Who Has the Force? Solving Conflicts for Multi User Mid-Air Gestures for TVs

Conference Paper · June 2016

DOI: 10.1145/2932206.2932208

3 authors:

Katrin Plaumann
Ulm University
8 PUBLICATIONS **41** CITATIONS
SEE PROFILE

David Lehr
Ulm University
3 PUBLICATIONS **5** CITATIONS
SEE PROFILE

Enrico Rukzio
Ulm University
220 PUBLICATIONS **2,954** CITATIONS
SEE PROFILE

Who Has the Force? Solving Conflicts for Multi User Mid-Air Gestures for TVs

Katrin Plaumann **David Lehr** **Enrico Rukzio**
Institute of Media Informatics
Ulm University, Ulm, Germany
<firstname>.<lastname>@uni-ulm.de

ABSTRACT
In recent years, mid-air gestures have become a feasible input modality for controlling and manipulating digital content. In case of controlling TVs, mid-air gestures eliminate the need to hold remote controls, which quite often are not at hand or even need to be searched before use. Thus, mid-air gestures quicken interactions. However, the absence of a single controller and the nature of mid-air gesture detection also poses a disadvantage: gestures preformed by multiple watchers may result in conflicts. In this paper, we propose an interaction technique solving the conflicts arising in such multi viewer scenarios. We conducted a survey with 64 participants, asking them about their TV viewing habits, experienced conflicts and opinions on conflict solving strategies. Based on the survey's results, we present a prototype for multi viewer gestural controls for TVs which solves possible conflicts.

ACM Classification Keywords
H.5.2 Information Systems: User Interfaces

Author Keywords
multi user gesture control;
mid air gestures; conflicts; TV gesture control; multi user;

INTRODUCTION
In recent years, mid-air gestures have become a viable alternative for controlling and manipulating digital content. One popular example is using gestures to control TVs instead of traditional remote controls [2, 5, 7, 8]. Low cost gesture recognition sensors and toolkits found their way into modern living rooms. Nowadays, even consumer products are shipped with gesture controls, allowing users to use gestures to navigate menus and manipulate content [6].

One of the main differences between gesture controls and traditional remote controls is that gesture controls do not require users to hold a dedicated device for interacting with the TV. This poses several advantages over traditional remote controls. For instance, remote controls are not always at hand and have to be grabbed first before a command can be given. This is omitted with gesture controls, where users just perform the gesture for a corresponding command. Another advantage of gesture controls is that they cannot be lost or hidden like remote controls. Additionally, gesture controls are more hygienic, since no surface needs to be touched by multiple persons.

However, gesture controls constitute a different problem, which was not at all a problem with traditional remote controls: it is now possible for more than one person to manipulate content at the same time, simply by simultaneously performing gestures. This leads to several conflicts, both technical and social. Technical conflicts arise since the system needs to interpret all given commands and decide which should be executed. Social conflicts arise when users perform contradicting gestures, but also when the solution for a technical conflict is not satisfying for all.

Despite of the substantial body of work covering mid air interaction and recent works showing that co-watchers influence interaction [5, 8], solving conflicts in multi user gesture control scenarios has not been extensively researched in the past. The few works focussing on this topic mostly focus on interaction on tabletops and collaborative work scenarios [1, 4]. To fill this gap in knowledge, we developed three conflict solving strategies based on previous work. Those strategies were assessed by 64 participants in an online survey. This survey also showed that gesture controls are a feasible alternative for remote controls, and participants reflected upon conflict prone scenarios. Based on the survey's results, we implemented a prototype preventing technical as well as social conflicts. Our contributions therefore are:

- An online survey assessing multi watcher scenarios and conflict prone situations for mid-air interaction for TVs as well as conflict solving strategies based on previous work
- A prototype implementation for solving and preventing conflicts for mid-air interaction applicable not only to TVs, but digital content in general

The remainder of this paper is structured as follows: First, we will give an overview over related work. Subsequently, we will describe possible interaction techniques for solving conflicts based on prior work, the survey we conducted to assess conflicts from user perspective, and our prototype for solving conflicts in multi user mid-air interaction scenarios.

