

Measuring the impact of multimodal behavioural feedback loops on social interactions

Ionut Damian, Tobias Baur, Elisabeth André

Angaben zur Veröffentlichung / Publication details:

Damian, Ionut, Tobias Baur, and Elisabeth André. 2016. "Measuring the impact of multimodal behavioural feedback loops on social interactions." In Proceedings of the 18th ACM International Conference on Multimodal Interaction - ICMI 2016: Tokyo, Japan, November 12 - 16, 2016, edited by Yukiko I. Nakano, Elisabeth André, and Toyooki Nishida, 201-8. New York, NY: ACM Press.
<https://doi.org/10.1145/2993148.2993174>.

Nutzungsbedingungen / Terms of use:

licgercopyright

Dieses Dokument wird unter folgenden Bedingungen zur Verfügung gestellt: / This document is made available under the following conditions:

Deutsches Urheberrecht

Weitere Informationen finden Sie unter: / For more information see:

<https://www.uni-augsburg.de/de/organisation/bibliothek/publizieren-zitieren-archivieren/publizieren>



Measuring the Impact of Multimodal Behavioural Feedback Loops on Social Interactions

Ionut Damian
Human Centered Multimedia
Augsburg University
Augsburg, Germany
damian@hcm-lab.de

Tobias Baur
Human Centered Multimedia
Augsburg University
Augsburg, Germany
baur@hcm-lab.de

Elisabeth André
Human Centered Multimedia
Augsburg University
Augsburg, Germany
andre@hcm-lab.de

ABSTRACT

In this paper we explore the concept of automatic behavioural feedback loops (BFL) during social interactions. Behavioural feedback loops (BFL) are rapid processes which analyse the behaviour of the user in realtime and provide the user with live feedback on how to improve the behaviour quality. In this context, we implemented an open source software framework for designing, creating and executing BFL on Android powered mobile devices. To get a better understanding of the effects of BFL on face-to-face social interactions, we conducted a user study and compared between four different BFL types spanning three modalities: tactile, auditory and visual. For the study, the BFL have been designed to improve the users' perception of their speaking time in an effort to create more balanced group discussions. The study yielded valuable insights into the impact of BFL on conversations and how humans react to such systems.

CCS Concepts

- **Applied computing** → *Computer-assisted instruction*;
- **Human-centered computing** → HCI theory, concepts and models; Ubiquitous computing;

Keywords

feedback loop; behaviour analysis; live feedback; social signal processing; realtime; social augmentation

1. INTRODUCTION

With the arrival of new technologies, human-to-human communication has never been easier. We can poke, tweet, share and like with a single tap on the screen of our smartphone and, in doing so, interact with hundreds, if not thousands of other persons in an instant. However, in spite of these changes to our social life (or, perhaps, because of), a large part of the population is still experiencing fear of face-to-face social interactions [15]. This is especially true for key-life-situations such as public speaking where recent

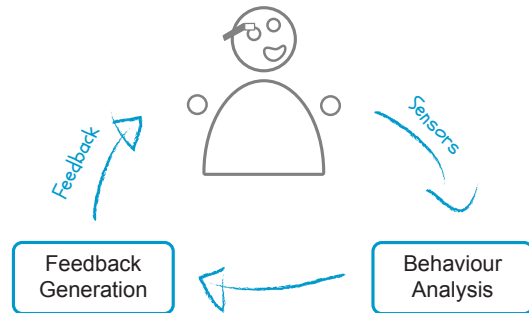


Figure 1: The behavioural feedback loop featuring its two main components: behaviour analysis and feedback generation.

estimates suggest that as much as 75% of the population experiences some form of anxiety or nervousness [7]. In such situations we often wish someone (or something) would give us tips on how to behave or what to say.

Automated realtime behavioural feedback loops (BFL) offer great promise in such scenarios by assisting the user in improving the quality of her or his behaviour and achieve more beneficial social interaction outcomes [8]. A BFL first analyses the behaviour of the user in realtime and then uses it to generate and deliver feedback to the user on how to improve its quality (see Figure 1). This is different from common behavioural training approaches which usually involve external persons (i.e coaches) providing feedback or reinforcement. The feedback loop itself is not a new mechanism, humans have always used environmental stimuli to adjust and adapt their behaviour in order to locate food, avoid danger or find a mating partner. However, only the recent boom in miniaturized sensing hardware, made the automation of this process possible.

Previous implementation of BFL, while preciously few, have already demonstrated it to be a viable mechanism for assisting users during critical situations such as public speaking [6, 16] and job interviews [5]. However, since BFL are meant to continuously augment the user's social interactions, it is only natural to ask how well can this process integrate within face-to-face interaction practices. Such social interactions have been mostly unaffected by the introduction of ever more diverse and ubiquitous technologies and an ever more intrusive media channel: We still look each other in the eyes and communicate using gestures and spoken language. Diverting from this archaic set of rules (e.g. texting while

This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in:

Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ICMI '16, November 12–16, 2016, Tokyo, Japan
ACM. 978-1-4503-4556-9
<http://dx.doi.org/10.1145/2993148.2993174>

talking, being a “glasshole”) is usually considered rude and impolite. Thus, one of the greatest challenges for BFL lies in minimizing the distraction effect of the technology and to seamlessly blend into the interaction.

