Identifying Mutual Engagement

Nick Bryan-Kinns
Queen Mary University of London
Centre for Digital Music
School of Electronic Engineering and Computer Science
Mile End, London. E1 4NS. UK.
nickbk@dcs.qmul.ac.uk

Fraser Hamilton
Fraser Hamilton Consulting Ltd
London, UK
fraser.hamilton@mac.com

Abstract. Mutual engagement occurs when people creatively spark together and enter a state of group flow. We present a characterization of mutually engaging interaction, discuss design features which contribute to mutually engaging interactions, and identify a set of measures for identifying mutual engagement in collaboration. A collaborative music editor's interface features are systematically manipulated in an empirical study of their effect on mutual engagement. Results of the study show that providing shared annotation mechanisms and awareness of identity of others significantly increases mutually engaging interaction.

Keywords: Mutual engagement; Collaboration; Design; Evaluation; Musical interface

1 Introduction

We are fundamentally social creatures, yet until recently, research into the role of technology in collaboration has focused on task oriented interaction. For example, research into Human-Computer Interaction has moved from studying office centric tasks of individuals to a focus on social and entertainment uses of computers (cf. Blythe et al., 2003). Similarly, research on the role of technology in collaboration has moved beyond the workplace in the last decade. For instance, Johnson and Hyde (2003) explored the collaborative work involved in solving a jigsaw puzzle. Similarly, in a recent special issue of the Computer Supported Collaborative Work (CSCW) journal there were investigations of non-'work' oriented collaborative activities such as collaborative and mobile gaming (e.g. Crabtree et al., 2007; Sall and Grinter, 2007), social co-ordination (e.g. Schiano et al., 2007), and the playful use of entertainment technologies in social situations such as trips to the zoo (O'Hara et al. 2007). Research has also diversified into collaborative virtual environments (e.g. Benford et al., 1995) where emphasis was placed on the role of embodiment in online experiences. The fruits of such research can be seen in contemporary popular online applications such as SecondLife which include full body avatars, awareness and presence mechanisms, and a variety of facial and gestural information. Interestingly, SecondLife is now itself the subject of study of collaborative work and mechanisms for supporting collaborative work (De Lucia et al., 2008). With these developments, the boundary between collaborative work and social interaction becomes blurred. Ijsselsteijn et al. (2003) provide a clarification of this boundary by distinguishing goal oriented from social and emotional oriented communication purposes. In their view CSCW informs design of support for the former, and research on connectedness informs design for the latter. However, current CSCW research points to a more porous and ill-defined boundary where work and social interaction overlap. This paper focuses on understanding design and evaluation in this little explored boundary space where collaborative activities are typically open-ended, social, and creative.

CSCW Research has also been used to inform the design of systems to support the collaborative artistic process such as WebStorm (Costa et al., 2007) which allows users to connect concepts and imagery across the web. Conversely, systems have been designed to support the collaborative process in music making (see Blaine and Fels, 2003 for a review), but these have typically not drawn on understandings of collaboration to inform their design, focussing rather on design of the musical production interfaces and issues such as complexity and expressivity.

In this paper we loosely use the term CSCW to refer to all research into the role of technology in collaboration. We explore the use of CSCW user interface design features to inform the design and evaluation of a collaborative music

system which aims to provide a mutually engaging collaborative experience. In particular, we focus on two key established design features: **awareness of identity**, and the role of supplementary (non-task) **communication channels**. In previous CSCW research, evaluations of such design features have primarily focused on the efficiency of interaction and collaboration, for example, examining what makes group work faster, logistically less complex, or less error-prone. We build on this to focus on the engagement between participants - the quality of the interaction between participants rather than the usability hurdles they need to overcome to collaborate.

2 Music Making

Music is a basic form of human expression found in all cultures. It can be both a cultural expression and a result of personal creativity. It conveys emotion, and transports us to different times and places. It is both public and private. Music making is frequently social and collaborative in nature (cf. Titon, 1996) yet there has been very little research on the topic in the CSCW literature. Trying to differentiate between the composition of music, its performance, and improvisation is problematic (cf. Bowers, 2002). For the purposes of this paper we consider music making as the act of a group of participants producing musical responses commensurate to the situation (a looser definition than Blum's characterization of group music improvisation, 1998).

Like face-to-face conversation, collaborative improvisation is traditionally co-present and multimodal, combining musical signals with verbal and visual cues. However, it has distinctive characteristics as a form of collaborative interaction. In particular, it is more strongly oriented toward:

- Mutual-engagement and aesthetic satisfaction rather than information exchange and task completion.
- Concurrent rather than sequential organization of contributions.
- Creativity and innovation with a resultant open-ended set of goals.
- Collaboration in which the process of interaction is itself a product for consumption by others within and outside the collaboration.
- Self-efficacy where perceived mastery of musical production mechanisms positively contributes to the experience.

These properties make musical collaboration uniquely interesting as a basic form of human interaction which has received little attention in CSCW literature, but which highlights generic issues such as engagement, innovation, and ensemble organisation that are important in a variety of collaborative contexts e.g. community development, team games, and brainstorming. Thus, understanding music making as an instance of collaboration could inform the design of products and services that support a wider range of collaborative situations. Also, it is worth noting that far from being a niche activity, music making, production, and enjoyment are a key part of the Creative Industries which contribute significantly

to national revenues. For instance, the United Kingdom government recently stated that 'within the UK, the Creative Industries sector contributes over 6.4% of UK Gross Value Added and is growing at a faster rate than the economy as a whole' (Technology Strategy Board, 2009).

Most musical devices do not explicitly support collaboration between participants, instead relying on physical proximity or other visual channels to convey co-ordinating information. Some notable exceptions are novel instruments such as squeezables (Weinberg and Gan, 2001), Jam-o-drum (Blaine and Perkis, 2000) and the tooka (Fels and Vogt, 2002) which are physically shared musical instruments. The design of these instruments is often informed by understandings of the nature of music performance and improvisation such as call and response patterns of interaction where one musician plays a musical motif which is then repeated by another member of the musical group. Research such as FMOL (Jordá, 2002), and WebDrum (Burk 2000) have begun to explore the collaborative and communicative requirements for group improvisation in geographically remote locations. This typically involves developing a shared visualisation of the music being produced and some communication support such as a text chat tool. See Blaine and Fels (2003) for a survey of the area of collaborative musical experiences, and Jordá (2005) and Barbosa (2003) for a survey of multi-user instruments.

Research communities exploring new forms of interactive support for music making such as NIME (New Interfaces for Musical Expression; Poupyrev et al., 2001) have conducted very little research into the evaluation of the interactive elements of the systems they develop (Fels, 2004), let alone the collaborative aspects (cf. Barbosa, 2003; Stowell et al., 2009) of these systems. For example, whilst Wanderley and Orio (2002) provide a much needed critique of HCI methods applicable to NIME, they reduce the evaluation to maximally simple tasks of controlling parameters on sound production rather than evaluating how expressive (cf. Dobrian and Koppelman, 2006) or engaging the experience is. The predominant form of evaluation in the field of creating new musical interfaces is introspective reflection on the idiosyncratic system. Moreover these new systems are typically only ever used (and possibly usable) by their creator. Such a paucity of understanding of design and evaluation of interactive music systems presents a great opportunity to explore existing understandings of design for collaboration from the CSCW literature in an under explored field. Moreover, the focus on aesthetics and engagement inherent in music making provides a new lens through which to observe collaboration and identify collaborative success.

3 Mutual Engagement

The general question we are interested in is: what characterizes mutual engagement in creative collaborations and how can we design technology to support it?

Intuitively, a key feature of creative collaborations is the *mutual engagement* between participants: the points at which people spark together, lose themselves in their joint action, and arrive together at a point of co-action 'where you are when you don't know where you are' (Tufnell and Crickmay, 1990). Such points are inherently difficult to identify and measure as the act of reflecting on mutual engagement undermines some of the characteristic qualities of the experience such as spontaneity. For our purposes, the distinguishing characteristic of mutual engagement is: *it involves engagement with both the products of an activity and with the others who are contributing to those products*. We argue that mutual engagement is essential for rich creative collaborations such as brainstorming, team based design, and improvisation.

Engagement itself can be characterized as a point at which participants feel that they are able to change and appreciate changes in the form (cf. Douglas and Hargadon, 2000) - it involves appreciation of possible contributions and anticipation of their outcomes. We see similar phenomena in accounts of *flow* (Csikszentmihalyi, 1991) - optimal experiences in which 'attention can be freely invested to achieve a person's goals' which results in a merging of action and awareness and consequent lack of self awareness and distortion of sense of time. In mutual engagement participants are similarly engaged with the product at hand, and also with others in the collaboration, which we could characterize as *group flow* (Sawyer 2003). Similarly, Miell and MacDonald (2000) propose that mutually engaged states are indicated by the 'exploration of the ideas of more than one person and the attempt to integrate these'.

Wenger's characterization of mutual engagement (1998) focuses on participants ability to 'engage with other members and respond in kind to their actions'. From Wenger's perspective, mutual engagement is about the work involved in learning how to interact with other people in an emerging community of practice. This entails the evolutionary development of identities, as well as 'establishing who is who, who is good at what, who knows what, who is easy or hard to get along with'. For Wenger, the focus is on the social 'work' that happens in mutually engaging communities.

In this paper we are concerned with identifying design features which have an effect on the mutual engagement between participants. As such we propose to operationalise the definition of mutual engagement as collaboration in which there is:

- **Evidence of engagement with the product of the joint activity** i.e. music in our domain. For example, participants' reports of feeling engaged with the product, a high quality product, focused contributions, or demonstrations of skills and expertise in creating contributions.
- **Evidence of engagement with others in the activity**. For example, more reports of feeling engaged with the group, coherent final joint products, co-location of contributions, mutual modification of work, discussions of

quality of the joint product, repetition and reinterpretation of others' contributions. Clearly this relies on participants' skills and expertise with the system.

3.1 Design for Mutual Engagement

We are concerned with exploring how CSCW design features can be used to inform the design of systems which encourage and support mutual engagement. CSCW has an extensive history of examining collaborative work and informing the design of collaborative systems, making it an ideal source of knowledge about designing for mutual *interaction* which we can add to by considering mutual engagement. In comparison to other fields such as computer games design, HCI, or Interaction Design, CSCW provides richer research into the nature of collaborative interaction, which compensates for its current lack of focus on engagement with and between participants. For example, studies have examined shared document editors such as ShrEdit (Dourish and Bellotti, 1992; Olsen et al., 1993), group decision support (e.g. Applegate et al., 1986), and collaborative brainstorming systems (Hymes and Olson, 1992). Typically the design of such systems has been informed by models of collaborative work and problem solving. Research has led to the characterisation of design features for collaboration such as awareness mechanisms, and the importance of shared representations. Indeed, Robertson (1997) identified the shared nature of representations as a key aspect of shared activities in a shared physical space. Similarly, Dourish and Bellotti (1992), and Gutwin and Greenberg (2002) expound the importance of awareness mechanisms which provide information about the shared activity. Typically, these have focussed on workspace awareness as the 'up-to-the-moment understanding of another person's interaction with the shared workspace' (Gutwin and Greenberg, 2002), including understanding who is in the shared workspace, what they are currently doing, what they have done in the past, and who they are interacting with. Conventionally the focus has been on geographically distributed support for activities such as manipulating documents, changing tasks and applications, and changing collaborative context e.g. grouping of participants. However, more recently an emphasis on co-located collaboration around tabletops has emerged (cf. Fischer, 2000, Terrenghi et al., 2006, Hilliges et al., 2007).