TVX'16, June 22-24, 2016, Chicago, IL, USA.

http://dx.doi.org/10.1145/2932206.2932208

RELATED WORK

Gesture Controls for TVs

There exists a substantial body of work covering gesture detection as such and especially for TVs. One of the many projects showing that in the context of watching TV, gestures detection works quite reliable was conducted by Lee et al. [2]. However, they solely focussed on the detection of gestures, not on interaction concepts. This was done by Vatavu [7], who developed a gesture alphabet for interacting with TVs using a user elicitation study. Yet they only regarded single watcher scenarios, while we focus on multi watcher scenarios and their possible conflicts.

Multi User Gestural Interaction

Morris et al. [3] researched cooperative gestures for co-located groupware. Cooperative gestures are gestures were the gesture of a single team member contributes to a command given by the whole team. This type of gestural interaction impose a certain degree of teamwork. Besides focusing on mid-air gestures, we also focus more on the conflicts arising from multi user gestural interaction than on cooperation.

In the context of watching TV, multi user scenarios were researched by Ruiz et al. [5]. They developed gesture alphabets for omnidirectional videos using user elicitation, including both single user and multi user scenarios. One of their findings was that the gestures used to interact are performed slightly different in multi user scenarios. However, they did not further analyse conflicts arising in multi user scenarios or how they could be solved.

Zoric et al. [8] further analysed gesture based interaction for controlling TVs in multi user scenarios. In their study, co-watchers could manipulate content with various gestures. Observations of participants' behaviours showed that performing gestures contributed to the social watching experience. Therefore, the authors argue, social needs should be considered when designing interaction concepts. While our scenario is similar to the scenario of Zoric et al., our work differs in focusing on solving conflicts.

Conflict Solving Strategies

To the best of our knowledge, the most elaborate work proposing conflict solving strategies for multi user gesture controls have been both conducted with touch based interaction on tabletops, and not mid-air gestures. FlowBlocks [1], for example, is an interface especially developed for crowd interaction around multitouch tabletops. Besides at set of constraints simplifying crowd interaction, FlowBlocks also prevents conflicts by increasing mutual awareness of other users intents and physically blocking actions represented by user interface elements.

Morris et al. [4] researched conflict scenarios in co-located co-operative work scenarios. Based on their observations of such scenarios, they argue that simply relying on social protocols does not sufficiently solve conflicts and thus propose several conflict solving strategies.

CONFLICT SOLVING INTERACTION TECHNIQUES

Based on the conflict solving strategies used in previous work, three different interaction techniques were developed. Those three techniques are *master user*, *rank*, and *voting*. In the following, each technique is explained in detail.

Master User

This technique is based on the proactive coordination policies proposed by Morris et al. [4]. It also corresponds to the interaction concept proposed in FlowBlocks [1], and a variant was implemented by Zoric et al. [8].

Comparable to holding a remote control, only one person is in control of the content, and is thus the master user. With this strategy, conflicts are completely avoided. The only thing that needs to be negotiated is who the master user is. This negotiation could occur at the beginning and the result being in effect for the complete watching session, or renegotiated like in Zoric et al. [8]. The first approach leads to a *permanent master user*, while the latter leads to *varying master users*.

The main advantage of this strategy is its conformity to using remote controls, leading to a system behaving according to user's expectations. Yet the interaction time is longer, since becoming the master user requires an additional interaction step.

Rank

Another possibility for solving conflicts is ranking users. In case of a conflict, the gesture performed by the higher ranked user will be regarded by the system. This strategy was previously described by Morris et al. [4], and requires to rank the users beforehand. *Rank* could be implemented by assigning a rank to each person in the household and leveraging for example facial recognition to identify them. This implies a *fixed rank* of all watchers, that is household or family members. Another more flexible implementation could require co-watchers to assign a new rank each time they watch TV. Instead of being set explicitly, ranks could be assigned implicitly, for example by assigning the first person starting to watch the highest rank while the second person joining is assigned to the second highest rank and so on.