To investigate this issue, we conducted an exploratory user study with 54 participants which were engaged in a group discussion. During the discussion, each participant was supported by a BFL which assisted the user in controlling her or his speaking time. The BFL’s main aim was to improve the balance and thus overall quality of the group discussion. During the study, we compared the effectiveness and disruptiveness of four different BFL types: auditory, tactile, visual (head-mounted) and visual (remote display). The BFL have been implemented using a custom built framework for designing and running BFL. At its core, the framework is composed out of two components. First, a behaviour analysis tool named SSJ which is capable of analysing and classifying social signals over multiple channels in realtime. Secondly, a feedback manager which is able to deliver multimodal feedback with the help of various output devices such as the Google Glass, the Myo armband or bone-conductance headphones.

Before presenting the implementation of BFL and the user study, we will first take a look at related work as well as discuss the general concept of BFL and what impact it can have on the user’s attention. Finally, we will discuss the results of the user study and conclude the paper by rehashing the main contributions and providing an outlook.

2. RELATED WORK

From a conceptual point of view, BFL share the goal of more common computer-assisted training mechanisms: help the user improve their behaviour. However, whereas in most such systems, this improvement happens during dedicated training sessions prior to the actual interaction (*a priori training*) or afterwards (*a posteriori training*), a BFL attempts to enhance the user’s skills during the actual event (e.g. a real job interview or when actually speaking in public) with the help of live feedback. To this end, it is neither a priori nor a posteriori training. This makes the disruption effect much more critical for BFL. During a training session, it is not a priority for the system to be comfortable or for the feedback to not distract, as training sessions are meant to be repeated multiple times and the user can learn to ignore such shortcomings. However, for instance, when using a BFL during a real job interview, if the feedback is even slightly distracting, it may negatively impact the outcome of the interview and the chances of getting a job.

Our literature survey revealed relatively few examples of such systems. Works such as Schneider and colleagues’ Presentation Trainer [11], Barmaki and Hughes’s TeachLive [2] and van der Linden’s MusicJacket [17], despite being theoretically capable of providing live feedback, have been designed with a priori training in mind and little attention has been given to the distraction effect of the systems.

A notable exception is the system introduced by Tanveer and colleagues [16]. Using the Google Glass, they analyse the audio signal of the user for speech rate and loudness, and display corrective visual feedback to the user in realtime on the Glass’s display. The aim of the system is to improve voice modulation and reduce monotonicity while speaking in public. Unfortunately, the system’s evaluation yielded ambiguous results both in terms of performance in alter-

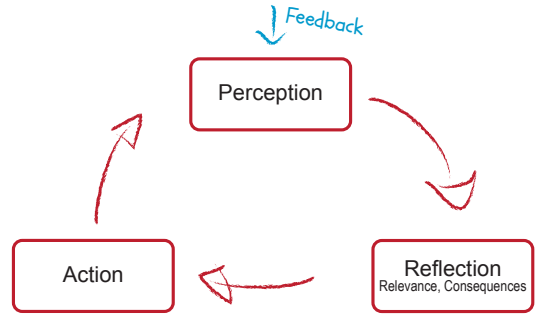


Figure 2: The three main phases of a feedback loop from the user’s point of view.

ing the users’ behaviours and distraction. More precisely, their analysis of the ratings of Mechanical Turk workers did not yield any significant effects in perceived behaviour quality or distraction. However, their self-report questionnaire data revealed that users preferred concrete what-to-do verbal instructions over a more abstract analytical graph-like feedback in terms of helpfulness, understandability and distraction.

A similar system proposed by Damian and colleagues [6], is able to deliver realtime feedback for public speaking scenarios using the visual or tactile [4] modality. However, despite promising initial results in terms of effect on behaviour quality, their evaluation is not exhaustive. Similarly to the work by Tanveer et al, they only measured the distraction effects using subjective questionnaires and semi-structured interviews. Although their visual feedback method differed from the ones implemented by Tanveer et al., they also report a high perceived helpfulness and a low degree of perceived distraction from the primary task. However, a direct comparison between different feedback methods is missing.

A more objective measurement of the distraction effect of live feedback was put forward by Ofek and colleagues [10]. In a user study they investigated what information delivery mode is better suited for providing users with secondary information during a human-to-human conversation. They tested three delivery modes, two visual and one auditory. Their results suggest that the audio modality is the most disruptive and least effective, and that delivering information visually in batches (rather than one at a time) is the least disruptive and most effective.

In this paper we aim to further explore the concept of automatic behavioural feedback loops and, for the first time, compare between different feedback methods spanning multiple modalities. We will place our focus on measuring the distraction effect of the feedback loop on the user, the user’s interlocutors and the social interaction itself.

3. BEHAVIOURAL FEEDBACK LOOPS

The feedback loop lies at the core of various psychological models, such as observational learning [12], operant conditioning [14] or social cognitive theory [1]. A feedback loop involves three main phases: perception, reflection and action. In the perception phase, the organism acquires information from an internal or external source. This information is then processed to evaluate its relevance, as well as any possible consequences this information may have for the organism.