We see mutually engaging activities as residing at the boundary between goal and emotionally oriented communication (cf. Ijsselsteijn et al., 2003), and see an opportunity for exploring the applicability of CSCW research this domain. There is very little work on designing for this form of interaction, and yet, as discussed in the previous section, there is some understanding of what mutually engaging collaborations might be like. From our characterization of mutual engagement and the nature of naturalistic co-located interaction (Clark and Brennan, 1991), we believe that the following features are key to supporting the collaborative aspects of mutually engaging interaction:

- **Mutual awareness** of actions. In normal conversation we are aware of who is contributing what by virtue of our co-location. We propose that awareness mechanisms (Gutwin and Greenberg, 2002) are important to the emergence of mutual engagement; for example, highlighting when new contributions to the joint product occur, and indicating who made contributions. We use the term *mutual awareness* to distinguish it from workspace awareness (Gutwin and Greenberg, 2002) as we are interested in awareness mechanisms which focus on creative interaction between people rather than workplace management. Awareness pertains to: who is contributing (i.e. representation of the identity of the contributor), what they are contributing (i.e. representation of what kind of contribution it is, and what its content is), and where they are contributing it (i.e. where in the shared, but possibly not visible, space and time are contributions happening).
- **Shared and consistent** representations. In conversation we share the same aural space, and to some extent the same visual space. This consistency should be retained in collaborative systems, for instance by ensuring that all participants' views on the joint product are the same. Similarly, Robertson (1997) identified the consistent nature of representations as a key aspect of shared activities in a shared physical space. We would expect that participants would find it easier to understand the state of the joint product, and the effect of their own and others' contributions when the representations are shared and consistent. The importance of both mutual awareness and shared and consistent representations is illustrated by similar key elements in other research aimed at supporting social interaction such as social translucence (Erikson and Kellogg, 2000).
- **Mutual modifiability**. From previous studies of collaborative tool use (Bryan-Kinns et al, 2004, 2007a, 2007b) it was clear to us that being able to modify each others' contributions is important for mutual engagement. Mutual modifiability implies an egalitarian approach to role assignment within the tool rather than explicitly enforcing role mechanisms – in such an approach participants co-ordinate their activity in a subtle and dynamic manner (cf. Dourish and Bellotti, 1992).
- **Annotation**. Being able to converse in and around a shared product (Churchill et al., 2000) and make references to aspects of the joint product (Dourish and Bellotti, 1992) has been shown to be beneficial for work oriented collaboration. We propose that similar mechanisms would contribute to supporting mutual engagement in collaborative creativity.

4 Designing a Collaborative Music Tool

We have developed a collaborative music tool referred to as Daisyphone (Bryan-Kinns, 2004) which provides a means of investigating user interface design features in the novel domain of remote group music making. Daisyphone's design

was informed by our mutual engagement design features outlined in the previous section, and acts as test-bed for understanding design for mutually engaging collaboration. In this section we provide a brief overview of its design and interaction, and highlight how the design features have been realized.

4.1 Daisyphone

Aside from the use of Daisyphone in this paper and studies reported elsewhere (Bryan-Kinns, 2004a, Bryan-Kinns et al., 2006, 2007a), Daisyphone has been publicly available for use over the internet since its launch on 25 Oct 2003 (Marks, 2003), and logs of ongoing public use are regularly analysed.

The underlying infrastructure of Daisyphone was inspired by WebDrum II's (Burk, 2000) use of a client-server architecture to share music and graphics. As such, Daisyphone provides a low bandwidth, semi-synchronous form of collaboration based around short (5 seconds; 48 beats), looping pieces of music. The client-server architecture ensures that each participants sees the same **shared and consistent** representation of the music and annotations. Daisyphone is built using Java to allow it to be accessed on a range of devices from desktop to handheld computers (Bryan-Kinns, 2004b). In this section we outline the representation and interaction with music, and support for remote collaboration.

The design of Daisyphone combines a circular representation of a loop of music with a moving play head as illustrated in figure 1. Notes are arranged clockwise around the circle, and a playhead (the grey straight line) rotates clockwise around the circle playing the notes it passes over. Notes are placed and removed by clicking the small circles, and participants can remove each others' notes, thus supporting **mutual modifiability**.

The pitch of notes decreases with distance from the centre along the 48 *spokes*, and the 12 note pitches have been selected to sound harmonious. Four different musical sounds are provided and represented by the square, round, diamond, and triangle shapes which players select by clicking on the central stamen of the daisy (this is a modal operation). Saturation of color represents the volume of the note, and hue indicates who contributed it - each player is assigned a unique hue when they join a Daisyphone session **supporting quasi-awareness of identity of contributions** (players are aware that there are a number of other participants, but not necessarily who they are). Participants can **annotate** the screen simply by clicking and dragging the mouse to create graphical lines which are shared between all participants. We chose graphical annotation over text or audio based methods as it provides an additional channel of communication whilst keeping the interaction focus on the mouse and screen (rather than typing), and not introduce a competing audio source. Moreover, it intuitively supports localization and ad hoc social interaction such as associating names with colours.

5 Study

In terms of our design features, cues to identity form a core part of providing mutual awareness of action – participants need to be aware of both who is contributing, and what they are contributing. In previous studies (Bryan-Kinns et al., 2004a, 2004b, 2006, 2007a) cues to the identity of other participants was repeatedly identified as an important design feature. For example, Bryan-Kinns and Healey (2007a) showed that providing cues to identity significantly affected participants' contributions to collaborative music making. Similarly, work such as Gutwin and Greenberg (2002) stress the importance of identity in collaboration with the 'who' category forming a key part of their framework for understanding workspace awareness. Conversely, Applegate et al. (1986) suggest that allowing anonymous contributions to a creative activity would increase the productivity of the group. We wanted to explore how changing cues to identity affected the mutual engagement of participants.

We also believe that annotation mechanisms are important to mutually engaging interaction as they provide a channel through which to critique and reflect on the group activity, as well as to conduct more light-hearted social interactions. As such we wanted to explore what effect providing no communication mechanisms (i.e. graphical annotation) beyond the shared music would have on participants' mutual engagement.

These two design features lead to two hypotheses:

- **H1**: Mutual engagement would be greater where participants had explicit cues to identity.
- **H2**: Mutual engagement would be greater where an additional channel of communication was provided - graphical annotation.

5.1 Independent Variables

Two independent variables were manipulated:

- A within-subjects factor of **Identity** (ID vs. No ID). In the ID condition, participants' contributions were distinguished by hue (each participant was assigned a unique hue), whereas in the No ID condition, all participants' contributions were grey.
- A between-subjects factor of **Annotation** (Annotation vs. No Annotation). In the Annotation condition, participants could 'draw' on the Daisyphone, and these graphical annotations were shared with other participants. In the No Annotation condition, no graphical annotation was supported, and so communication could only occur through the music.

5.2 Dependent Variables

In order to identify mutually engaging interaction we developed a number of dependent variables derived from the indicators of points of mutual engagement

outlined at the start of this paper. These are grouped into Participant reports, Content assessments, and Interaction assessments outlined below, and detailed in the following sections.

Participant Reports provide measures of participants' subjective experience of the collaboration. As mutual engagement is a subjective experience, we would expect such self-reporting to provide the strongest indicator of mutual engagement:

- **Quality** measure: participants' reports of their assessment of the quality of the final product and the collaboration itself.
- **Preference** measure: participants' reported preferences for different conditions.

Content Assessments provide measures of the quality of the collaboration and final group product:

- **Musicality** measure: we would expect more evidence of musicality in the final product when participants were mutually engaged.
- **Communication** measure: analysis of the topics of annotations. Where participants are mutually engaged we would expect evidence of discussion of the joint product and each others' contributions, e.g. critique of the quality of the joint product.
- **Attunement** measure: analysis of whether participants were able to appreciate, and make changes to others' contributed patterns. When participants are mutually engaged we would expect them to start to mimic each others' contributions.

Interaction Assessment gives objective measures of the interaction itself.

- **Contribution** to joint production measure: number of notes contributed. We would expect more contributions where participants are more mutually engaged.
- **Mutual Modification** measure: number of deletions of participants' own notes, and other participants' notes. In mutually engaging interaction we would expect to see evidence of participants modifying each others' contributions.
- **Proximal** interaction measure: closeness of participants' contributions to each others' contributions. When there is mutual engagement, we would expect to see participants' contributions closer together, indicating a willingness to work together, as opposed to working away from each other in their own areas.

5.2.1 Participant Reports

We developed a questionnaire to identify participants' subjective assessment of the interface and their experience of the collaboration. First we asked participants for two ratings of the **quality** of each session - how much they felt they

contributed versus the group contributed, and how well they thought the group collaborated.

We also asked participants to compare the two sessions so that we could understand which session they had a **preference** for. Our previous experience of asking for ratings of interfaces indicated that ratings per interface (for example, ratings of how much a participant said they enjoyed a particular interface) do not yield reliable differences between interface designs (Bryan-Kinns et al, 2007a). We propose that this was not because there were no differences, but because the measures were not sensitive enough, or the wordings of the ratings did not elicit a strong enough reaction. We propose that this is especially true for interfaces which aim to support creative, aesthetic, and exploratory activities such as music making where preference for an interface may be subjective and conflated with the quality of the experience and end product. As a result of this, our questionnaire in this study focused on asking participants to make a number of comparisons between the two interfaces they used, for example, indicating with which interface they felt most "out of control", rather than asking them to rate how out of control they felt for each interface. Our comparisons aimed to capture: a) satisfaction with the product; b) feelings of enjoyment or flow (cf. Csikszentmihalyi, 1991); c) sense of collaboration; d) usability. For each of the statements below, participants were asked to indicate which session they felt the statement was most applicable to. We developed particularly strongly worded comparisons in order to elicit strong responses from the participants.

a) "The best jingle was produced"
   "I felt satisfied with the result"

b) "I enjoyed myself the most"
   "I felt out of control"

c) "I felt most involved with the group"
   "I understood what was going on"
   "Other people ignored my contributions"

d) "The interface was frustrating"
   "The interface was most complex"

5.2.2 Content Assessments

The primary aim of collaborating through Daisyphone is to create pleasing compositions together. Clearly, assessing the 'pleasing nature' of a composition would be somewhat subjective based on personal preference, cultural background, experience, and taste (cf. Desmet, 2003). Instead we focus on the **musicality** of the composition - evidence that there was some musical intention

behind the contributions - some attempt at making *music*. This provides a less subjective assessment of its quality as we are assessing whether notes fit into an overall musical scheme for the piece, rather than whether we like it per se.

We developed four methods of judging musicality: the most powerful method is simply listening to the final composition; an experienced musician could also judge the musical score of the final composition; a judge with experience of Daisyphone could rate the image of the final Daisyphone; and we could judge overviews of activity in each session to identify whether the it is indicative of random contributions, or more focused attempts at music making.

We developed a coding scheme for analyzing the topics of participants' **communication** through annotations based on previous coding schemes of collaborative interaction developed by Olsen et al. (1993), Applegate et al. (1986), and Bryan-Kinns et al. (2007b). The purpose of the coding scheme is to identify differences in the content of annotations between conditions. When participants are mutually engaged we would expect them to discuss the quality of the joint product. Our coding schemes categorizes topics of annotations as follows:

System related - where participants discuss technical problems with the Daisyphone system itself, or the experimental set up e.g. "Did it crash?".

Presence and Identity - statements about people's presence and their identity within Daisyphone e.g. "Me Zaki".

Query presence and identity - questions about participants' identity and presence in Daisyphone e.g. "who's the blue" or "Who is this, IDENTIFY YOURSELF"

Quality Judgment - comments about the quality of the music produced e.g. "nice!", or "it's completely random"

Task organization - discussion about the process of completing the task e.g. "Let's bang it out", or "Next time choose a section"

Social - non-task related discussion e.g. "haha", or smileys

We developed a coding scheme for differentiating participants' **attunement** to each others' actions - evidence of sensitivity to the actions and intentions of others in the group. We identified four broad levels of attunement from acknowledgement, to mirroring, transformation, and complementing others' contributions. Evidence of complementary contributions indicates the highest level of mutual engagement. Assessing attunement can only be achieved by listening to the jingle as it emerges, and observing the placement of contributions. It is critical to distinguish between Acknowledgement (placing notes taking into account others' contributions) and Complementing (placing notes to add to other's contributions). This is most reliably differentiated by listening to the effect of the new contribution - if it adds *musically* to someone else's contributions then it is complementary, if it is merely placed with spatial awareness (e.g. by dividing the composition area into different segments) then it is acknowledgement.

- **Acknowledgement**. With the lowest level of mutual engagement, participants show that they are aware of the contributions of another. This indicates a very basic, logistical, level of mutual engagement. In Daisyphone this would be illustrated by contributions which take into account other people's already present contributions by, for example, not writing over their notes. Note that we regard this as a low level of mutual engagement as the contributions are not musically integrated with each other - the result is not a co-ordinated musical piece, but at least it is not anarchy.
- **Mirroring**. Higher levels of mutual engagement are indicated by participants mirroring, or mimicing, others' contributions thus demonstrating that they themselves are able to produce it. This shows a level of mutual engagement in that participants are able to appreciate and reproduce others' contributions. In Daisyphone we could identify this when musical patterns are repeated verbatim around the Daisyphone. Intuitively, an example of this in conventional music making would be a call-and-response pattern of improvisation where one musician plays a musical motif which is repeated by another. The musical scales and sounds used in Daisyphone mean that patterns can also be repeated to create chords, and progressions of others' musical motifs as part of the joint composition, not only call-and-response patterns.
- **Transforming**. Building upon someone else's tune (taking it and transforming it to something new) indicates a higher level of mutual engagement as it involves an appreciation of others' contributions, and an ability to make changes to the form. In Daisyphone this would be indicated by repetition of musical patterns with some musical modifications.
- **Complementing**. Adding notes with musical intent to a phrase already formed by another participant indicates the highest level of mutual engagement - an ability to appreciate the form and make additions in-situ. In Daisyphone adding to, or augmenting someone else's tune musically would indicate complementing behavior.