Rank leads to shorter interaction times, since users only need to execute the control gesture. Also, as long as the watcher's ranks are obvious, the system behaves as users would expect. However, especially with *fixed ranks*, and depending on the implementation of the ranking process, watchers could feel discriminated and being without a chance to control the content.

Voting

As in the co-located cooperative work scenario with tabletops described by Morris et al. [4], voting could also be applied for mid-air interaction for TVs. As soon as a user performs a gesture, the system would ask all users to either *approve* of or *veto* against the gesture. When approval is needed, all watchers agreeing to the gesture have to perform an approval gesture. If this is the majority, the associated action is triggered. When users are allowed to veto, they also perform a special gesture to express their will. If this is the majority, the associated action is not triggered.

Obviously, this is the most democratic strategy for solving conflicts, since every user can express their opinion. However, the time it takes to execute an action (or not) is longer, and the outcome not as predictable as with the other two proposed conflict solving strategies.

Incidence of discussions about ...

Number of co watchers (amount of answers)	1 (5)	2 (27)	3 (19)	4 (11)	>4 (2)	Total (64)
Remote control as such	2	2	2	4	4	2
Volumn	2	2	3	3	4.5	3
Channel	5	3	4	4	4.5	4
Menu settings	1	1	1	1	1	1

Table 1. The median ratings of incidence of discussions about remote controls as such, as well as volume, channel and menu settings. Participants gave their rating on six point Likert scales, with 1 resembling "never" and 6 resembling "always". The results are given per number of average watchers and in total.

Participants found mid air interaction as means for controlling TVs ..

Age groups (amount of answers)	<18 (21)	18-20 (20)	21-24 (14)	25-39 (3)	>39 (6)	Total (64)
imaginable	4	4	4.5	4	3	4
applicable	4	4	4	4	2.5	4
preferable	3	3	3	3	2	3

Table 2. The medians for the Likert scale items regarding how imaginable, applicable, and preferable mid-air interaction for controlling TV is, with 1 resembling "not at all" and six resembling "definitely". The results are given per age group and in total.

SURVEY

Participants and Procedure

To further asses multi watcher scenarios, behaviour and conflict prone scenarios, we conducted an online survey. The link to the survey was made available to members of our local institution and local clubs through mailing lists. Participation was voluntary. 64 persons participated in the survey, of whom 30 were female. The average age was 21.75 (8.97 SD). The average amount of persons sharing a TV with our participants was 3.47 (1.13 SD). Participants reported to have at average 2.67 (0.96 SD) co-watchers.

After giving consent and providing us with their demographic data, participants answered several questions regarding their opinion on gesture controls for TVs, possible conflict scenarios, and which of the three previously mentioned conflict solving strategies they preferred. Questions were answered on six point Likert scales. Additionally, participants could express their opinion, upsides and downsides in free text fields for each Likert scale.

Results

Potential conflicts in multi-watcher scenarios

Table 1 shows the median ratings regarding incidence of discussions, with 1 resembling "never" and 6 resembling "always". As Table 1 shows, an increase of co-watchers is accompanied by an increase of discussions. Discussions most often cover the selection of channel, volume settings, and the remote control as such.

When asked about how conflicts were solved, participants answers showed that a variety of strategies were applied. Those strategies included rank, where the most senior family member decided, and fixed master user, where the owner of the remote control decided. A third strategy was to discuss e.g. the channel and find a solution satisfying for all co-watchers. If a compromise was not satisfying for all, the concerned participants tended to avoid further conflicts by e.g. starting to watch their channel on a different device, do something else or stay and accept the decision.

