.

It was announced that the gesture for start was bad. Also few participants said that using one hand instead of both would have been as good, or even better. Five participants had problems in performing the gesture for start. Three test persons used their index fingers, instead of the side of their hands, to show the gesture for pause.

7.4.3 Dynamic versus static gestures

Figure 26 shows the means for the static and dynamic gesture sets. When comparing the dynamic and static gesture set, only the simplicity for start and the intuitiveness for stop differ significantly. No participant rated higher for the static gesture for start than for the dynamic one. Furthermore, all but three participants voted with a 5 for the dynamic gesture, while the ratings for the static gesture varied. For the stop gesture we observed the opposite. For this gesture the static one is significantly more intuitive. Also the simplicity differs significantly. The simplicities of the gestures for next and back have identical values.

7.4.4 Missing functions

In the last part the participant could give their opinion about functions that were missing so far. Two persons did not miss additional functions at all. The other participants provided many suggestions. Most of the participants would like to have a random function that can be toggled on or off. As gesture it was proposed to shake both hands as when dicing. Another frequently mentioned function was

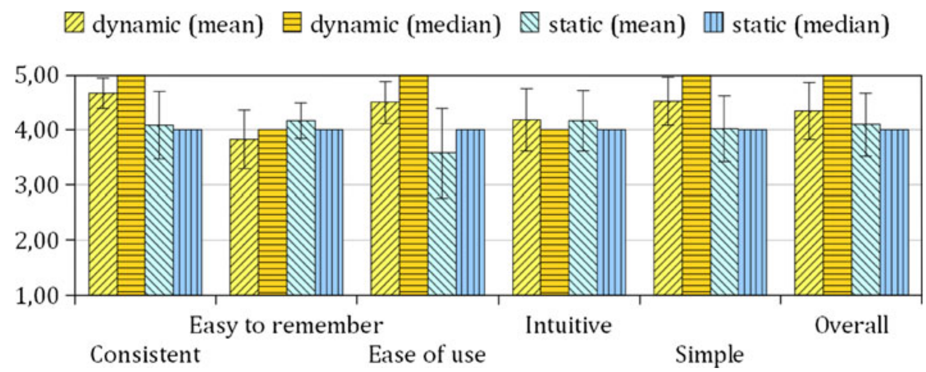

Fig. 23 Concluded results of ratings for the gesture sets

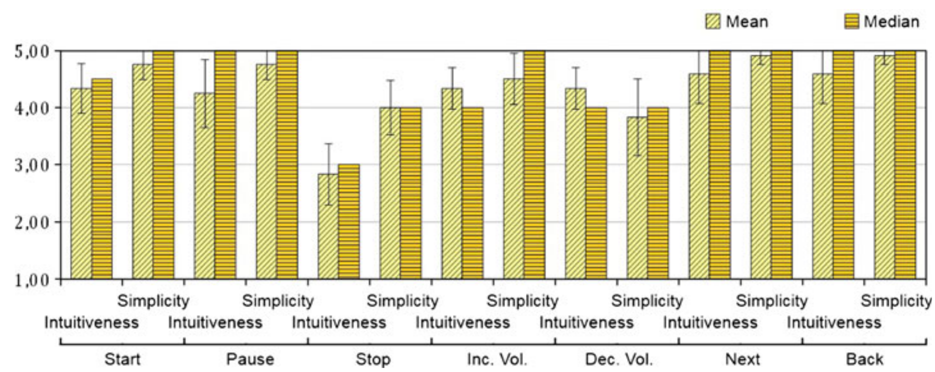

Fig. 24 Feedback for dynamic gestures with a confidence interval 95%

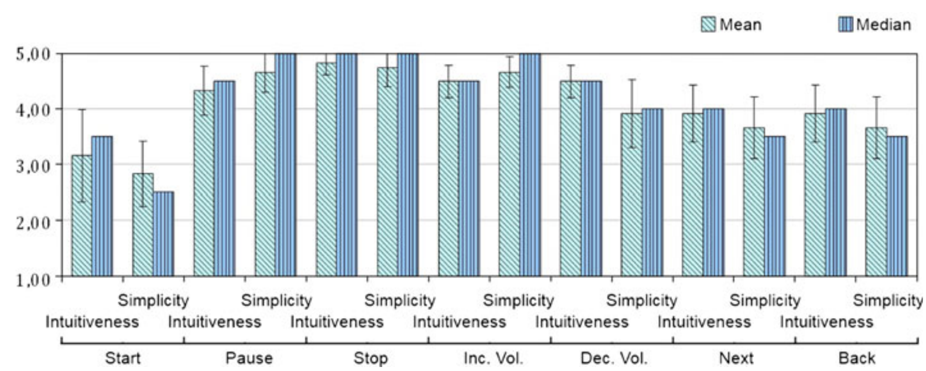

Fig. 25 Feedback for static gestures with a confidence interval 95%

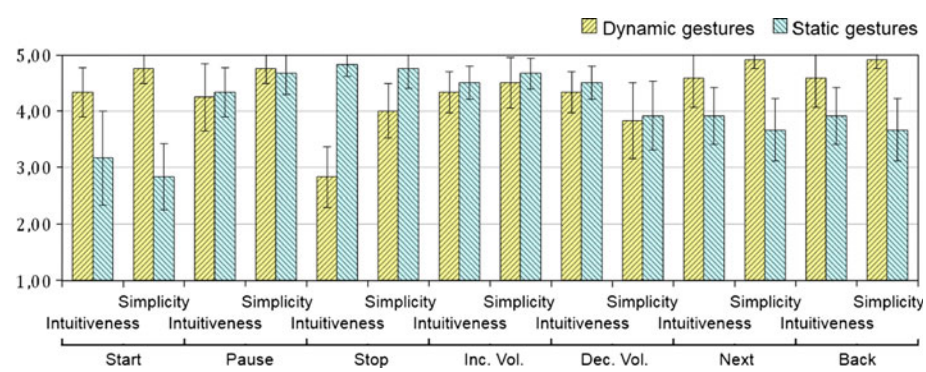

Fig. 26 The means of static and dynamic gestures for simplicity and intuitiveness with a confidence interval 95%

repeat. What this function exactly should do differs from toggling on or off the repetition of the current track to switching between different modes of what should be repeated via a gesture. The proposed repeat modes were do not repeat, repeat current track, repeat current album, repeat current playlist and repeat tracks of current artist. As gesture people suggested sometimes to use a circling hand. A few persons proposed more complex functions, like editing the playlist or to tag a track. Therefore, they would use a small set of gestures in the same way the few buttons on a MP3-Player are used to do that. As example it was suggested to create a simple menu, where a person could interact with using the next and back gestures for moving left or right within the menu. To select a menu-item the start gesture could be used whilst the stop gesture was proposed to be used to cancel an operation. That way only one extra gesture for opening a menu was needed to do more complex tasks.

7.4.5 Additional user feedback