Finally, an action is executed to place the organism in a more desirable state. This progression is illustrated in Figure 2

One very successful example of such automated feedback loops are dynamic traffic speed displays. These technology-enhanced traffic signs include a radar and a display to give drivers realtime feedback on their speed. They were first used in Garden Grove, California out of frustration over the poor success rate of conventional methods to change driving behaviour near schools. To the surprise of everyone, despite their simplicity and redundancy (they provide the drivers with information they already know), average speeds decreased by roughly 10%. In this example, in the perception phase, the driver sees the display and extracts the presented information. The driver then analyses the information, recognizes it refers to her or his behaviour, therefore deeming it relevant, and then remembers the consequences the information can have for her or him (e.g. possible accident or speeding ticket). Finally, the driver executes an action (reduces the speed of the vehicle) to change to a more desirable state (safety).

In our case, we look at behavioural feedback loops (BFL) relative to their power to change social behaviour by increasing the user’s awareness of her or his own behaviour, providing information on the quality of the current behaviour and clarifying what actions are needed to shift to a more desirable behavioural state. To achieve these effects, we propose a two-step pipeline (see Figure 1) where the behaviour of the user is first analysed in realtime and then, based on the user’s current behaviour quality, feedback describing the current behavioural state is automatically generated and delivered to the user. The feedback should be subtle enough to avoid disrupting the social interaction itself, but detailed enough to allow the user to evaluate its relevance and correctly extract the possible consequences.

From a psychological point of view, a BFL represents a secondary cognitive task which is executed in parallel to a primary task, i.e. the social interaction itself. The quality of the executions of the primary and secondary tasks is governed by how attention is distributed among them. Thus, we can say that the BFL is distracting the user and disrupting the social interaction if the execution of the feedback perception task draws too much attention from the primary task.

According to distributive attention models [9, 19], tasks can be carried out in parallel without quality degradation as long as enough processing resources are available. Thus, in order to reduce the amount of distraction caused by the behavioural feedback loop, the feedback delivery mechanism needs to be economical with its consumption of resources. More specifically, amount of consumed resources, frequency of processing and resource overlap with the primary task need to be minimized. One way to achieve this, is to deliver simple feedback with a low frequency using modalities which are not in use by the primary task. For example, in the case of social interactions which rely heavily on visual and audio tasks, tangible feedback might be less disruptive.

4. BFL FRAMEWORK

We implemented an Android-based software framework which allows the design, creation and execution of BFL. It consists of two main components: behaviour analysis and feedback management.

4.1 Behaviour Analysis

The first phase of a BFL is analysing the user’s behaviour. To this end, we implemented SSJ, an open source java library for recording, analysing and classifying social signals in realtime specifically designed for the Android platform. SSJ is able to run on virtually all devices in the Android ecosystem including common smartphones, smart glasses (e.g. Google Glass, Lumus DK-40, Epson Moverio) and smart watches, allowing an unprecedented flexibility and mobility for doing social signal processing.

SSJ has been inspired by Wagner and colleagues’ Social Signal Interpretation Framework [18] for windows, and thus borrows several design principles from it. More specifically, SSJ is build around a very modular design, with each processing step being encapsulated in a distinct component. There are three major types of components: sensors, transformers and consumers.

A sensor handles the connection and the data retrieval from various sensor hardware. Currently, SSJ supports several sensing hardware including standard Android inertial measurement units (acceleration, gyroscope and magnetometer), gps, standard Android cameras and microphones, the empatica armband, the myo armband and the microsoft band 2. Transformers are signal processing steps which transform the incoming data stream. Some examples of transformers currently in SSJ are: audio pitch, audio loudness, movement energy, physiological arousal, moving average and butfilter. For the audio processing transformers, SSJ makes use of TarsosDSP [13] and PRAAT [3]. The consumers are end-of-signal components, often entrusted with either signal classification or communication tasks.

SSJ applications are called pipelines and are formed out of a sequence of components through which data streams flow from component to component. Whereas the general principle of a pipeline is to maintain a constant data flow from left to right, parallel data streams (e.g. generated by multiple sensors) are also possible. In this case, all data streams are synchronized to a central clock. This way, even if a sensor misses frames or is not able to maintain the required sample rate, the data streams will still be consistent and comparable. For example, if a BFL requires both a heart rate sensor and an accelerometer to determine whether the user is stressed, SSJ will maintain the data streams coming from these two sensors synchronized so that any event recorded by the heart rate sensor can be matched to an event from the accelerometer. Furthermore, SSJ also allows multiple pipelines on different devices to be synchronized over the network. For instance, data recorded by the smart glasses on the user’s head is synchronized with the data recorded by a smart watch.

Despite its all-java core, SSJ is able to run complex signal processing pipelines in realtime at high update rates (e.g. 60 Hz) and thus ensures minimal output latency. SSJ is already being used in multiple research projects and is freely available for download¹.

4.2 Feedback Management

The delivery of the feedback to the user is handled by the feedback manager. It receives behaviour events from SSJ and based on a predefined feedback strategy, delivers feedback to the user with the help of various output devices.