5.2.3 Interaction Assessment

We developed several measures of interaction in the collaboration.

The **contribution** measure indicates how active a participant is simply by counting the number of notes they contributed.

The amount of **mutual modification** is indicated by the number of other people's notes a participant deletes.

Promixal interaction indicates how close a participant places their notes to other notes (i.e. co-location). We also developed three measures of proximal interaction for each new note contributed. Each of these is measured with respect to notes already contributed by the participant (referred to as **Self** notes), and notes contributed by other participants (referred to as **Other** notes):

1) **Number** of notes already on the same spoke.

2) **Distance** to the nearest note on the same spoke.
3) **Proximity** of nearest note not on the same spoke.

For example, in figure 2 a new note A is added by the blue participant using a circular note. There are 2 notes by that participant already on the same spoke, and no notes by anyone else. The *distance* to the nearest note on the same spoke is 3 notes (illustrated by line 2). The *proximity* of the nearest note contributed by this participant, but not on the same spoke is 2 notes away (line 1). The proximity of the nearest note contributed by another participant is 24 notes away (line 3). For the new note B added by the green participant using triangles, there are no other notes on the same spoke, the proximity of the nearest Self note is 4 (line 5), and the proximity of the nearest Other note is 17 (line 4).

5.3 Participants

Final year Computer Science students at the first author's institution were recruited through advertisements to take part in the experiment as part of their course, but not offered any incentives to take part. 39 of a possible 80 participants took part (28 males, 11 females; aged from: 20 to 29; mean age: 22, average computer literacy: expert; average musical ability: intermediate; none were professional or trained musicians; none had used Daisyphone before). Participants' musical preferences ranged from Hip Hop (most popular) to Latin (least popular) as illustrated in figure 3. All participants were assigned randomly to groups of three participants.

5.4 Procedure

The study took place in our lab with each participant in the group of three physically separated from each other so that they could not see or hear other participants. Each participant sat at a PC running Daisyphone and wore headphones, and a facilitator managed the study and had access to an additional PC on which they could view the interaction.

An introduction lasted 15 minutes in which participants were briefed that they were trialing the Daisyphone software in order to see how well it works in different situations. Additionally they were informed that the software records all actions using made using the interface, and that this data would be used for research purposes. They were then asked to sign consent forms for subsequent use of software logs, and questionnaire responses. They were also asked to complete a demographic questionnaire. All data was held anonymously.

Participants were instructed that their tasks were to work together as a group to jointly remotely compose two jingles for the Olympic games in 2012 (the games will be located quite close to the first author's institution). They were given instructions and free use of Daisyphone in a 15 minute individual acclimatization session. They then worked for up to 20 minutes on each composition – from previous studies we found that 20 minutes was typically the maximum time

people would spend creating one short loop (5 seconds) in Daisyphone. Participants were told that they could stop the task at any point. They were also told that their interaction would be judged for collaboration and that they should aim to collaborate effectively to create the tune (as we had previously found this motivation to have a positive effect on the collaboration; Bryan-Kinns et al., 2007a).

In the study, ID (the presence or absence of colour to indicate identity) was randomly changed between sessions, and Annotation (whether participants had the ability to contribute graphical annotations) was randomly assigned to half of the groups. The participants were not told which conditions they were in. After both joint remote composition sessions had taken place, a post-task questionnaire composed of comparisons, ratings of the participants' perceptions of their experience, and open questions, was completed individually by the participants for up to 10 minutes. These questionnaires were aimed at identifying whether participants experienced mutual engagement in their interaction.

6 Study Results

All participants undertook the sessions for the full 20 minutes each session. The following section details results by dependent variable.

6.1 Participant Reports

These provide measures of participant's subjective, self reported experience of the collaboration.

6.1.1 Participant Quality Measure

When asked, participants did not rate their contributions to the jingle differently to the rest of the group's contribution for any of the four conditions (mean rating of approximately 3 'About equal' on a 5 point Likert scale). Table 2 and figure 5 illustrate that there were also no differences in participants' ratings of how well they thought the group collaborated together when they had No Annotation. When participants had Annotation, there was a significantly higher rating of the group's collaboration on a 4 point Likert scale when they had ID (Wilcoxon Signed-Rank test; $W = 96$; $z = 3$; $p = 0.0027$) in comparison to when they had Annotation and No ID.

6.1.2 Participant Preference Measure

A significant number of participants (31 of 39) reported that they noticed a difference between the two interfaces ($Chi^2_{(1)} = 12.42$; $p = 0.0004$).

Table 1 gives details of the results of the post task questionnaire comparison questions where participants were asked to compare the two interfaces they used (ID and No ID). Numbers of agreements with the comparative statements for ID

and No ID interfaces and the significance of the results when a Chi^2 test was applied are given (significant differences are highlighted in bold; $p < 0.05$). For example, when participants had Annotation, 15 participants said that they made the best jingle when they had ID versus 5 participants saying that they made the best jingle when they had No ID, and this was a significant difference which we interpret as a significant preference for the ID condition when they had Annotation. These results are also summarized in figure 4 where percentages of agreements with statements for the ID interface are given. The results indicate that there was an interaction effect between Annotation and ID; when participants had No Annotation, there were no significant preferences for the ID or No ID interfaces, when they had Annotation there were significant preferences for the ID condition except for *complexity of interface* and *frustration with the interface* where there were no significant differences in preferences.

We also examined the textual rationale given for the preference measures of "The best jingle was produced" and "I felt satisfied with the result". For each participant these responses were either concerned with the quality of the composition (e.g. “Ended up with a coherent tune”), or the quality of the collaboration (e.g. “More contribution and co-operation amongst participants”).

6.2 Content Assessments

These provide measures of the quality of the collaboration and the final group product.

6.2.1 Musicality Measure

Two independent judges listened to each jingle and rated them for evidence of musicality on a 5 point Likert scale form “Not musical at all” to “A collaborative effort where music fitted together as a whole”. Table 3 and figure 6 illustrate the average ratings for the two judges. The average rating of jingles was significantly higher for participants with Annotation (mean: 3.78) than No Annotation (mean: 2.66) (Wilcoxon Mann-Whitney; $U_A = 125.5$; $z = -2.11$; $p = 0.0174$) (Cohen’s kappa: 0.75). The mode of ratings when participants had Annotation was 4, and when they had No Annotation was 2.

Figure 7 illustrates the difference between musical and non-musical jingles’ scores where each stave represents one particular instrument sound in Daisyphone (circle, square, triangle, diamond). Visually inspecting these three scores it is clear that figure 7a is a musical jingle as there are a limited number of notes played across the different instruments, and there is some repetition of musical motifs.

Figures 7b and 7c both illustrate non-musical jingles which have large numbers of discordant notes spread across multiple instruments, little repetition of musical phrases, and no co-ordination between instruments. With Annotation, 13 of 14 compositions’ scores were judged to look musical. Where participants had

No Annotation, 5 of 12 compositions' scores were judged to look musical. We visually inspected the final compositions in Daisyphone and found the same results. Figure 11 illustrates the final compositions created by each group in Daisyphone.

Finally, we examined overviews of the total interaction in each session as illustrated in Figure 12 which gives an overview of each session. In these overviews, the size of the circles is proportional to how often a user contributed a note at that point. Comparing these overviews to the final compositions, we can identify evidence of musicality. For example, figure 8a illustrates an overview of one session whose final jingle is illustrated in figure 8b.

We propose that where participants were engaged with the music the overview is more evenly distributed as participants contribute notes and then attempt to edit them into the correct position (e.g. Figure 8a) allowing for pauses between notes, rather than contributing many shapes and drawings which would tend to fill the whole overview and not sound musical as illustrated in Figure 9. From examining the overviews, ten of 14 jingles were judged to be musical when participants had Annotation, verses two of 12 when participants had No Annotation.

In addition to helping us identify musical interaction, the overviews provide indications of evidence of proximal interaction and mutual modification which are also captured by measures of the Interaction discussed later in this section.

It is worth noting that as the participants had no fixed starting location or orientation, it was not possible to statistically analyse the spatial spread of contributions as, for example, undertaken by Pinelle et al.'s (2008) analysis of the territoriality of tabletop use by seated participants.

6.2.2 Communication Measure

There was no significant difference in the number of participants' textual contributions per group between ID (mean: 14.7) and No ID (mean: 12.5). Nor were there any significant differences in the number of letters in each contribution made by participants between ID (mean: 8.2 letters) and No ID (mean: 10.1 letters).

There were also no significant differences in the proportions of topics of annotations made by participants, but there were trends as illustrated in figure 13 and table 4. In particular, we see that Task organization was a more frequent topic when participants had No ID (mean: 28%) as opposed to ID (mean: 21%). Conversely, there was more discussion of the quality of the music produced when participants had ID (mean: 18%) then when they had No ID (mean: 12%). Social interaction occupied a similar amount of discussion for ID (mean: 27%) and No ID (mean: 26%).

6.2.3 Attunement Measure

We examined each session for evidence of the different forms of attunement by stepping through the interaction using our log playback tool and judging each new contribution using our coding scheme. We found that both Annotation and No Annotation interfaces had similar proportions of sessions showing evidence of Acknowledgement, Mirroring, and Transformations. However, there were proportionally twice as many sessions which showed evidence of complementing where participants had Annotation (9 of 14; 65%) versus No Annotation (4 of 12; 33%). There was no difference between sessions with ID and No ID. Overall, the number of instances of Acknowledgement, Mirroring, Transformation, and Complementing are too small to provide any statistically significant results.

In the rest of this section we illustrate some of the identified examples of attunement.

Acknowledgement. Figure 14 illustrates a session where participants acknowledged each others' contributions by keeping their contributions apart from each other (but they were not judged to be musically complementary). Most sessions showed some evidence of Acknowledgement.

Mirroring. Figure 15 illustrates an example of mirroring highlighted by black ellipses where the participant using the diamond shapes has mirrored the another participant's contributions.

Transformation. Figure 16 illustrates the light grey participant copying the dark grey participant's pattern of a descending series of notes in a modified manner - they remove the second note in the pattern, and move the first notes to a lower pitch. They also use two different instruments (triangle and diamond shape) rather than one consistent shape through the whole pattern as the dark grey participant did.

Complementing. Figure 17a illustrates a situation in which the light grey participant laid out three triangles in sequence which the darker participant has musically complemented by adding triangles just before two of the notes. In figure 17b, the dark grey participant had contributed a sequence of triangles around the daisy which the light grey participant has complemented with two diamonds and a triangle which musically complement the previous notes.

6.3 Interaction Assessment

These give objective measures of the interaction between participants.

6.3.1 Contribution Measure

As table 5 and figure 18 illustrate, there were significantly more notes contributed by individual participants when they had No Annotation mechanisms compared to when they had Annotation mechanisms (Wilcoxon Mann-Whitney; $U_A = 331$; $z = -3.99$; $p < 0.0001$). Figure 19 illustrates that there was a non significant interaction between Annotation and ID as there was a trend that

participants with No Annotation contributed more notes when they had No ID compared to when they had ID.

There were also significantly more notes contributed by groups (3 participants) who had No Annotation mechanisms (Wilcoxon Mann-Whitney; $U_A = 0$; $z = 2.76$; $p = 0.0029$). Moreover, the overall number of notes remaining at the end of the sessions (number of notes minus number of deletions) per group was significantly higher for those with No Annotation mechanisms (Wilcoxon Mann-Whitney; $U_A = 31$; $z = -2.11$; $p = 0.0174$).

Comparing participants with ID and no ID, there were no significant differences for the overall mean number of notes, nor the number of notes contributed by individual participants, nor the number of notes contributed by groups. Also, there was no significant difference between the overall number of notes remaining for each group at the end of each session (i.e. the 'size' of their jingles) for those with and without ID.

6.3.2 Mutual Modification Measure

As table 6 and figure 20 illustrate, individual participants who had No Annotation mechanisms deleted significantly more notes (Wilcoxon Mann-Whitney; $U_A = 316$; $z = -3.56$; $p = 0.0002$). However, there was no significant difference between the total number of notes deleted by groups. As illustrated in table 7 and figure 21, there were significantly more Self Deletions than Other Deletions made by each participant in the Annotation condition (Wilcoxon Signed-Rank test; $W = 101$; $z = 1.75$; $p = 0.0401$), but no significant difference for the No Annotation condition. There were no significant differences in the overall ratios of Self:Other deletions, nor Notes:Deletions for any of the conditions. For participants with No Annotation, the ratio of Self:Other deletions was significantly higher when participants had ID i.e. there were significantly more Self than Other deletions (Wilcoxon Signed-Rank test; $W = 77$; $z = 1.81$; $p = 0.0351$).