Two participants said that a visual feedback was very important. Also two times it was said that it would be best to mix both function sets. One person emphasised that it did not make sense to use gestures in any context. Another person wanted to confirm the recognition of a gesture, before the system starts the function. Moreover, it was said that it could be nicer to only use one hand for the gestures.

7.5 Discussion

In the first part of this evaluation we ensured that all the basic functionality is covered by the designed gesture sets. More functionality such as a shuffle or repeat toggle, as mentioned some times in the third part, could be added in the future, but so far gestures for the most wanted functionalities were created.

In the second part of this evaluation the defined gesture sets were tested against each other. On average the dynamic set of gestures was rated higher, than the static gestures. But this result is not significant, as only a few gestures differ significantly, but not the whole sets. To make the dynamic gesture for stop more simple and intuitive, using the corresponding static gesture combined with a movement towards the webcam could be tried. For the gesture for start, a gesture that is inspired by the dynamic gesture could be used. Maybe it is more intuitive and simple to point towards the webcam, than create a triangle with your hands.

The creativity of the participants in the third part of this evaluation was surprising. Many functions and their gestures were suggested. Apart from simple function as e.g. repeating a track, or choose the next track randomly, some participants also suggested more complex functions, as seen in the results, that need menu-structures. It would be interesting to get to know if the participants decided to use menus with gestures, because they are used to menu structures or because it is intuitive. Second, in future work it could be tested if participants would rather use the proposed menu structure for more complex functions or physical artefacts as we suggested in Sect. 4.

As this evaluation is part of the final step of our proposed procedure, we were able to evaluate two sets of gestures against each other. The most critical part when doing so is to choose the right gestures out of the variety of gestures proposed by the users in the previous step. Because we got a positive result for the most gestures, it seems to be a good approach to use the gestures that were suggested the most as a first approach and then replace the gestures that do not seem to fit into the set of gestures. On the other hand, it could be better to use other characteristics of gestures to split them into different groups. Defining static and dynamic gestures did not give a significant difference. A revised set of gestures that will use both kinds of gestures will be introduced in Sect. 7.6.

As part of this step in the process the participants also gave qualitative feedback. Most participants did not want different gestures. Also other functionalities were not missed too much by the participants. However, some people suggested functions and gestures for functions that were not mentioned in the previous steps. This shows that the users should have the opportunity to give qualitative advices at this step to confirm the outcomes of a previous step.