Gesture controls as alternative

The median ratings of gesture controls as alternative input modality are shown in Table 2. The results are given split by age group and in total. As can be seen, most participants could well imagine to use gestures to control their TV and also find this applicable. Yet still, they are not tending to prefer gesture controls over traditional remote controls. This contradiction was clarified in the free text fields. Participants were concerned with technical issues, making mid-air interaction unreliable and error prone. Gestures recognised as but not being intended to be control gestures were seen as problematic. Gestures should be easy to perform and simple. Most participants expressed the wish to use mid-air gestures when traditional remote controls were not at hand, impracticable or inconvenient.

Possible conflict scenarios for multi-user gesture controls

Figure 1 shows various scenarios potentially harbouring conflicts, and the percentage of participants rating them as conflict prone. As can be seen, most participants found scenarios involving contrary commands of two watchers most conflict prone. Note that the majority of participants rated scenarios were both watchers performed the same gesture as not prone of conflicts, although it could result in a conflict when the system executes the command associated to the gesture twice. This should be accounted for by system implementers, for example by ignoring a gesture when performed at the same time or shortly after the same gesture was performed.

Preferred conflict solving technique

Figure 2 depicts participants ratings regarding the usefulness of the previously presented conflict solving techniques. 56% of the participants at least tended to see *Voting* as useful. However, the number of participants rating this strategy as not useful

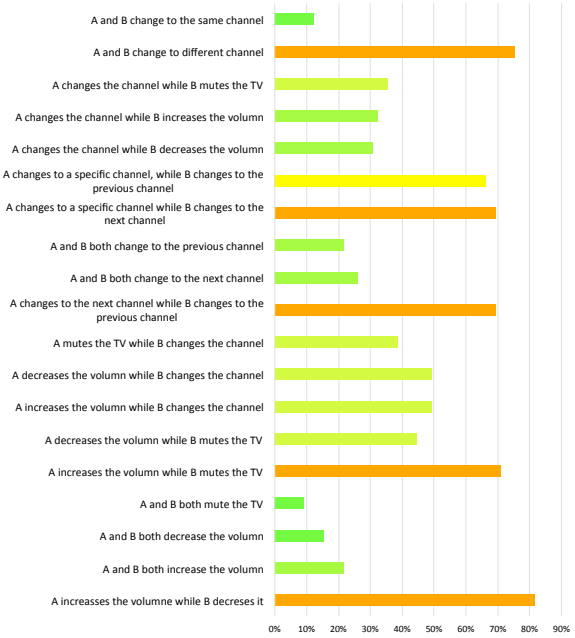

Figure 1. Potential conflict scenarios and the percentage of participants rating them as potential conflict scenarios.

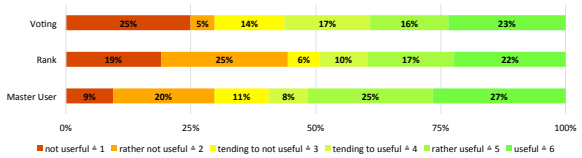

Figure 2. Rating of the usefulness of each conflict solving strategy. Participants rated the usefulness on a six point likert scale ranging von 1 (not usefull) to 6 (useful).

was the largest. This is substantiated by the free text comments: participants acknowledged the fairness and entertainment factor, but also criticize its impracticability and the time it takes to come to a decision.

Rank, on the other hand, was tended to be seen as useful and not useful by nearly the same amount of participants (49% and 51%, respectively), yet this is the strategy rated as at least "rather not useful" by most participants. Interestingly, most of the participants expressing a positive opinion on rank where either in the youngest or in the oldest age group. As positive aspect they mentioned that it is always clear who is in control. Yet most of the comments were rather negative. Participants were concerned that the ranking of persons would cause more conflicts than solve. It was seen as unfair, and enabling or strengthen power games between the co-watchers.