There was a trend that participants with No ID made more deletions than those with ID (Wilcoxon Signed-Rank test; $W = -201$; $z = -1.45$; $p = 0.0735$).

The graphical overviews outlined previously also provide a general indication of participants' editing activities - the larger the circles, the more notes contributed at that point (and so, by implication, deleted as there can only be one note at any one position). It is useful to compare the final composition to the overview to identify how much work actually contributed to the final piece. For example, in figure 10c it is evident that a lot of work was carried out by the purple participant in the centre of the piece, but this was not evident in the final composition illustrated in figure 10d. However, the Mutual Modification measures discussed above provides a more focused analysis of participants' editing activities than this subjective visual interpretation.

6.3.3 Proximal Interaction Measure

Figure 22 and table 8 illustrate that when participants contributed notes in the No Annotation condition, there were significantly more notes already contributed by the participant on the spoke than when they were in the Annotation condition (Wilcoxon Mann-Whitney; $U_A = 351$; $z = 3.79$; $p = 0.0001$). Similarly, there were significantly more Other participants' notes on the spoke contributed to in the No Annotation condition. There were no significant differences between ID and No ID conditions.

Table 9 and Figure 23 illustrate that whilst there is not a significant difference between the number of notes already on a spoke when a note is contributed, there is a trend that when there is No ID, there are more Other participants' notes on the spoke (Wilcoxon Signed-Rank test; $W = 187$; $z = 1.35$; $p = 0.0885$). For participants with Annotation, there was also a trend for participants with ID to have more Self notes on spokes they contributed to (Wilcoxon Signed-Rank test; $W = -68$; $z = -1.36$; $p = 0.0869$).

Figure 24 and table 10 illustrate that there was no significant difference in the Distance to notes already on spokes when participants made contributions. When participants had Annotation, the Proximity of notes previously placed by participants (Proximity of nearest Self note) was significantly higher than when they had No Annotation (Wilcoxon Mann-Whitney; $U_A = 1046$; $z = -3.46$; $p = 0.0003$) i.e. they placed their notes further away from their own notes when they had Annotation. Similarly, Proximity of Other people's notes was significantly larger when participants had Annotation (Wilcoxon Mann-Whitney; $U_A = 1125.5$; $z = -4.29$; $p < 0.0001$) i.e. they placed notes further away from Other people's notes when they had Annotation.

The graphical overviews discussed previously illustrate the proximal co-location of participants' contributions, and whether this was consistent throughout the session. For example, the overview in Figure 10a indicates that the blue and green participants occupied two halves of the composition, whereas the red overlapped with them slightly. Similarly, in figure 10b there is a strong separation between all 3 participants who take a third of the composition each. The overviews also illustrate separation around the composition (not just sections). For instance, figure 10c shows that the purple participant mostly worked in mid range of the composition, whereas the green predominantly worked on the edges. There are examples of both kinds of spatial separation for all participants regardless of whether they had Annotation or not. Analyzing such overviews gives a more subjective measure of proximal interaction than the analysis of interaction reported above.

6.4 Summary of Results

In summary, the following **significant observations** were made for participants who had **Annotation** mechanisms:

- They had **higher rated jingles** than those with No Annotation mechanisms.
- On inspection, more jingles were judged to be **musical** when participants had Annotation.
- Participants expressed a **significant preference for the ID condition** when they had Annotation, but no preference when they had No Annotation i.e. could not graphically annotate. This included ratings of how well they felt the **group collaborated**, **satisfaction** with the product, **enjoyment**, feeling of **control**, **understanding** of what was happening, and perception that **others were taking notice** of their contributions.
- They contributed **fewer notes** than those with No Annotation mechanisms.
- They **deleted fewer** notes than those with No Annotation mechanisms.
- They **deleted more of their own** (Self) notes than Other participants' notes. This was not the case for participants who had No Annotation mechanisms.
- Their final **tunes were smaller** (contained less notes).
- The **spokes contained fewer notes** of their own (Self) and Other when contributions were made.
- Their **contributions were placed further away** from their own (Self) contributions, and Others' contributions.

In addition, we found the following **trends** which were not significant.

- Where participants had the ability to **Annotate and had ID**, they made **fewer annotations about task organization**, and **more annotations about the quality of the music produced** than when they had No ID.
- There were **more** sessions showing evidence of **complementary** contributions by participants with **Annotation**.
- Participants with **No ID made more deletions** than those with ID.
- When participants contributed with **No ID** there were **more Other people's notes on spokes** they contributed to than participants with ID.
- Participants had **more Self notes on spokes** they contributed to when they had **Annotation and ID** as opposed to Annotation and No ID.
- Finally, participants with **No ID contributed more notes** in total than those with ID, but this was not significant.

7 Discussion

The questionnaire results show that participants had a strong preference for the condition where they had Annotation and ID. Annotation was the primary factor, and ID the secondary factor. Furthermore, analysis of the musicality of final compositions and participants' collaboration indicates that participants who had Annotation produced the most musical and coherent jingles. Analysis of participants' annotations gives some indication that when participants had ID they focused on the music they were producing rather than organizing the task itself. Drawing from these three observations we propose that participants who had **Annotation and ID were most mutually engaged**. This supports hypotheses H2

that where participants had annotation they were most mutually engaged. As ID was a secondary factor, this weakly supported H1 - that where participants had cues to identity they were most mutually engaged. Possibly the provision of annotation mechanisms which allow social interaction such as writing one's own name increases a sense of mutual engagement, but, as we discuss below, there are many other factors which contribute to mutual engagement. Further work should explore other possible explanations for our results such as whether part of the reason for increased mutual engagement when graphical annotation was provided was an increased sense of accountability (cf. Erikson and Kellogg, 2000) that could be developed through the graphical annotations. For example, there has been some evidence in this, and previous studies, of participants writing their names around the Daisyphone (cf. Bryan-Kinns, 2004a). This could contribute to an increased sense of accountability, which may contribute to the mutual engagement. Alternatively, the tagging of contributions could be an indicator of a person's pride in their contribution - an engagement with the product itself, which may in turn contribute to mutual engagement. However, this naming, or 'tagging' of contributions was not consistent across sessions, and was not significantly different when participants did or did not have ID.

In terms of collaborative interaction, participants with Annotation and ID were significantly less active (as indicated by note addition and deletion) than others, and produced jingles with significantly fewer notes. This indicates that increased contributions and activity does not mean increased mutual engagement, nor better products (in this case, music). This was not what we expected, but in our examination of the interaction, judging of quality of interaction and final product, and participant reports of preference, it is clear that increased contributions actually indicated boredom, or lack of engagement with the activity and other members of the group. Moreover, the fact that final jingles had significantly more notes when participants had No Annotation, and there was significantly more activity indicates that participants were simply contributing more but we suggest that the interaction was *less focused*. So, in this study, **fewer contributions indicates more focused and mutually engaging interaction**. One possible explanation for this can be drawn from theories of *flow* (Csikszentmihalyi, 1991) where rich and rewarding activities include periods of reflection, observation, and contemplation of contributions rather than mechanically and repeatedly making contributions. Such a theory may be used to suggest that there would be less time spent on contributing, and more on contemplation in mutually engaging interactions. Alternatively, theories of conversational interaction, and in particular the establishment of *common ground* in conversation (Clark, 1996) where conversational effort expended in order to establish mutual belief about situations, may help to explain our findings. We could, for example, consider the musical interaction as a conversation in which increased contributions indicated a difficulty in establishing a musical common ground between participants. Indeed,

we could view the music making activity as rather ambiguous and therefore *tightly coupled* in which 'the more common ground the participants have, the less interaction is required to understand the situation and what to do' (Olson and Olson, 2000). By implication, this explanation suggests that increased contributions indicate a lack of common ground between participants. This could provide an explanation for the increased sense of mutual engagement when there were fewer contributions. An alternative explanation may be that when participants have Annotation and ID, and so are more accountable for their actions, they are less likely to experiment with the product. However, the fact that participants with Annotation and ID were consistently judged to have produced higher quality products does not support this line of argument. Further research is needed to explore how these models and theories could be used to inform design of systems intended to support mutual engagement.

The observation that participants deleted proportionally more of their own notes than others' notes when they had Annotation indicates that they were not involved in as much mutual modification as those with No Annotation. However, given that there were significantly fewer tuneful jingles produced by participants with No Annotation this indicates again that whilst they may have been deleting more of other participants' contributions, they were not actually engaged with the product at hand. Creating a coherent, tuneful jingle, is, in this case, a stronger indicator of mutual engagement than simply deleting other participants' notes. There is a tension here between our dependent measures of mutual modification on the one hand, and ratings of preference, content, and attunement on the other. In this case, we conclude that **mutual modification does not imply mutual engagement**.

Participants with Annotation had significantly fewer notes on spokes they contributed to, and their contributions were placed further away from other notes than participants with No Annotation. Given the significantly higher number of notes in the final jingles of participants with No Annotation, this difference is simply a result of there being more notes on the Daisyphone. Therefore we conclude that such **measures of proximity are only applicable when there are similar numbers of contributions**. For example, there was a trend for participants to have more Self notes on spokes they contributed to when they had Annotation and ID as opposed to Annotation and No ID. In this comparison there were fewer notes contributed in the Annotation and ID condition, so here the larger number of Self notes on spokes contributed to indicates **more frequent focused interaction with their own notes when they had Annotation and ID**.

The results of this study help us understand what mutual engagement is and how we may identify mutually engaging interaction. For mutual engagement using Daisyphone we are looking for evidence of both of the following:

Being engaged with the music being jointly produced. Examining the final jingle and the contributions over time gives us an indication of the **musicality**

of the interaction. The **questionnaires** also provide an indication of which condition participants felt most engaged with the music. Also, examining the topics of **annotations** provides an indication of engagement with the music when participants discuss its quality. A **decreased** amount of **activity** seems to indicate increased focus and engagement with the music in this study. However, whilst increased activity seemed to indicate lack of engagement as exemplified by non-musical jingles, determining when participants are contributing so little that they are no longer engaged is a topic for future research.

Being engaged with each other. **Questionnaires** provide an indication of which conditions participants felt engaged with each other - where others did not ignore their contributions, and they felt involved with the group. Logs of interaction can be analysed to identify points where participants compliment each others' contributions (**attunement)** which indicates mutual engagement. By implication, ratings of final compositions also gives an indication of how well participants engaged with each other. In this study, analysis of the interaction did not provide indicators of engagement with others - the **proximity** measures were essentially swamped by the varying numbers of contributions. Also, examining ratios of Self to Other **deletions** did not yield conclusive results - we would have expected people who were mutually modifying contributions to delete proportionally more Other notes than Self notes.

A key question with respect to this study is the applicability of the results. We believe that this study has shed light on which metrics are reliable for the task of collaborative music making of short loops. However, there are several features of music making which differentiate it from other creative collaborative activities which will impact the generalisability of our approach. The key difference is the temporal dimension of music. For a loop of music, the most important judgement of quality is whether it sounds *good* - musically coherent over its temporal length. As such, creating a coherent musical piece over the extent of the loop is a stronger indication of mutual engagement than spatial co-location or mutual modification. Specifically, for the piece to be coherent participants must understand and respond to each others' contributions over the full length of the piece, not simply co-locate their notes. Therefore, for musical tasks, we propose that identifying points of Transformation and Complementary contributions gives a stronger indication of mutual engagement than spatial co-location. For other, spatially oriented, domains such as brainstorming or collaborative sketching, this may not be the case. A key concern with focusing on identifying points of attunement is the analytic overhead of analyzing each contribution to judge whether it is Acknowledgement, Mirroring, Transformation, or Complementary.

We have demonstrated that the applicability of the CSCW design features of awareness and localized annotation in the new domain of group music interaction.

Whilst the current study took place in a controlled environment, we would expect the results to be applicable to more creative improvisational interaction in naturalistic settings. For example, the design features and mutual engagement assessment tools proposed in this paper would certainly be applicable to designing ad hoc exertion interfaces for interactive performances (cf. Sheridan and Bryan-Kinns, 2008). They may also be applicable to co-located collaboration with multiple Daisyphones, or one shared interactive surface such as a tabletop.

We would also like to explore features of *dis*-engagement in creative activities, and what effect factors such as context, personality, social interaction, and so on have on collaboration.

Finally, it is worth mentioning that the findings and conclusions we presented here relate to design features for collaboration. We have not explicitly examined participants' engagement with Daisyphone itself. Whilst there was no significant difference between perceptions of the usability of the interface in different conditions (questionnaire: *frustrating* and *complex* interface), it should be noted that participants (especially with Annotation) were able to create musical jingles in a short period of time. We kept the musical nature of the interface constant across conditions to prevent the interface itself becoming a confounding variable in the study.