¹<http://hcmlab.github.io/ssj/>

Similarly to SSJ, the feedback manager is also tailored to the Android platform and is thus able to work with various output devices including smartphones, smart glasses or smart watches.

The feedback manager is able to deliver both unimodal and multimodal feedback over three modalities: visual, auditory and tactile. Visual feedback can be rendered on smartphones, tablets or on head-mounted displays such as the Google Glass and takes the form of functional icons which can change in shape and colour. The tactile feedback consists of vibrations delivered through the Myo arm-band whereas the audio feedback uses audio files to play back sounds.

The feedback manager models a feedback event as a reaction to a behavioural cue according to the current feedback strategy. The strategy is defined using a simple xml file, allowing easy adjustments of the feedback mechanisms without requiring to rebuild the app. A sample feedback strategy is illustrated below:

```
<feedback>
<class type="tactile">
  <behaviour name="SpeechDuration" sum="true"
    ev="VoiceActivity" sender="SSJ"/>
  <event from="90" to="999" duration="500"
    intensity="70" lock="10000" multiplier="1.2"/>
</class>
</feedback>
```

In this example, the manager will trigger a tactile event (i.e. vibration) with a duration of 500 ms and an intensity of 70, when the total speech duration of the user exceeds 90 seconds. A feedback strategy can also be configured to track the user's response to a feedback event and, if necessary, adapt. In the example above, the feedback is configured to increase in intensity every 10 seconds if the user continues to speak.

The architecture of the component is modular and easily expandable to support new output devices or new behaviour cues. Furthermore, similar to SSJ, the feedback manager is also open source and freely available for download².

5. EFFECTS OF BFL ON SOCIAL INTERACTIONS

To get a better grasp on what impact a BFL can have on human-human interactions, we conducted a user study with 54 participants. For this, we designed and implemented an exemplary BFL aimed at improving the balance of group discussions by providing feedback to the participants regarding their speaking duration. The BFL was deliberately chosen to be simple in an effort to reduce the impact of personal variations in feedback interpretation and behaviour adaptation. To main goal of the study was to investigate the degree to which the BFL disrupts the social interaction both in terms of objective paralinguistic statistics and subjective questionnaire-based measurement. Furthermore, we explored the impact of not only the presence of the BFL, but also that of four different feedback delivery methods: auditory, tactile, visual (head-mounted) and visual (remote display). To this end, we formulate the following research questions:

RQ1. *Does the BFL improve the social interaction? How do the feedback methods compare?*

RQ2. *Does the BFL disturb the user? How do the feedback methods compare?*

RQ3. *Does the BFL disturb the other interlocutors? How do the feedback methods compare?*

5.1 Procedure

The user study featured a 4x2 mixed design. Table 1 summarizes the study design. Each evaluation session consisted of two group discussions between four participants, held on the same day back to back. During both discussions, each participant was given one of four output devices. Each participant from one group was given a different device which was chosen by the experimenter prior to the session and accounted for sensory-impairments (e.g. color-blind persons were not given visual feedback devices).

During the first discussion, i.e. the control condition, the BFL were inactive. The feedback loops were activated before the start of the second group discussion, i.e. the experimental condition. The participants were told that the BFL was active either in the first or the second discussion, based on chance. The group discussions were mostly non-moderated. The experimenter intervened only at the beginning of each discussion round, to propose a conversation topic (chosen based on the interests of the participants), as well as in the case the discussion ended prematurely, in which case the experimenter would propose a new topic. Each group discussion was designed to last 10 minutes. If the duration of the discussion exceeded 10 minutes, the experimenter would again intervene and stop it.

During each discussion round, each participant had three tasks. First, she or he was asked to participate in the discussion in a normal fashion while attempting to keep the discussion balanced, i.e. all participants should talk for roughly two minutes. Secondly, they were told that the devices will help them maintain this balance by giving them feedback on how much time they had left to talk. Thirdly, similarly to the study design of Ofek and colleagues [10], each participant was given a wireless clicker which they had to press every time they noticed one of their peers received feedback while talking. From a psychology point of view, the first task, the discussion, represents the primary task. The secondary task is either the feedback task (when the participant was talking) or the clicker task (when not talking). To achieve this, feedback was only delivered when the participant was talking and the clicker was only allowed to be used when not talking. Prior to the first discussion round, each participant received an extensive introduction on the study tasks and the function of the output device.

5.2 Measures

After each session, the participants filled out a questionnaire with items on the perceived disturbance of their own feedback device as well the devices of the others. All items consisted of a 7-point Likert scale where 1 was labelled as "I do not agree" and 7 as "I agree fully". In total, we analysed 13 items, with 4-5 items for each research question (Q_{x-y} = item y of RQ_x). The Table in Figure 5 list all the items of the questionnaire.

In addition to the questionnaires, video and audio data has been recorded for each session. For this, a webcam was installed in the room and each participant wore a close-talk

²<http://hcm-lab.de/logue/>



Figure 3: Setup of user study showing four participants, each wearing an output device: Myo armband (A), Aftershokz Bluez 2S bone conduction headphones (B), Google Glass (C), Microsoft Surface 2 Pro (D).