The third alternative, *master user*, was rated the most at least "tending to useful" and the least "not useful" conflict solving strategy. Participants mentioned that the concept was not so different to traditional remote controls. This was both seen as positive and negative. Positive, because conflicts are avoided, communication between co-watchers is promoted, and falsely detecting control gestures is avoided. Negative aspects are that still only one watcher at a time could control the TV, and this person should have to somehow release the control to other users.

Discussion

The survey results show that conflict scenarios exist. So far, those conflicts were mostly solved through social protocols. Conflicts are also expected to occur when mid-air gestures are used. Conflict prone scenarios are mostly changing channels and volume settings. Changing channels and volume settings might also be the most often performed use cases when watching TV.

Regarding mid-air gestures as input modality for TVs, our results show that mid-air gestures are at least today not likely to replace traditional remote controls. For this, especially the technical concerns are too great. However, mid-air interaction could well serve as an equal alternative to traditional remote controls. The decision between gestures and remote controls could be made depending on the situation, with users choosing the input modality more convenient for each situation.

As for conflict solving strategies, our results show that albeit being the closest to traditional remote controls, *master user* is a feasible solution. It is most accepted by users, and prevents not only social but also technical conflicts by recognising control gestures only from one person at a time. However, attention needs to be paid to a proper implementation. Instead of *fixed maser user*, users are more likely to accept *varying master user*, were the master user can change over time.

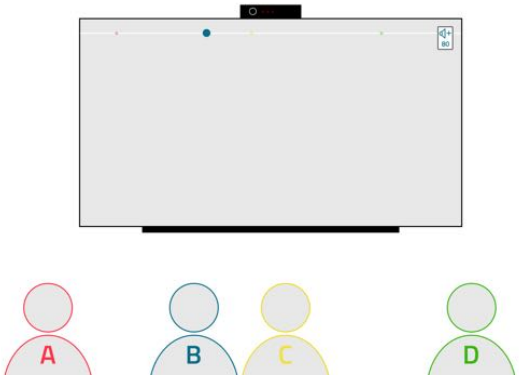

Figure 3. The design of our proposed user interface. At the top of the video, a line is displayed. On this line, a sphere for every watcher is shown. The spheres' positions resembles the relative position of the viewers to each other. The current master user is marked with a larger sphere. Upon the detection of a gesture, the sphere changes its opacity from the default 45% to 100% when the gesture is completed. Icons in the upper right corner visualize the action triggered by the detected gesture.

PROTOTYPE

The prototype user interface was designed based on the results of our survey, thus using the varying master user strategy for preventing conflicts. That means, that watchers have to perform a special gesture to become master user. They have the option of releasing control by performing another special gesture. To avoid one watcher never releasing control, we set a timeout, after which control automatically is released. Gesture recognition was implemented using a Microsoft Kinect 2 and its SDK. Video playback was implemented as a WPF application. In the following, the used gestures and the design of the user interface are explained. Figure 4 shows a screenshot of the prototype.

Gestures

Since changing channels and volume were seen as most conflict prone interactions, changing to the next and previous channel as well as increasing and decreasing the volume were implemented. For changing to the next and previous channel, swiping to the left and right respectively were implemented. The volume could be increased by moving the hand up and decreased by moving the hand down. Those gestures are similar to the ones describe by Ruiz et al. [5] and Vatavu [7] for the same functions. By default, no one is master user. Watchers can become master user by performing a special gesture: they have to outstretch their arm, form a fist and move their fist back to their shoulder. After a configurable timeout, the master user is revoked of their privileges and all watchers can become master users again. It is also possible to actively return ones master user privileges by performing the reverse gesture: moving the closed fist from the shoulder forward until the arm is outstretched, and opening the fist.