8 Conclusions

In this paper we defined mutual engagement and provided a series of measures which could be used to identify mutually engaging interaction. We exemplified the effect CSCW user interface design features had on participants' mutual engagement and showed that providing shared annotation mechanisms and awareness of identity increases mutual engagement. Importantly, in this study we identified that mutually engaging collaborations involve less activity, but more focused interaction. For group music making itself we propose that mutual modification is not as strong an indicator of mutual engagement as coherence and musicality of the final product. These results provide us with new insights into the design of more engaging collaborative systems. The general finding that less interaction may be more engaging will provide a different slant on collaborative system design. The finding that simply editing each others' contributions does not indicate engagement will help us to find better measures of engagement. We see the work as contributing directly to the CSCW and NIME fields of research, with a playful, yet powerful contribution to make to interaction design in general.

Future work will investigate at which point participants disengage from a collaboration through reduced interaction, and seek to generalize the results to other domains where the temporal dimension of music is not as important. We will also investigate the additional factors that may contribute to mutual engagement such as social context, prior experience, and group dynamics.

9 Acknowledgements

We thank the students and facilitators involved in this study. This work was supported by EPSRC grant Engaging Collaborations (GR/S81414/01). We also thank the reviewers for their useful critique of the content of this paper.

10 References

Applegate, L., Konsynski, B. R., and Nunamaker, J. F. (1986). A group decision support system for idea generation and issue analysis in organization planning. In *Proceedings of the 1986 ACM conference on Computer-supported cooperative work*, pp. 16-34.

Barbosa, A. (2003). Displaced soundscapes: A survey of network systems for music and sonic art creation. *Leonardo Music Journal,* Vol. 13, 53–59.

Benford, S., Bowers, J., Fahlen, L., Greenhalgh, C., and Snowdon, C. (1995). User Embodiment in Collaborative Virtual Environments. In *Proceedings of CHI 1995*, pp. 242-249.

Blaine, T. and Perkis, T. (2000). The jam-o-drum interactive music system: A study in interaction design. In *Proceedings of DIS 2000*, Brooklyn, New York, J. Karat and J. Thackara, Eds. ACM, New York, USA, pp. 165–173.

Blaine, T. and Fels, S. (2003). Contexts of collaborative musical experiences. In *Proceedings of the 2003 Conference on New Interfaces for Musical Expression (NIME-03),* Montreal, Canada, F. Thibault, Ed. Faculty of Music, McGill University, Montréal, Québec, Canada, pp. 129–134.

Blum, S. (1998). *Recognizing Improvisation*. University of Chicago Press, Chapter 1, 27–45.

Blythe, M. A., Monk. A. F., Overbeeke, K., and Wright, P. C. (2003). *Funology From Usability to Enjoyment*. Kluwer Academic Publishers, London, UK.

Bowers, J. (2002). *Improvising machines*. M.S. thesis, Masters in Music by Research, University of East Anglia, Norwich, UK.

Bryan-Kinns, N. (2004a). Daisyphone: The Design and Impact of a Novel Environment for Remote Group Music Improvisation. In *Proceedings of DIS 2004*, Boston, USA, pp. 135-144.

Bryan-Kinns, N. (2004b). Mobile group music improvisation. In *Proceedings of Engagability and Design 2004*, Birmingham, UK.

Bryan-Kinns, N., & Healey, P. G. T. (2006). Decay in Collaborative Music Making. In *Proceedings of NIME 2006*, Paris, France, pp. 114-117.

Bryan-Kinns, N., & Healey, P. G. T. (2007a). Exploring Mutual Engagement in Creative Collaborations. In *Proceedings of Creativity and Cognition 2007*, Washington, USA, pp. 223 - 232.

Bryan-Kinns, N., Healey, P. G. T., Papworth, D. and Vaduuva, A. (2007b). Cues to Mutual Knowledge. In *Proceedings of ECSCW 07*, Limmerick, 2007.

Burk, P. 2000. Jammin' on the web - a new client/server architecture for multi-user musical performance. In *Proceedings of International Computer Music Conference (ICMC 2000),* Berlin, Germany.

Churchill, E. F., Trevor, J., Bly, S., Nelson, L., and Cubranic, D. (2000). Anchored Conversations. Chatting in the Context of a Document. In *Proceedings of CHI 2000*, ACM Press, pp. 454-461.

Clark, H. H. and Brennan, S. E. (1991). Grounding in Communication, in L. B. Resnick, J. Levine, and S. D. Behrend (eds.): *Perspectives on Socially Shared Cognition*, American Psychological Association, pp. 127-149.

Clark, H. H. (1996). *Using Language*. Cambridge University Press, UK.

Costa, P., Duarte, P., and Costa, C. J. (2007). WebStorm: Mixing Brainstorming with Art in the Web. In *Proceedings of SIGDOC'07*, pp. 170-175.

Crabtree, A., Benford, S., Capra, M., Flintham, M., Drozd, A., Tandavanitj, N., Adams, M., and Farr, J. R. (2007). The Cooperative Work of Gaming: Orchestrating a Mobile SMS Game, *Computer Supported Cooperative Work*, 16, pp. 167–198.

Csikszentmihalyi, M. (1991). *Flow: The Psychology of Optimal Experience*. Harper Collins.

De Lucia, A., Francese, R., Passero, I., and Tortora, G. (2008). SLMeeting: supporting collaborative work in Second Life. In *Proceedings of the Working Conference on Advanced Visual interfaces AVI '08*, pp. 301-304.

Desmet, P. (2003). Measuring Emotion: Development and Application of an Instrument to Measure Emotional Responses to Products. In Blythe, M. A., Monk. A. F., Overbeeke, K., and Wright, P. C. (eds.) *Funology From Usability to Enjoyment*. Kluwer Academic Publishers, London, UK.

Dobrian, C., and Koppelman, D. (2006). The 'E' in NIME: Musical expression with new computer interfaces. In *Proceedings of New Interfaces for Musical Expression (NIME)*. IRCAM, Centre Pompidou, Paris, France, pp. 277–282.

Douglas, Y. and Hargadon, A. (2000). The pleasure principle: Immersion, engagement, flow. In *Proceedings of Hypertext 2000*, F. Shipman, Ed. ACM, New York, USA, pp. 153–160.

Dourish, P., and Bellotti, V. Awareness and Coordination in Shared Workspaces. In *Proceedings of ACM Conference on Computer-Supported Cooperative Work (CSCW'92),* 107-114. (1992).

Erikson, T., and Kellogg, W. A. (2000). Social Translucence: An Approach to Designing Systems that Support Social Processes. *ACM Transactions on Computer-Human Interaction (TOCHI) Vol. 7 , No.1*, pp. 59 - 83.

Fels, S. and Vogt, F. (2002). Tooka: Explorations of two person instruments. In *Proceedings of the 2002 Conference on New Instruments for Musical Expression (NIME-02)*, Dublin, Ireland.

Fels, S. (2004). Designing for intimacy: Creating new interfaces for musical expression. *Proceedings of the IEEE*, Vol. 92, No. 4, 672–685

Fischer, G. (2000). Symmetry of ignorance, social creativity, and meta-design. *Knowledge-Based Systems*, Vol. 13, Issues 7-8, 1 December 2000, pp. 527-537

Gutwin, C., and Greenberg, S. (2002). A Descriptive Framework of Workspace Awareness for Real-Time Groupware. *Computer Supported Cooperative Work*, Vol. 11, 411-446.

Hilliges, O., Terrenghi, L., Boring, S., Kim, D., Richter, H., and Butz, A. (2007). Designing for Collaborative Creative Problem Solving. In *Proceedings of Creativity and Cognition 2007*, pp. 137-146.

Hymes, C. M. and Olson, G. M. (1992). Unblocking brainstorming through the use of a simple group editor. In *Proceedings of the 1992 ACM Conference on Computer-Supported Cooperative Work CSCW '92*, pp. 99-106.

Ijsselsteijn, W., van Baren, J., Romero, N., and Markopoulos, P. (2003). The Unbearable Lightness of Being There: Contrasting Approaches to Presence Engineering. In *Proceedings of SPIE*, Vol. 5150.

Jordá, S. (2002). Improvising with computers: A personal survey (1989-2001). *Journal of New Music Research,* Vol. 31, No. 1.

Jordá, S. (2005). Multi-user instruments: Models, examples and promises. In *Proceedings of the 2005 International Conference on New Interfaces for Musical Expression (NIME05),* Vancouver, BC, Canada, pp. 23–26.

Johnson, H. and Hyde, J. (2003): Towards Modeling Individual and Collaborative Construction of Jigsaws Using Task Knowledge Structures (TKS), *ACM Transactions on CHI*, vol. 10, no. 4, December 2003, pp 339–387.

Marks, P. (2003). Will Jamming be the New Texting? *New Scientist, 180, 2418* (25 Oct 2003), p. 25.

Miell, D. and MacDonald, R. (2000). Children's creative collaborations: The importance of friendship when working together on a musical composition. *Social Development 36*, 348-369.

O'Hara, K., Kindberg, T., Glancy, M., Baptista, L., Sukumaran, B., Kahana, G., and Rowbotham, J. (2007). Collecting and Sharing Location-based Content on Mobile Phones in a Zoo Visitor Experience, *Computer Supported Cooperative Work*, Vol. 16, pp. 11–44.

Olson, J. S., Olson, G. M., Storrøsten, M., and Carter, M. (1993). Groupwork close up: a comparison of the group design process with and without a simple group editor. *ACM Transactions on Information Systems*, Vol. 11, No. 4, pp. 321-348.

Olson, G. M., and Olson, J. S. (2000). Distance Matters, *Human-Computer Interaction, Vol. 15, Issue 2*, pp. 139 - 178.

Pinelle, D., Nacenta, M., Gutwin, C., and Stach, T. (2008). The Effects of Co-Present Embodiments on Awareness and Collaboration in Tabletop Groupware. In *Proceedings of Graphics Interface Conference 2008*, Ontario, Canada, CHCCS/ SCDHM & ACM, pp. 1-8.

Poupyrev, I., Lyons, M. J., Fels, S., and Blaine, T. (2001). New interfaces for musical expression. In *CHI '01 extended abstracts on Human factors in computing systems*, Seattle, USA. M. M. Tremaine, Ed. ACM, New York, USA, pp. 491–492.

Robertson, T. Cooperative Work and Lived Cognition: A Taxonomy of Embodied Actions. In *Proceedings of ECSCW*, pp. 205-220.

Sall, A., and Grinter, R. E. (2007). Let's Get Physical! In, Out and Around the Gaming Circle of Physical Gaming at Home, *Computer Supported Cooperative Work*, Vol. 16, pp. 199–229.

Sawyer, K. (2003). *Group creativity: Music, theater, collaboration*. Lawrence Erlbaum Associates (LEA), NJ, USA.

Schiano, D. J., Elliott A., and Bellotti, V. (2007). A Look at Tokyo Youth at Leisure: Towards the Design of New Media to Support Leisure Outings, *Computer Supported Cooperative Work*, Vol. 16, pp. 45–73.

Sheridan, J. G. and Bryan-Kinns, N. (2008 in press). Designing for Performative Tangible Interaction. *International Journal of Arts and Technology. Special Issue on Tangible and Embedded Interaction*.

Smith, H., and Dean, R. (1997). *Improvisation, hypermedia, and the arts since 1945*. Harwood Academic Publishers, Amsterdam B. V.

Stowell, D., Robertson, A., Bryan-Kinns, N., & Plumbley, M. D. (in press). Evaluation of live human-computer music-making: quantitative and qualitative approaches, *International Journal of Human Computer Studies*.

The Technology Strategy Board (2009). *Creative Industries Technology Strategy 2009 - 2012*. Innovate UK, HM Government.

Terrenghi, L., Fritsche, T., and Butz, A. (2006). The EnLighTable: Design of Affordances to Support Collaborative Creativity. In *Proceedings of Smart Graphics 2006, Lecture Notes in Computer Science*, Vol. 4073, Springer Berlin / Heidelberg, pp. 206-217.

Titon, J. T. (1996). *World of Music*. Schrimer Books, New York, USA.

Tufnell, M. and Crickmay, C. (1990). *Body Space Image* , First ed. Dance Books, London, UK.

Wanderley, M. M., and Orio, N. (2002). Evaluation of input devices for musical expression: Borrowing tools from HCI. *Computer Music Journal*, Vol. 26, No. 3, 62–76.

Weinberg, G. and Gan, S.-L. (2001). The squeezables: Toward an expressive and interdependent multi-player musical instrument. *Computer Music Journal,* Vol. 25, No. 2, pp. 37–45.

Wenger, E. (1998). *Communities of Practice*. Cambridge University Press, Cambridge, UK.