Table 1: Participant distribution across conditions and devices

| | | between-subject | | | |
|----------------|----|-----------------|-------|-----|---------|
| | | tactile | audio | hmd | monitor |
| within-subject | CC | 13 | 13 | 14 | 14 |
| | EC | 13 | 13 | 14 | 14 |

microphone. To aim of this data, especially the high quality audio recordings, was to allow us to analyse whether the behaviour of the participants changed between conditions. We decided against installing any more sensors to avoid artificially altering the behaviour of the participants and keep the interaction as natural as possible.

5.3 Participants

A total of 54 persons (mean age = 24.6, M/F = 40/14), split over 14 sessions, participated in the user study. Due to last-minute drop-outs, two sessions had to be carried out with only 3 participants. The gender distribution across sessions was as follows: MMMM = 4, MMMF = 7, MMFF = 1, MFF = 2. The distribution of participants over the conditions is detailed in Table 1. Most participants were students of either computer science, physics, mathematics, law or economics. Users were compensated with snacks and an online shop gift card raffle.

The participants were recruited two to three weeks prior to the evaluation sessions from various sources and, in most cases, had no prior connections between each other. The recruitment consisted of an online questionnaire which included items on demographics, session scheduling and interests. The data gathered from the recruitment questionnaire was used to create balanced discussion groups in terms of interests and gender.

5.4 Apparatus

For the evaluation study, we used four output devices: Myo Armband (tactile feedback, Figure 3-A), Aftershokz Bluez 2S bone conduction headphones (auditory feedback, Figure 3-B), Google Glass (head-mounted visual feedback, Figure 3-C) and Microsoft Surface 2 Pro (remote visual feedback, Figure 3-D). We built a BFL for each device. Each BFL tracked the amount of time a user has spoken using a head-worn close-talk microphone and a simple signal-to-

noise ratio-based voice activity detector. Based on this information, feedback containing the remaining speaking time was delivered to the user. The total speaking time for each participant per discussion was set at 120 seconds.

To keep the feedback consistent across the devices, feedback was delivered at the same intervals for every device. More precisely, feedback was delivered once the user reached 75%, 50%, 25% and 10% remaining speaking time. In the case of the auditory and the tactile devices, the feedback was discrete, i.e. it was only delivered at the before-mentioned intervals. For the visual devices, the feedback was provided continuously while the user was talking, however, it was only updated at the specified intervals.

The tactile feedback was generated by the Myo armband and consisted various vibrations patterns delivered to the user's forearm. More precisely, one short vibration was delivered at 75% remaining speaking time, two short vibrations at 50%, three short vibrations at 25% and one long vibration at 10%. If the user passed 0%, the Myo would deliver one long vibration every time the user started an utterance.

To provide auditory feedback, we used a pair of bone conduction headphones. The largest advantage of these headphones is their ability to propagate sound waves through the skull without disrupting the normal hearing process, and thus allowing the user to participate normally in the conversation. The auditory feedback was verbal for the 75%, 50%, 25% and 10% marks and consisted of a simple read-out of the remaining speaking time. Once the remaining speaking time was close to zero, the verbal feedback is replaced by a beeping sound which continued after the time has elapsed if the user proceeded to speak.

For both visual devices, the feedback was delivered using simple colour and text coding. More precisely, a feedback event consisted of the remaining time displayed in text form in the middle of the screen. The text was displayed on a coloured background which varied from green to red on a YUV spectrum according to the remaining speaking time. In order to only deliver the visual feedback to the user when she or he was talking (see Section 5.1), the display was only turned on when the user was speaking. The participants were positioned in a way which permitted only one user was to see the display of the tablet. For the Google Glass, the outside of display was covered using thick paper.

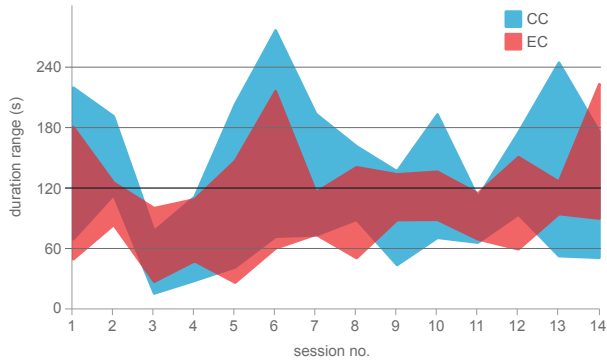


Figure 4: Speaking duration ranges (min - max) for all groups between conditions.

5.5 Results

Objective Measurements.

After the completion of the study, the audio data of each participant was processed using the PRAAT [3] toolbox to extract speaking duration, speechrate, loudness and pitch information. Overall, each participant spoke on average for 112.80 seconds during the control condition (CC) and 105.98 during the experimental condition (EC). This difference was however, not statistically significant. Nevertheless, when looking at the standard deviation (SD) of speaking duration within each group, we found significant differences between the conditions. More precisely, a paired sample T-Test revealed that the SD is significantly ($p = 0.017$) smaller in EC ($M = 30.92$) than in CC ($M = 45.07$). Similarly, the mean maximum speaking duration over all discussion groups is significantly ($p = 0.016$) smaller in the EC ($M = 143.53$) when compared with the CC ($M = 175.60$). These effects can be seen in Figure 4 where the duration ranges (min - max) for all discussion groups are plotted. One-Way ANOVAs revealed no significant differences between the devices for both conditions in terms of speaking duration, speechrate, loudness and pitch.