7.6 Revised design

To derive a final consistent gesture set, we combined the sets of gestures. Therefore, we chose the gestures that were observed to be the simplest and most intuitive ones for their function. After that the definition of some gestures were changed to be more clear. To increase the chance of recognition for a system, we tried to define gestures that have unique hand shapes so that the movement is in most cases optional. The gestures introduced in the following might be less comfortable for a left-hander. That could be avoided, by also defining the gestures for start and change volume to work with the left hand.

7.6.1 Start

This gesture is defined as the dynamic one in Sect. 6. The difference is that the movement is not needed. Therefore, it is important to point with the right index finger towards the webcam.

7.6.2 Pause

This gesture is the same as the static one. It is similar to the gesture for stop so that they go well together. Also it is similar to the symbol for pause so that it is easy to remember.

7.6.3 Stop

Because the static gesture was rated higher than the corresponding dynamic one, the static gesture is used for this function.

7.6.4 Next track

For this function the dynamic gesture is used. To the previous definition is added that the side of the hand must be the from the right hand, to make this gesture easier to recognize by the system.

7.6.5 Previous track

Analogous to next track. The movement is to the left and the left hand is used when performing the gesture.

7.6.6 Increase volume

When ignoring the movement, the dynamic and static gestures are the same for this function. To make it easier to recognise, only the right hand should be used for this gesture.

7.6.7 Decrease volume

Analogous to increase volume, but this time the index finger of the right hands points to the bottom.

8 Conclusions and future work

In this paper we describe a process to derive gestures from strong user involvements. We refine the process proposed by Nielsen et al. [24] with the aim of collecting more information from each of the user studies associated. In contrast to Nielsen et al.'s approach, the outcome of each step is validated in the subsequent step and the final gesture set is determined by comparing different gesture sets in a controlled experiment. The process is applied to the design of an interface to control music playback using free-hand gestures. In step 1, based on an explorative user study the usage context is concretised and an initial set of necessary functions is collected. In step 2, we successfully validated the function set in the subsequent user study and collected a large number of gestures for the functions using participatory design techniques. In step 3, two gesture sets are derived and evaluated in step 4 in a comparative study.

The main contribution of the paper is twofold: first, we have shown that the added validation in step 2 ensures a suitable function set. Second, the definition of more than one gesture set and their experimental comparison in step 3 and 4 provide a reference framework to judge the individual gestures. This proved to be beneficial since it is not necessary to exclude promising candidates for gestures in the early design phases. As a result of the case study we also provide a consistent and validated set of gestures. These do not only prove that the refined process can successfully be used to define functionalities and a set of free-hand gestures, but the result set can also serve as a realistic benchmark for gesture recognition algorithms.

While we intentionally focused on developing a gesture set for controlling music playback the found gestures can be transferred to other application domains. Controlling the playback of similar multimedia content such as videos and multimedia presentations requires an analogue set of functions. Therefore, it can be assumed that the developed gestures can also be used for such applications. We are particularly interested in using the gesture set for controlling music playback on mobile phones facilitating accelerometer-based gesture recognition. Furthermore, the gesture set can serve as a starting point when developing gestures for specific applications that require a larger number of functions.

Throughout this paper, we intentionally refrain from considering technical limitations such as the performance of gesture recognition techniques. In future work it should be analysed how the consideration of technical limitations from early on affects the outcome of the process. Future work should also tell if the user preferences for gestures change over time. In particular, we are interested in how the gestures we found perform in long-term studies. Furthermore, it should be investigated how a predefined set of gestures perform compared with gestures sets that are defined by individual users.

References

1. Akers, D.: Wizard of oz for participatory design: inventing a gestural interface for 3d selection of neural pathway estimates. In: Ext. Abstracts CHI, pp. 454–459. ACM (2006)
2. Baudel, T., Beaudouin-Lafon, M.: Charade: remote control of objects using free-hand gestures. In: Proc. Commun ACM, vol. 36, pp. 28–35 (1993)
3. Bergman, J., Kauko, J., Keränen, J.: Hands on music: physical approach to interaction with digital music. In: Proc. Mobile HCI, pp. 33:1–33:11. ACM, New York (2009)
4. Bolt, R.: Put-that-there: voice and gesture at the graphics interface. In: Proc. SIGGRAPH, pp. 262–270. ACM (1980)
5. Choi, E., Bang, W., Cho, S., Yang, J., Kim, D., Kim, S.: Beatbox music phone: gesture-based interactive mobile phone using a tri-axis accelerometer. In: Proc. ICIT, pp. 97–102. IEEE (2005)
6. Description of the microsoft kinect. http://www.xbox.com/kinect/ (2011)
7. Freeman, W., Roth, M.: Orientation histograms for hand gesture recognition. In: Proc. IWAFGR, vol. 12, pp. 296–301 (1995)