11 Tables

Question	Condition	ID	No ID	$Chi^2_{df=1}$	p
The best Jingle					
	Annotation	**15**	5	4.06	**0.0439**
	No Annotation	10	6	0.56	0.4543
I felt most involved with the group					
	Annotation	**17**	3	8.45	**0.0037**
	No Annotation	9	5	0.64	0.4237
I enjoyment myself the most					
	Annotation	**17**	4	6.86	**0.0088**
	No Annotation	11	3	3.5	0.0614
I understood what was going on					
	Annotation	**15**	5	4.06	**0.0439**
	No Annotation	8	4	0.76	0.3833
I felt satisfied with the result					
	Annotation	**15**	5	4.06	**0.0439**
	No Annotation	9	6	0.26	0.6101
The interface was most complex					
	Annotation	8	10	0.06	0.8065
	No Annotation	5	8	0.3	0.5839
The interface was frustrating					
	Annotation	6	12	1.38	0.2401
	No Annotation	5	9	0.64	0.4237
I felt out of control					
	Annotation	3	**15**	6.72	**0.0095**
	No Annotation	6	9	0.26	0.6101
Other people ignored my contributions					
	Annotation	2	**13**	6.66	**0.0099**
	No Annotation	4	8	0.76	0.3833

Table 1: Questionnaire results

Ratings (4 points Likert scale)	Standard dev		Mean	
	ID	**No ID**	**ID**	**No ID**
Annotation	0.98	0.96	2.52	1.62
No Annotation	1.08	0.68	2.17	1.67

Table 2: Participants' ratings of their group's collaboration

	Standard Deviation		Mean	
	ID	**No ID**	**ID**	**No ID**
Annotation	1.15	0.94	3.71	3.86
No Annotation	1.69	0.84	2.83	2.50

Table 3: Average ratings of final compositions

	No ID						ID					
	System related	**Presence**	**Query presence**	**Quality Judgemen**	**Task**	**Social**	**System related**	**Presence**	**Query presence**	**Quality Judgemen**	**Task**	**Social**
Total number of annotations on this topic	3	12	6	14	27	24	1	16	8	20	24	31
Mean number of annotation on this topic	0.43	1.71	0.86	2.00	3.86	3.43	0.14	2.29	1.14	2.86	3.43	4.43
Mean number of contributions as a % of total	3%	19%	8%	13%	28%	27%	1%	20%	10%	18%	21%	26%
Standard deviation of number of contributions as a % of total	5%	23%	14%	12%	25%	15%	3%	22%	8%	15%	20%	16%

Table 4: Topics covered by the annotations

Contributions	Standard dev		Mean	
	ID	No ID	ID	No ID
Annotation	57.6	54.8	77.6	76.7
No Annotation	82.2	129.7	181.8	226.8

Table 5: Mean number of notes contributed per participant

Deletions	Standard dev		Mean	
	ID	No ID	ID	No ID
Annotation	62.9	49.49	55.4	60.2
No Annotation	116.1	169.1	150.2	182.6

Table 6: Number of deletions

Deletions	Standard dev		Mean	
	Self	**Other**	**Self**	**Other**
Annotation	57.1	72.5	70.9	44.7
No Annotation	116.4	189.4	166.1	167.8

Table 7: Number of Self and Other deletions

Notes on same spoke	Standard dev		Mean	
	Self	**Other**	**Self**	**Other**
Annotation	0.73	0.22	0.59	0.26
No Annotation	0.68	0.39	1.08	0.64

Table 8: Number of notes already on same spoke

Notes on same spoke	Standard dev		Mean	
	Self	**Other**	**Self**	**Other**
ID	0.83	0.29	0.90	0.39
No ID	0.58	0.31	0.78	0.51

Table 9: Number of notes on same spoke by ID

	Standard dev				Mean			
	Same spoke Self	**Same spoke Other**	**Proximity Self**	**Proximity Other**	**Same spoke Self**	**Same spoke Other**	**Proximity Self**	**Proximity Other**
Annotation	1.14	1.66	2.33	3.12	2.54	3.76	3.90	7.18
No Annotation	0.91	0.88	0.95	1.50	2.34	4.07	2.61	4.78