Graphical User Interface

Particular attention was paid to giving appropriate feedback to watchers. First, it should always be clear who the master user is. We therefore decided to show representations of users at the top of the video, as can be seen in Figure 3. Every user is

Figure 4. A screenshot of the final application. The spheres on the top of the video represent the watchers and their relative positions to each other. The watcher represented by the orange sphere is the master user, thus the sphere is larger than the others. The sphere is opaque, what indicates that a control gesture was detected, and the icons on the right upper corner of the video show that the volume was adjusted.

resembled through a uniquely coloured sphere. The positions of the spheres resemble the relative positions of the detected watchers to each other. The sphere of the master user is clearly distinguishable from the others through its larger diameter (blue sphere in Fig. 3). The second concern was giving appropriate feedback when gestures are recognized. Therefore, the already present spheres were leveraged. When no gesture is detected, the user's sphere has an opacity of 45%. When a gesture is detected, this changes up to full 100% opacity. This allows user to tell whether a gesture is detected, and how big it should be. When a gesture is recognized, the action triggered is further visualized through icons at the upper right corner.

CONCLUSION

This paper focused on conflict solving for multi user mid-air interaction for TVs. Mid-air interaction as input modality for TVs is getting more and more attention, and multi user scenarios are likely to occur when gestures are used to control TVs. Despite having several advantages over traditional remote controls, conflicts are more likely to arise with gesture controls. Thus, we proposed three conflict solving interaction strategies based on previous work. Further, we assessed multi watcher behaviour, potential of gesture controls and conflict prone scenarios in an online survey with 64 participants. Based on the results, we implemented a prototype preventing social and technical conflicts.

Albeit we focused on interaction with TVs in this paper, our proposed prototype and conflict solving strategy could also be used for manipulating other digital contents. For example, viewing images, viewing and manipulating 3D graphics, or manipulating objects in virtual reality. Also, the conflict solving strategies presented in this paper can not only be applied to gesture controls, but also to voice input and usage of multiple hardware controls like mobile devices.

Future plans regarding conflict solving strategies for multi user mid-air interaction include evaluating our proposed prototype in a user study, and implementing it for other digital content.

ACKNOWLEDGMENTS

This work was conducted within the context of the projects SenseEmotion (BMBF) and SFB/TRR 62 (DFG).

REFERENCES

1. Florian Block, Daniel Wigdor, Brenda Caldwell Phillips, Michael S Horn, and Chia Shen. 2012. FlowBlocks: a multi-touch ui for crowd interaction. In *Proc. of UIST '12*. ACM, 497–508.
2. Sang-Heon Lee, Myoung-Kyu Sohn, Dong-Ju Kim, Byungmin Kim, and Hyunduk Kim. 2013. Smart TV interaction system using face and hand gesture recognition. In *Proc. of ICCE '13*. 173–174.
3. Meredith Ringel Morris, Anqi Huang, Andreas Paepcke, and Terry Winograd. 2006. Cooperative gestures: multi-user gestural interactions for co-located groupware. In *Proc. of CHI '06*. ACM, 1201–1210.
4. Meredith Ringel Morris, Kathy Ryall, Chia Shen, Clifton Forlines, and Frederic Vernier. 2004. Beyond social protocols: Multi-user coordination policies for co-located groupware. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*. ACM, 262–265.
5. Gustavo Alberto Rovelo Ruiz, Davy Vanacken, Kris Luyten, Francisco Abad, and Emilio Camahort. 2014. Multi-viewer Gesture-based Interaction for Omni-directional Video. In *Proc. of CHI '14*. ACM, New York, NY, USA, 4077–4086.
6. Samsung. 2015. Samsung Smart TV - TV has never been this smart. (2015). http://www.samsung.com/us/experience/smart-tv/ Last accessed: 2015 - 01- 20.
7. Radu-Daniel Vatavu. 2012. User-defined Gestures for Free-hand TV Control. In *Proc. of EuroiTV '12*. ACM, New York, NY, USA, 45–48.
8. Goranka Zoric, Arvid Engström, Louise Barkhuus, Javier Ruiz Hidalgo, Axel Kochale, and others. 2013. Gesture interaction with rich TV content in the social setting. (2013).