The click data of the participants was also recorded and statistically analysed. When comparing between CC and EC, a paired sample T-Test revealed a significant ($p = .014$) increase in clicks per discussion (CC: $M = .34$, $\text{Sum} = 24$, EC: $M = .97$, $\text{Sum} = 68$). However, ANOVA tests revealed no significant differences between devices for both discussions.

Subjective Measurements.

Statistical analysis on the questionnaire data also revealed interesting effects. When comparing the participants' rating against the scales midpoint (i.e. four), one-sided T-Test with applied Bonferroni-Holm correction yielded 8 significant differences. In particular, for the first research question where we investigated the effectiveness of the BFL, one item ($M_{Q1-2} = 4.72$, $SD_{1-2} = 1.73$) was rated significantly above the middle and two items ($M_{1-1} = 3.22$, $SD_{1-1} = 1.63$, $M_{1-3} = 2.89$, $SD_{1-3} = 1.25$) were rated as significantly below the middle. No items which were meant to measure the disruption effect on the user (RQ2) received ratings sig-

nificantly different than the middle value. On the other hand, all items of RQ3 were we measured the disruption effect on the other users scored significantly below the middle ($M_{3-1} = 3.02$, $SD_{3-1} = 1.89$, $M_{3-2} = 3.39$, $SD_{3-2} = 1.93$, $M_{3-3} = 2.07$, $SD_{3-3} = 1.31$, $M_{3-4} = 1.46$, $SD_{3-4} = 1.04$, $M_{3-5} = 1.91$, $SD_{3-5} = 1.48$).

First, a multivariate ANOVA revealed significant differences between the devices for Pillai's Trace, Wilks' Lambda, Hotelling's Trace and Roy's Largest Root. We followed this with univariate tests for each dependent variable. These yielded significant differences for the items Q3-1 ($M_{tactile} = 2.46$, $M_{audio} = 2.15$, $M_{hmd} = 4.21$, $M_{monitor} = 3.14$) and Q3-2 ($M_{tactile} = 4.00$, $M_{audio} = 2.15$, $M_{hmd} = 4.29$, $M_{monitor} = 3.07$), with participants rating the Google Glass as most disruptive. Significant differences between devices were also found for the items Q2-2 ($M_{tactile} = 5.15$, $M_{audio} = 3.00$, $M_{hmd} = 4.07$, $M_{monitor} = 3.57$) and Q2-3 ($M_{tactile} = 2.36$, $M_{audio} = 2.07$, $M_{hmd} = 4.20$, $M_{monitor} = 3.40$).

Furthermore, when looking at how participants rated the devices of the other interlocutors, a multivariate ANOVA revealed significant differences between the devices. Followup univariate tests showed significant differences for the item "The mere presence of the other person's {DEVICE} disturbed me" (Q3-5) when comparing between devices ($p = .001$). Participants felt that the Google Glass worn by the other interlocutors was the most disturbing ($M_{tactile} = 1.39$, $M_{audio} = 1.70$, $M_{hmd} = 2.57$, $M_{monitor} = 1.98$). The mean values of each questionnaire item are displayed in Figure 5.

5.6 Discussion

Overall, the study yielded some valuable insights into how BFL integrate with social interactions. The main findings are summarized in Table 2.

RQ1: Does the BFL improve the social interaction? How do the feedback methods compare.

To measure the first research question, we first investigated whether the behaviour of the participants in the experimental condition (system was on) was different than that in the control condition (system was off). This way we were able to observe that the behavioural feedback loops did have a positive effect on the behaviour of the users during the group discussions. More precisely, the group discussions in the experimental condition were more balanced in terms of speaking time distribution between users than the discussions in the control condition. This can be observed in Figure 4, where the speaking duration range for EC is "narrower" than for CC. Upon closer inspection, we found that the BFL had a larger impact on the more active persons, significantly reducing the maximum speaking duration across all sessions. However, no such effect can be observed on the minimum speaking duration, suggesting that the less active persons were not particularly affected by the feedback. Interestingly, these effect was not noticed by the participants themselves. Only one (Q1-2) out of RQ1's four items was rated significantly above the middle value (four) and two were rated significantly below the middle value.

When comparing between devices, we were not able to find any significant differences both in terms of objective and subjective measurements. This leads us to believe that all feedback delivery mechanisms, regardless of modality, were similar effective.