Table 10: distances to nearest notes

12 Figures

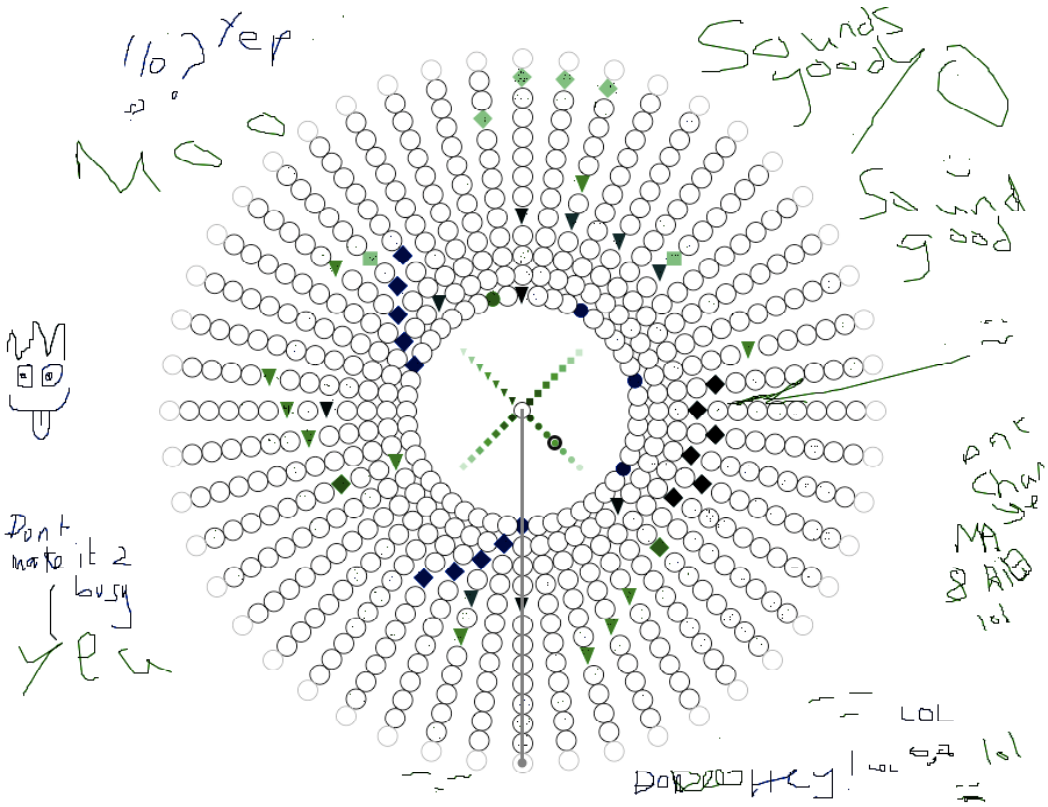


Figure 1 - the Daisyphone user interface

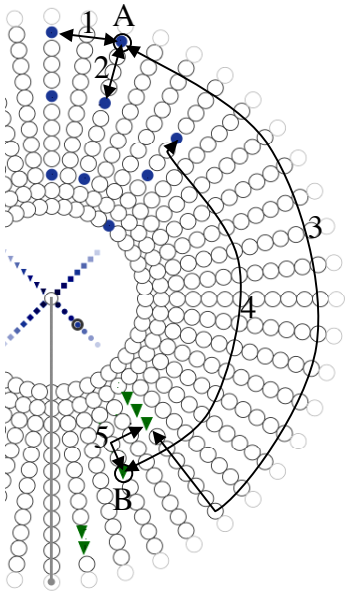


Figure 2: Example of proximity of interaction

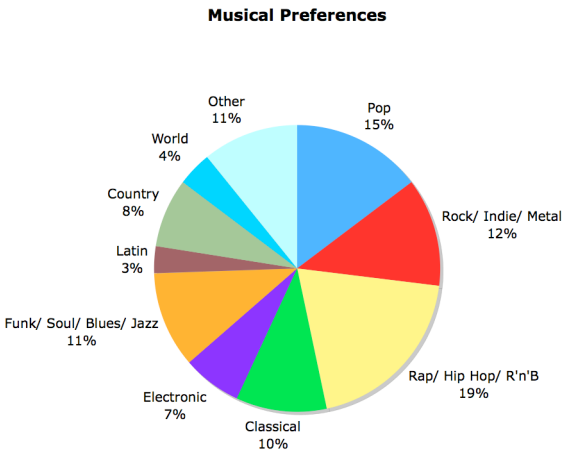


Figure 3: Musical Preferences of participants

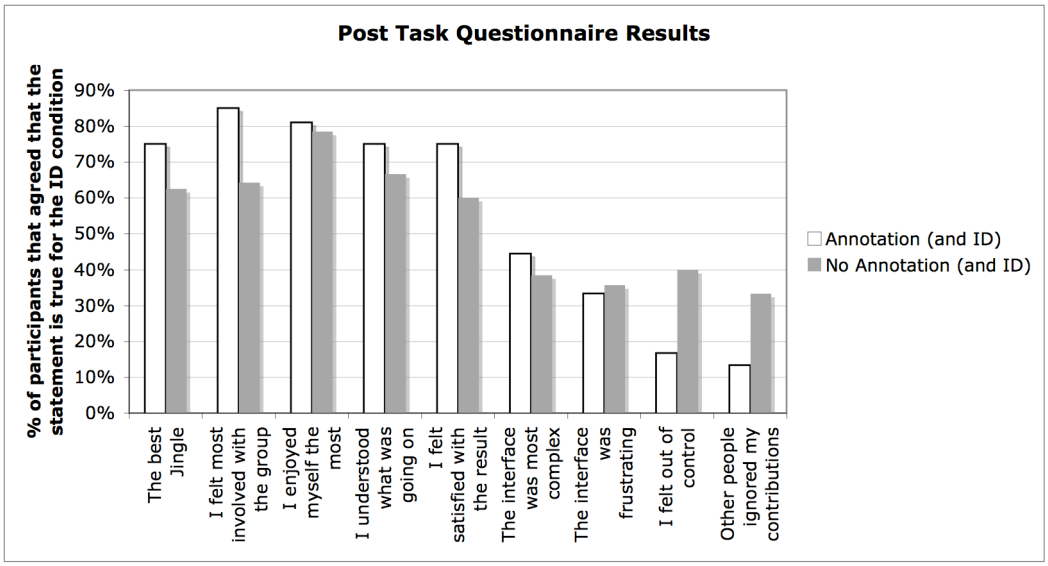


Figure 4: Post task questionnaire results

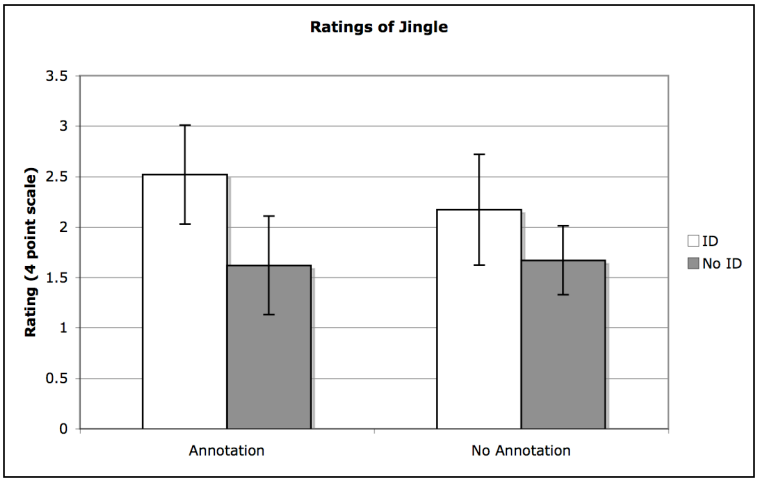


Figure 5: Participants' ratings of their group's collaboration

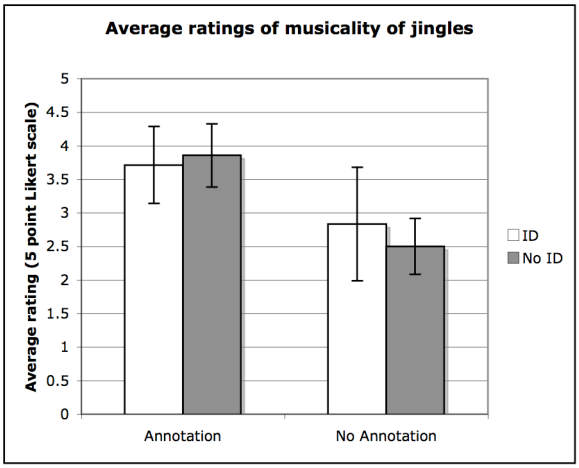


Figure 6: Average ratings of musicality of final jingles

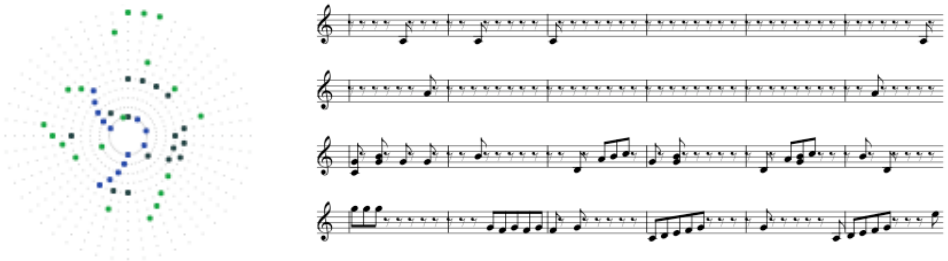


Figure 7a: Musical jingle

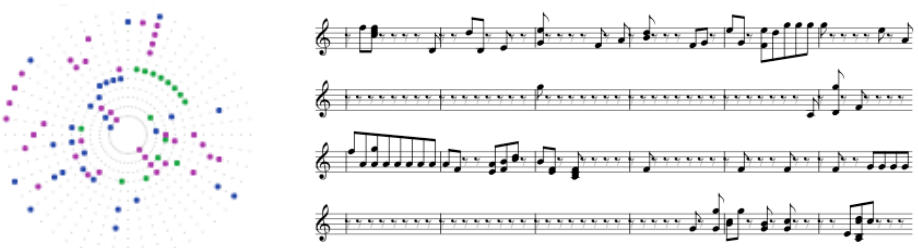


Figure 7b: Non-musical jingle

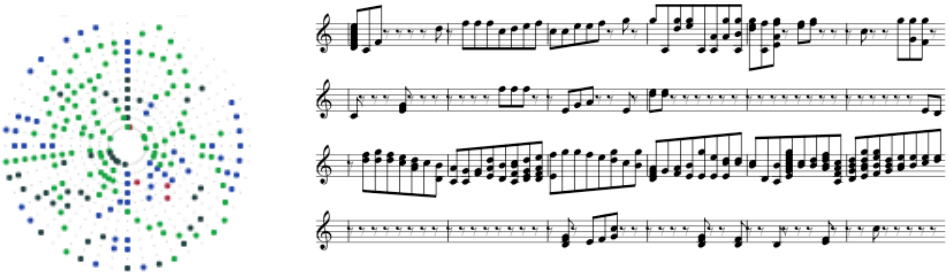


Figure 7c: Non-musical jingle

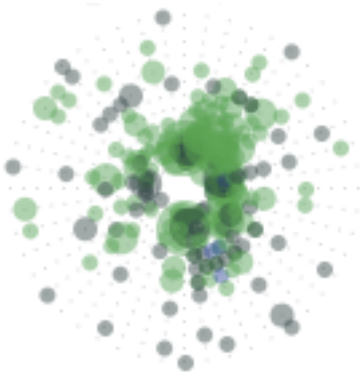


Figure 8a: Overview b: Final jingle

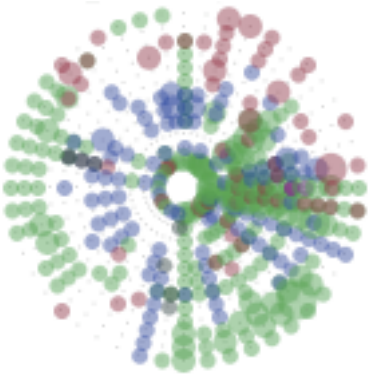
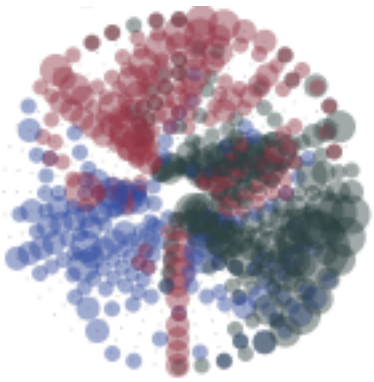


Figure 9: Overview of non-musical interaction

Figure 10a: Spatial separation (2)



b: Spatial separation (3)