8. Freeman, W., Weissman, C.: Television control by hand gestures. In: Proc. IWAFGR, pp. 179–183 (1995)
9. Gavrila, D.: The visual analysis of human movement: a survey. Comput Vision Image Underst **73**(1), 82–98 (1999)
10. Giles, J.: Inside the race to hack the kinect. New Sci **208**(2789), 22–23 (2010)
11. Hayafuchi, K., Suzuki, K.: Musicglove: A wearable musical controller for massive media library. In: Proc. NIME, pp. 259–262. ACM (2008)
12. Henze, N., Boll, S.: Snap and share your photobooks. In: Proc. MM, pp. 409–418. ACM (2008)
13. Henze, N., Boll, S.: Designing a cd augmentation for mobile phones. In: Proc. CHI, pp. 3979–3984. ACM (2010)
14. Hesselmann, T., Boll, S., Heuten, W.: Sciva—designing applications for surface computers. In: Proc. EICS. To appear. ACM (2011)
15. Hofmann, F., Heyer, P., Hommel, G.: Velocity profile based recognition of dynamic gestures with discrete hidden markov models. In: Gesture and Sign Language in Human–Computer Interaction, pp. 81–95 (1998)
16. Kela, J., Korpipää, P., Mäntyjärvi, J., Kallio, S., Savino, G., Jozzo, L., Marca, S.: Accelerometer-based gesture control for a design environment. Pers Ubiquit Comput **10**(5), 285–299 (2006)
17. Kranz, M., Freund, S., Holleis, P., Schmidt, A., Arndt, H.: Developing gestural input. In: Proc. ICDCS, pp. 63–63. IEEE (2006)
18. Kray, C., Nesbitt, D., Dawson, J., Rohs, M.: User-defined gestures for connecting mobile phones, public displays, and tabletops. In: Proc. Mobile HCI, pp. 239–248. ACM (2010)
19. Liang, R., Ouhyoung, M.: A real-time continuous gesture recognition system for sign language. In: Proc. FG, pp. 558–567. IEEE (1998)
20. Masui, T., Tsukada, K., Siio, I.: Mousefield: A simple and versatile input device for ubiquitous computing. In: Proc. UbiComp, pp. 319–328 (2004)
21. Mistry, P., Maes, P.: Sixthsense: a wearable gestural interface. In: Proc. SIGGRAPH ASIA. ACM (2009)
22. Moeslund, T.B., Störring, M., Granum, E.: A natural interface to a virtual environment through computer vision-estimated pointing gestures. In: Gesture and Sign Language in Human–Computer Interaction, pp. 59–63 (2002)
23. Morris, M., Wobbrock, J., Wilson, A.: Understanding users' preferences for surface gestures. In: Proc. GI, pp. 261–268. Canadian Information Processing Society (2010)
24. Nielsen, M., Störring, M., Moeslund, T., Granum, E.: A procedure for developing intuitive and ergonomic gesture interfaces for hci. In: Gesture-based Communication in Human–Computer Interaction, pp. 105–106 (2004)
25. Pylvänäinen, T.: Accelerometer based gesture recognition using continuous hmms. In: Pattern Recognition and Image Analysis pp. 639–646 (2005)
26. Rico, J., Brewster, S.: Usable gestures for mobile interfaces: evaluating social acceptability. In: Proc. CHI, pp. 887–896. ACM (2010)
27. Schlömer, T., Poppinga, B., Henze, N., Boll, S.: Gesture recognition with a wii controller. In: Proc. TEI, pp. 11–14. ACM (2008)
28. Stenger, B., Woodley, T., Cipolla, R.: A vision-based remote control. In: Computer Vision, pp. 233–262 (2010)
29. Symbian moove—gesture controlled music player. http://www.eyesight-tech.com/ (2010)
30. Wobbrock, J., Morris, M., Wilson, A.: User-defined gestures for surface computing. In: Proc. CHI, pp. 1083–1092. ACM (2009)
31. Wobbrock, J., Wilson, A., Li, Y.: Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes. In: Proc. UIST, pp. 159–168. ACM (2007)
32. Wu, J., Pan, G., Zhang, D., Qi, G., Li, S.: Gesture recognition with a 3-d accelerometer. In: Proc. Ubiquitous Intelligence and Computing, pp. 25–38 (2009)
33. Wu, Y., Huang, T.: Vision-based gesture recognition: a review. In: Gesture-Based Communication in Human–Computer Interaction, pp. 103–115 (1999)