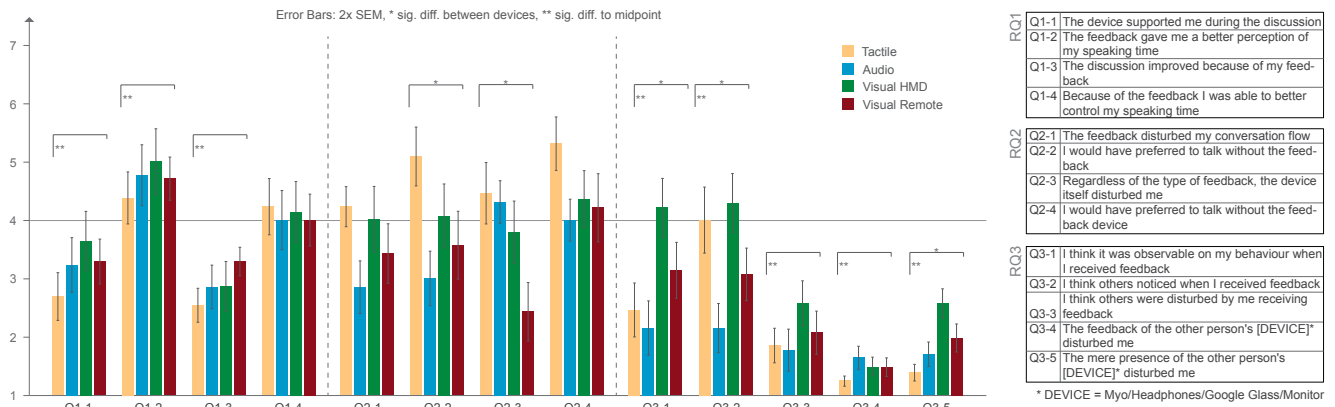


Figure 5: Mean values for post-session questionnaire.

RQ2: Does the BFL disturb the user? How do the feedback methods compare.

For RQ2 we first looked at the paralinguistic features and how they changed between conditions. If the BFL disturbed the participants we would expect some differences in terms of speaking rate or voice modulation. However, we were unable to find any evidence of such an effect.

Looking at the differences between devices, we were unable to find any significant differences in terms of paralinguistic features. However, the analysis of the questionnaire data did reveal that the participants found the Google Glass to be the most disturbing device and the tablet the least. This is not very surprising and can be explained by the high intrusion factor and field of view obstruction of the Google Glass and the low intrusion factor and the ordinariness of the tablet. Our analysis also revealed that the vibrotactile feedback was rated as the one they would have preferred to talk without the most (Q2-3). This effect is more interesting as it goes against our expectations. More precisely, since in a social interaction we rely most heavily on our vision and hearing, we expected the tactile feedback to be the least disruptive and the visual and auditory feedback to be the most. Our results may be explained by the fact that the users were not accustomed to this type of feedback and were startled by the vibrations. Also of interest is the fact that audio feedback was at the opposite end of Q2-3. This stands in contrast to the results reported by Ofek et al. [10], as they suggested that during vocal tasks, audio feedback is particularly disruptive. However, it is important to point out that in our scenario, audio feedback events were delivered in most cases only 4-5 times over a period of 2 minutes and consisted of a single short word. In Ofek’s study, up to 100 words were delivered during a 5 minute window. Furthermore, Ofek did not use bone conduction speakers. For our future work we plan to look closer at these effects and compare between multiple feedback strategies for each modality.

RQ3: Does the BFL disturb the other interlocutors? How do the feedback methods compare.

To measure how disturbing the device is for the others, we looked at the clicker data and the questionnaires. While the clicker data is not directly able to tell us if the BFL disturbs others, it does tell us if other persons are aware of BFL as the participants were instructed to press the clicker every time

they thought the speaker received feedback. Looking at this data we noticed an increase in overall amount of clicks in the experimental condition, which does show that the participants were in fact able to detect the presence of the BFL. We can thus state that the BFL’s impact on the participants’ behaviour was sufficient to be observable from the outside. However, the questionnaire data analysis yielded that this effect was not considered disruptive since all five items were rated significantly below the scale’s midpoint. Here we do want to note that the overall amount of clicks was quite low and during some sessions, the clickers were not used at all. When confronted by this fact, most participants said that they were so engaged in the discussion, they forgot about the clicking task. To this end, although successfully used by Ofek et al. [10], we were disappointed by this metric and thus find it difficult to recommend for other natural social interaction studies.

Although, generally speaking the BFL were not considered disturbing, we found that some devices were rated less disturbing, and others more. In particular, we found significant differences for how disruptive participants believed the feedback was (Q3-1 and Q3-2). Here, users generally thought that the audio feedback was least noticeable by others, whereas the feedback delivered on the Google Glass was the most noticeable. The participants rated that the tactile and auditory feedback as least likely to cause any unwanted noticeable changes in behaviour. When asking the opinion of the other users (i.e the participants not wearing the device in question), we found that they were significantly more disturbed by the Google Glass than the other devices (Q3-5). These results are in line with the “glasshole” phenomenon, according to which persons who use the Google Glass in public may be perceived as “creepy or rude”. This term attracted so much interest in 2014 that Google was forced to release guidelines on how to behave when wearing the Google Glass³.