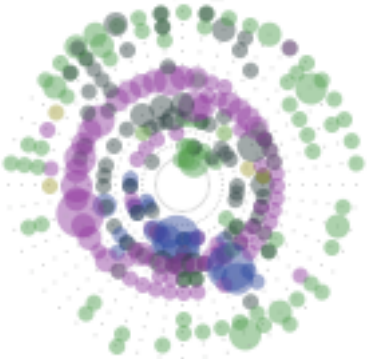


Figure 10c: Spatial separation around rings of the Daisyphone

d: Final composition

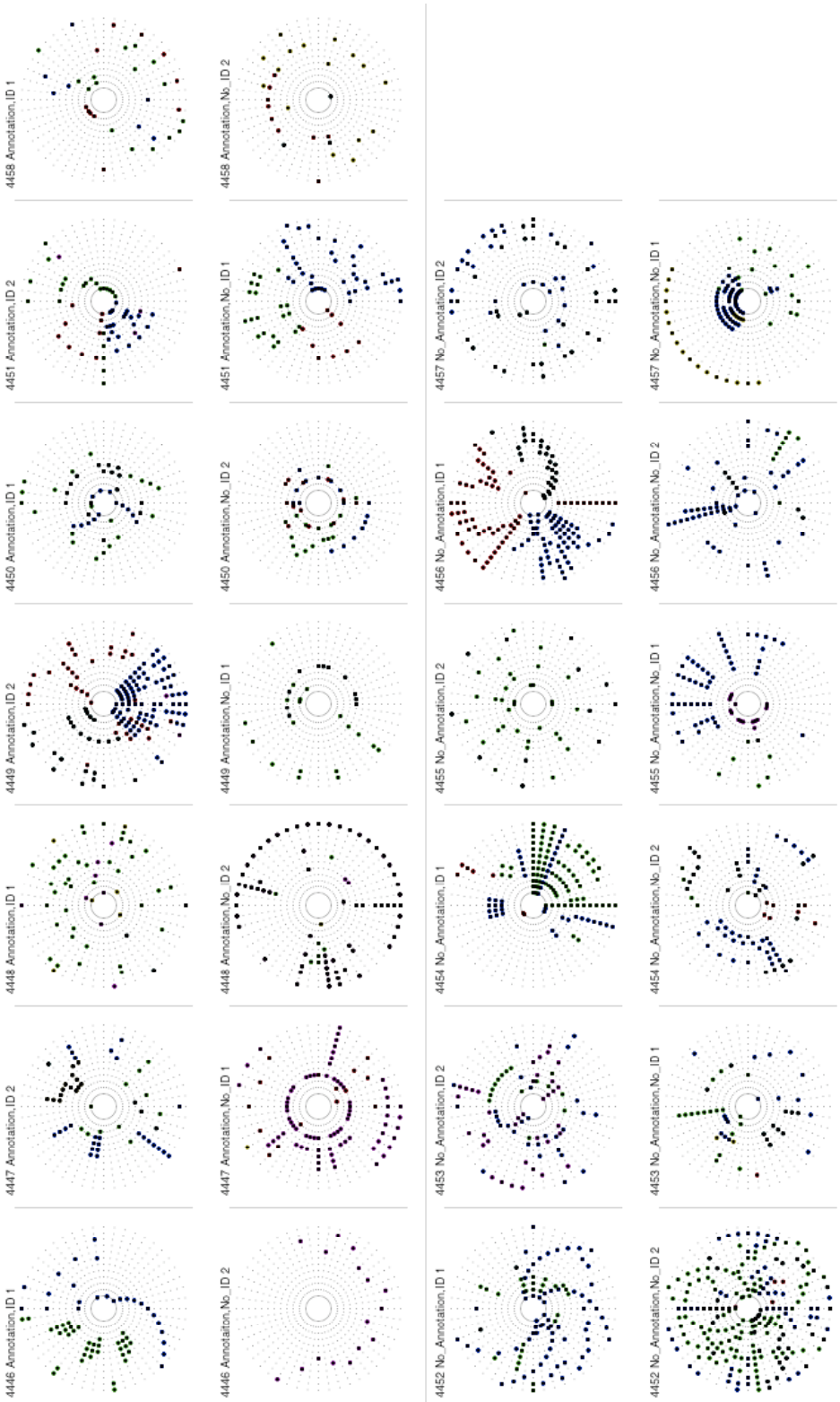


Figure 11: Final placements of notes

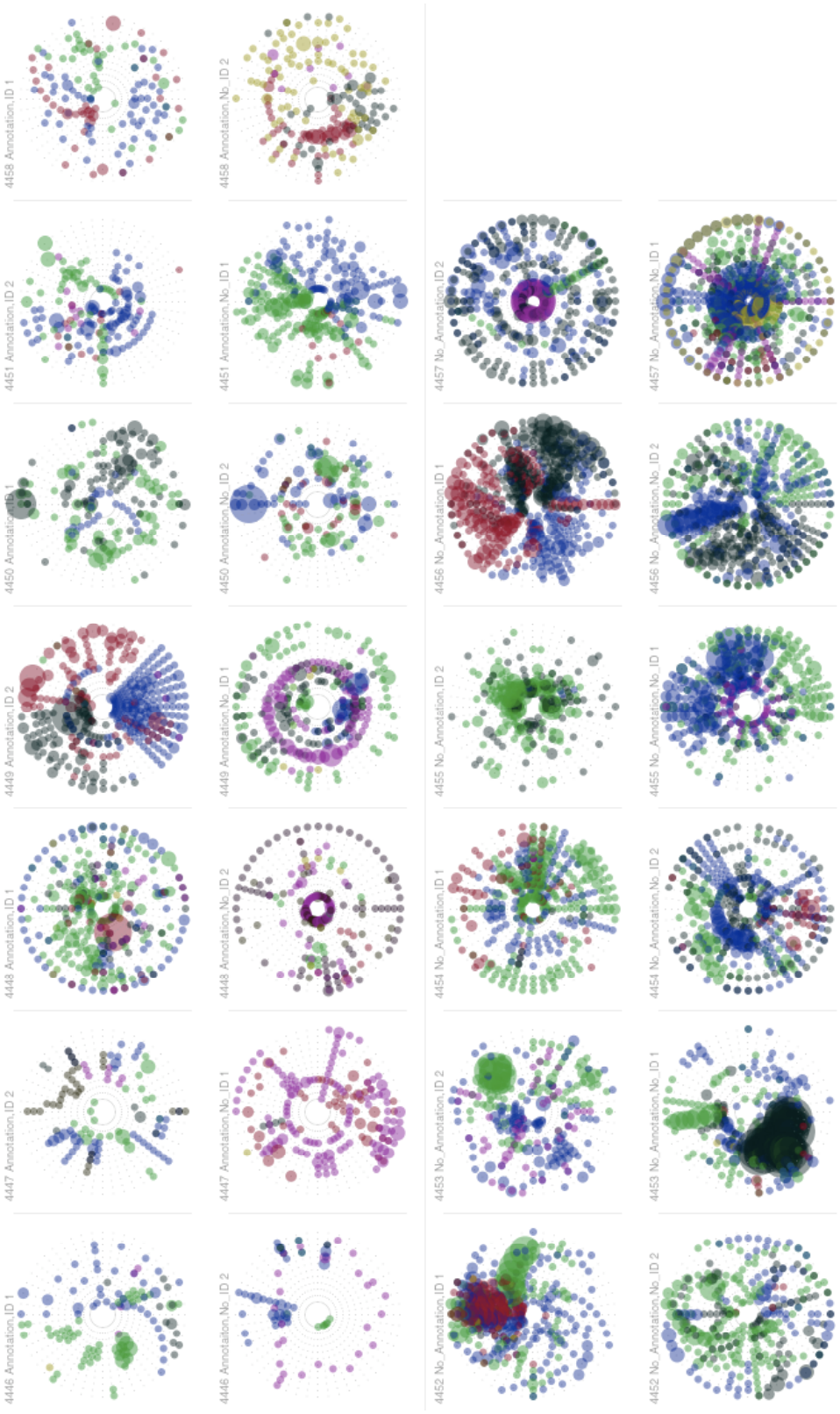


Figure 12: Activity summary

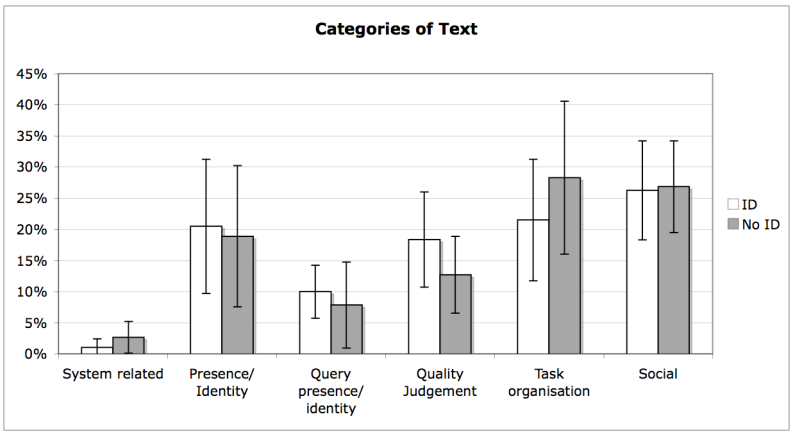


Figure 13: Topics covered by the annotations

Figure 14: Acknowledgement

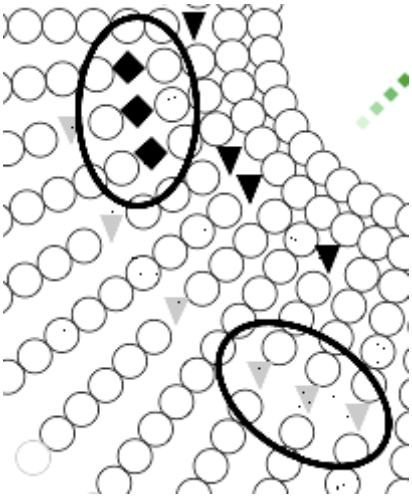


Figure 15: Example of Mirroring

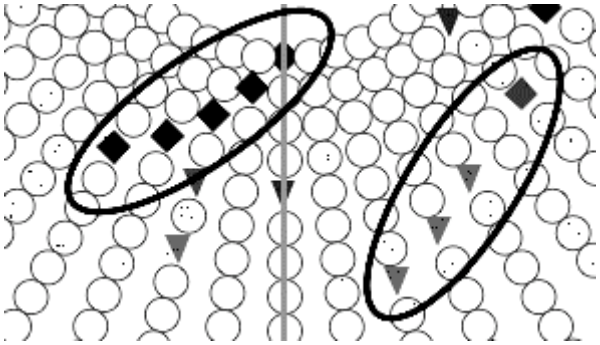
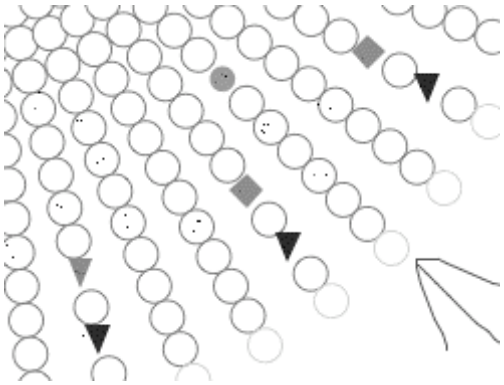


Figure 16: Example of Transformation

Figure 17a: Complementing in sequence



17b: Complementing in parallel

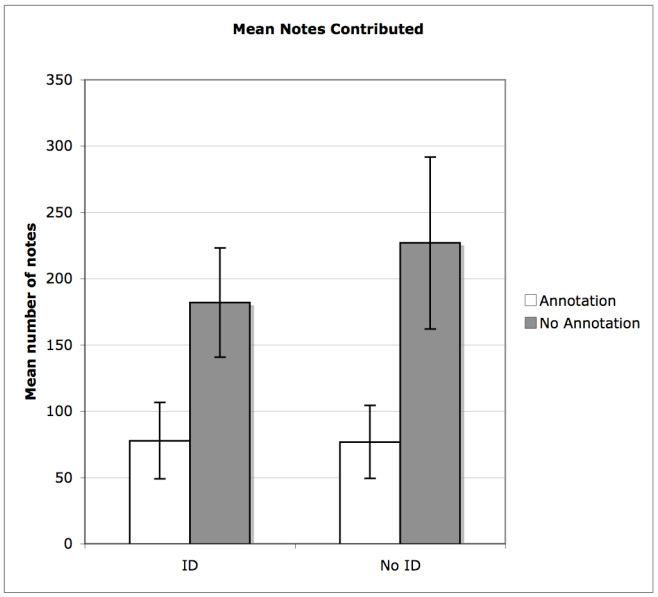


Figure 18: Mean number of notes per participant

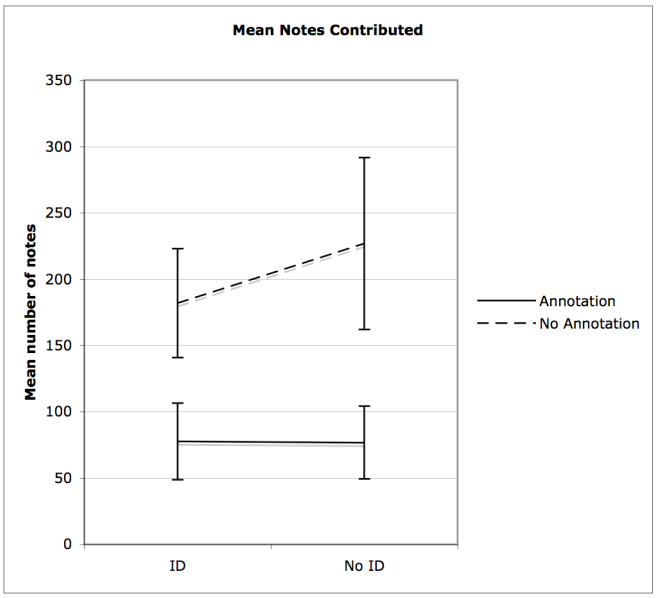


Figure 19: Interaction of mean number of notes contributed per participant

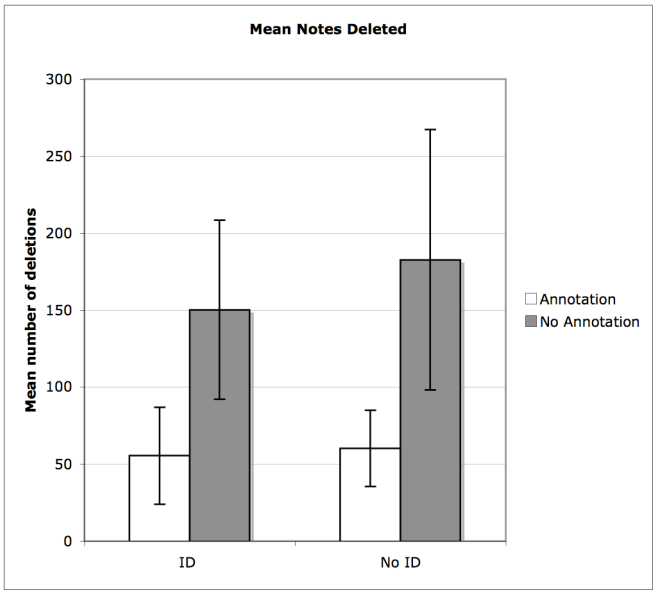


Figure 20: number of deletions

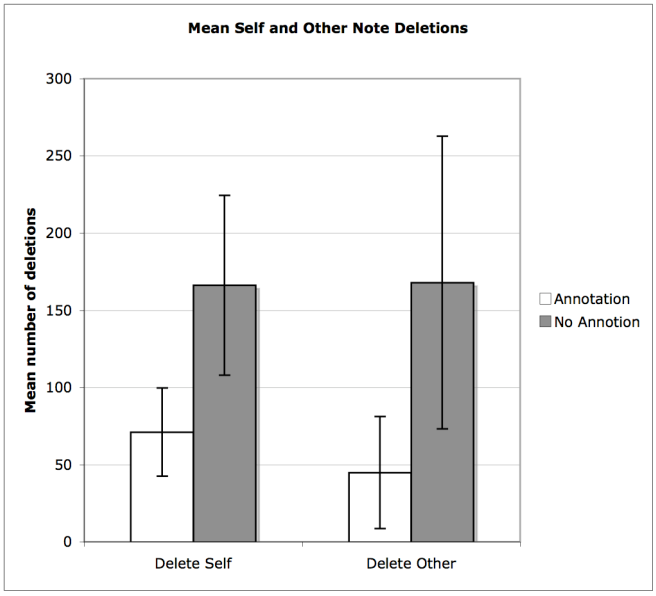


Figure 21: Number of Self and Other deletions

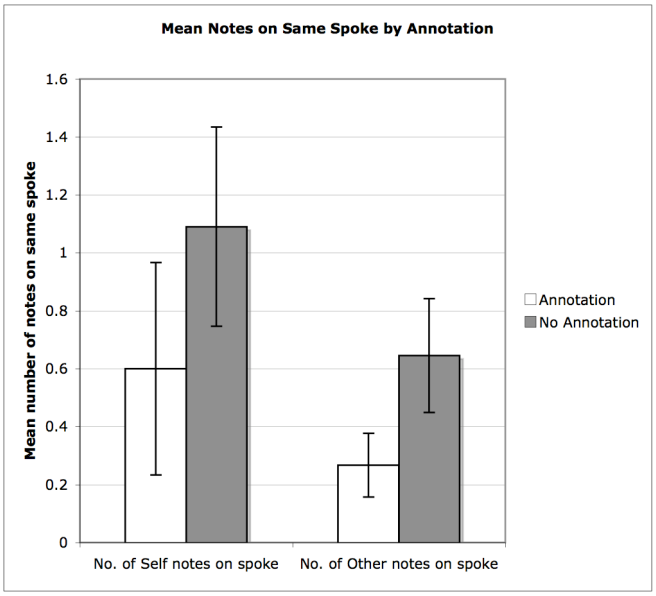


Figure 22: number of notes on the same spoke

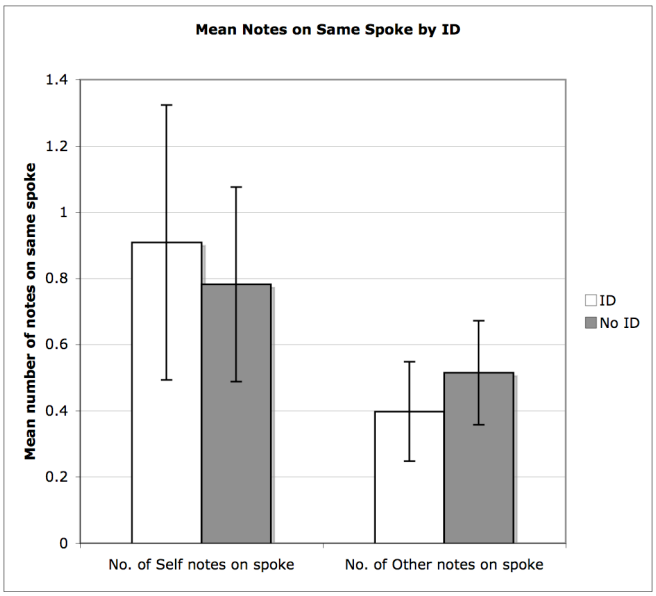


Figure 23: number of notes on same spoke by ID

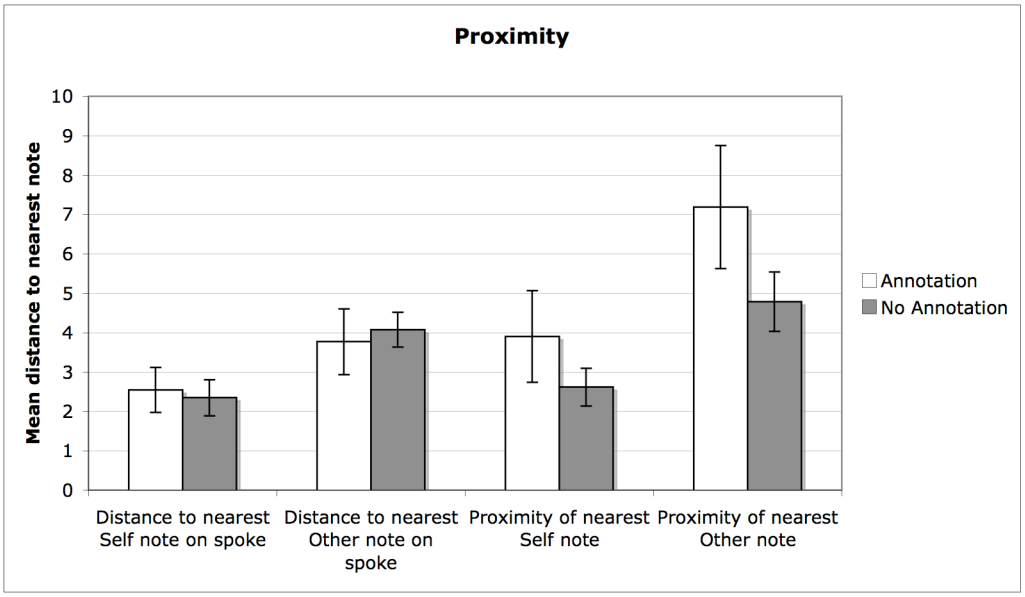


Figure 24: distances to nearest notes