6. CONCLUSION

The main contribution of this work lies in the exposition of behavioural feedback loops (BFL) as automated realtime multimodal processes, able to assist the user in improving her or his behaviour during social interactions. Further-

³<https://sites.google.com/site/glasscomms/glass-explorers>

Table 2: Summary of user study findings

| | General effects | Effects between feedback methods |
|------------|--|--|
| RQ1 | Discussions were more balanced when BFL were active | All feedback methods performed similarly well |
| RQ2 | BFL were not considered particularly disturbing or undisturbing by users | HMD was rated most disturbing device, tablet least. Tactile feedback was rated most disturbing feedback method, audio least. |
| RQ3 | Presence of BFL was observable by others (but was generally considered not disturbing) | HMD and HMD visual feedback are most disturbing to others. |

more, in an effort to encourage further research into BFL, we created an open source software framework for designing, building and executing BFL in a mobile environment. We then conducted a user study with 54 participants where we investigated the impact of four different BFL spanning three modalities on face-to-face human interactions. The BFL were designed to assist the users in keeping a balanced discussion, i.e. all participants should speak for roughly the same amount of time. The feedback was delivered either using tactile (Myo armband), auditory (headphones), visual head-mounted (Google Glass) or visual remote (common monitor) stimulation. The study yielded that although objectively effective, humans are sceptical towards the usefulness of such systems. Moreover, our participants found the Google Glass as a feedback delivery device to be the most disturbing. Interestingly, when rating the feedback methods themselves, participants found the vibrotactile feedback to be the most disturbing.

As part of our future work, we intend to explore different feedback strategies for each modality, e.g. other visualisations for visual feedback or other vibration patterns for tactile feedback. Furthermore, we would like to also test more advanced feedback delivery mechanisms, such as dynamically adapting the feedback strategy to the user’s reaction, simultaneously sending feedback over multiple channels or more intelligently controlling when the feedback is delivered based on current activity and cognitive state.

Acknowledgments

The work was partially funded by the German Federal Ministry of Education and Research (BMBF) under the project “GLASSISTANT” (FKZ 16SV7267K). We thank the students Christian Brummer, Tim Buchwald, Yang Peng Liu, Davin Geiger, Veli Hasanca, Sekip Tuncer, Nicola Hauptmann, Philipp Nowak, Yulia Khlyaka, Matthias Landsgeßell, Dominik Brindl and Marziye Tanha for their work on the user study.

7. REFERENCES

- [1] A. Bandura. Social foundations of thought and action: A social cognitive theory, 1986.
- [2] R. Barmaki and C. E. Hughes. Providing real-time feedback for student teachers in a virtual rehearsal environment. In *Proc. ICMI*. ACM, 2015.
- [3] P. Boersma and D. Weenink. Praat: doing phonetics by computer. <http://www.praat.org/>, retrv. 2013.
- [4] I. Damian and E. André. Exploring the potential of realtime haptic feedback during social interactions. In *Proc. TEI*. ACM, 2016.
- [5] I. Damian, T. Baur, B. Lugin, P. Gebhard, G. Mehlmann, and E. André. Games are better than books: In-situ comparison of an interactive job interview game with conventional training. In *Proc. AIED*, LNCS, pages 84–94. Springer, 2015.
- [6] I. Damian, C. S. Tan, T. Baur, J. Schöning, K. Luyten, and E. André. Augmenting social interactions: Realtime behavioural feedback using social signal processing techniques. In *Proc. CHI*, pages 565–574, 2015.
- [7] C. Hamilton. *Communicating for results: A guide for business and the professions*. Thomson/Wadsworth, Belmont, Calif., 9th ed. edition, 2011.
- [8] J. G. Hollandsworth, R. Kazelskis, J. Stevens, and M. E. Dressel. Relative contributions of verbal, articulative, and nonverbal communication to employment decisions in the job interview setting. *Personnel Psychology*, 32(2):359–367, 1979.
- [9] D. Navon and D. Gopher. On the economy of the human-processing system. *Psychological Review*, 86(3):214–255, 1979.
- [10] E. Ofek, S. T. Iqbal, and K. Strauss. Reducing disruption from subtle information delivery during a conversation: Mode and bandwidth investigation. In *Proc. CHI*, pages 3111–3120. ACM, 2013.
- [11] J. Schneider, D. Börner, P. van Rosmalen, and M. Specht. Presentation trainer, your public speaking multimodal coach. In *Proc. ICMI*. ACM, 2015.
- [12] S. J. Shettleworth. *Cognition, Evolution, and Behavior*. Oxford University Press, 2009.
- [13] J. Six, O. Cornelis, and M. Leman. Tarsosdsp, a real-time audio processing framework in java, 2014.
- [14] B. F. Skinner. The behavior of organisms: an experimental analysis, 1938.
- [15] M. B. Stein and D. J. Stein. Social anxiety disorder. *The Lancet*, 371(9618):1115–1125, 2008.
- [16] M. I. Tanveer, E. Lin, and M. Hoque. Rhema: A real-time in-situ intelligent interface to help people with public speaking. In *Proc. IUI*, pages 286–295, 2015.
- [17] J. van der Linden, R. Johnson, J. Bird, Y. Rogers, and E. Schoonderwaldt. Buzzing to play: Lessons learned from an in the wild study of real-time vibrotactile feedback. In *Proc. CHI*, pages 533–542. ACM, 2011.
- [18] J. Wagner, F. Lingensfelder, T. Baur, I. Damian, F. Kistler, and E. André. The social signal interpretation (SSI) framework - multimodal signal processing and recognition in real-time. In *Proc. ACM MULTIMEDIA*, 2013.
- [19] C. D. Wickens. Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science*, 3(2):159–177, 2002.