Augmenting Face-to-Face Collaboration with Low-Resolution Semi-Ambient Feedback

THÈSE Nº 4895 (2010)

PRÉSENTÉE LE 17 DÉCEMBRE 2010 À LA FACULTÉ INFORMATIQUE ET COMMUNICATIONS RECHERCHE ET APPUI POUR LA FORMATION ET SES TECHNOLOGIES PROGRAMME DOCTORAL EN INFORMATIQUE, COMMUNICATIONS ET INFORMATION

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Khaled BACHOUR

acceptée sur proposition du jury:

Prof. J. Huang, président du jury Prof. P. Dillenbourg, Dr F. Kaplan, directeurs de thèse Prof. L. Bannon, rapporteur Dr D. Gatica-Perez, rapporteur Prof. K. Höök, rapporteur



Résumé

L ES CONVERSATIONS FORMENT une part importante de notre vie quotidienne, dans un grand nombre de contextes et en particulier dans les situations de travail et d'apprentissage collaboratif. Participer activement à une discussion implique une grand attention, raison pour laquelle les ordinateurs sont généralement considérés comme envahissant. Ils tendent en effet à occuper une part non négligeable de l'attention de leurs utilisateurs et réduisent donc leur capacité à tenir une conversation. Plusieurs causes expliquent ce phénomène, en particulier la présence d'un écran vertical qui agit comme une barrière entre les gens, et des interfaces telles que les claviers ou souris qui obligent les utilisateurs à interagir de manière explicite avec les outils informatiques. De nouvelles formes d'ordinateurs offrent cependant un solution, grâce à des machines capables de se dissimuler dans notre environnement. Ces outils réalisent leur fonction sans nécessiter d'action directe de leurs utilisateurs, leur permettant ainsi de se concentrer sur leurs tâches sans être déconcentrés.

Dans cette thèse, nous proposons un système qui permet d'explorer le rôle que ces nouveaux types d'outils prennent dans les conversations en face-à-face. Notre table interactive, *Reflect*, écoute les conversations se déroulant autour d'elle grâce à des microphones et présente sur sa surface des informations relatives à la participation de chaque personne, de façon discrète et non contraignante. Nous tentons de répondre à un certain nombre de questions concernant le potentiel d'un tel dispositif pour améliorer la qualité des situations collaboratives. Un effet auquel nous nous intéressons en particulier concerne la capacité de cet outil à modifier le comportement des participants à une conversation et les conditions qui favorisent un tel changement. Nous examinons également si un effet peut avoir lieu sans que les utilisateurs n'en soient conscients. De plus, nous observons l'utilisation de la table *Reflect* dans un contexte authentique, plus précisément dans le cadre de la formation en communication. Nous décrivons enfin les étapes dans la conception qui ont été nécessaires pour faciliter la transition des situations contrôlées, créées en laboratoire, à une utilisation en conditions réelles.

Pour répondre à ces questions, deux études expérimentales ont été conduites avec notre système. La première étude montre de quelle manière et sous quelles conditions *Reflect* peut être utilisée pour encourager une participation équilibrée dans une situation collaborative. La deuxième étude teste l'impact de la table sur la façon dont les participants à une conversation s'expriment. Les résultats nous ont permis d'identifier les difficultés inhérentes à la création de ce types d'effets, ainsi que des différences dans les réactions des participants en fonction de leur sexe.

Dans le cadre de l'utilisation de la table *Reflect* pour la formation en communication, nous observons les changements qui doivent être apportés au système par rapport au contexte très contrôlé du laboratoire. Nous discutons finalement de la façon dont la table est perçue par les personnes l'ayant utilisée dans ce contexte authentique.

Mots-clés: Informatique Omniprésente, Interaction Homme-Machine, Travail Coopératif Supporté par Ordinateur, Apprentissage Collaboratif Supporté par Ordinateur.

Abstract

CONVERSATIONS ARE A DAILY HUMAN ACTIVITY, and a crucial part of face-to-face collaboration in virtually any context. It is also a highly engaging activity that requires the full attention of participants involved in it. This is why computers have generally been perceived as intrusive in the world of human conversation, for they take some of their user's attentive focus, reducing their capacity to engage with the other. However, computers today are no longer limited to pieces of technology that we place in front of us or hold in our hands while we interact directly with them via keyboards, touch screens or other input devices. Some computers now hide in our environment, avoiding our attention, achieving whatever function is required of them without us even knowing they are there, and leaving us to focus on the tasks that are important to us.

We present a system to explore the role computers can take in face-to-face conversations within the context of this new computing paradigm. Our interactive table, which we call *Reflect*, monitors the conversation taking place around it via embedded microphones and displays relevant information about member participation on its surface in a discreet and unobtrusive manner. We raise several questions about how such a device can be used to improve the quality of face-to-face collaboration. In particular, we explore whether or not this system is capable altering user behavior and under what conditions this is possible. We also examine whether or it is possible to achieve a change in user behavior while remaining unobtrusive. In addition, we look at the use of such a device outside the scope of face-to-face collaboration by examining its role in the world of communication training. Finally we study the transition process and the design changes needed to bring such a device out of the laboratory and into the real world.

To answer these questions, we describe two user studies conducted on the *Reflect* table. In the first study, we show how the table can be used to promote balanced partcipation and we examine the conditions under which this is possible. In the second study, we test such a system's ability to change the way people speak during a conversation, and show some of the difficulties in achieving that, as well as some differences in how male and female users respond to such a device.

We then take the *Reflect* table outside of the laboratory and explore its use in the real world. We explore the changes to the system design that are needed for such a transition to take place. We also show how the table is perceived by users outside the scope of a laboratory study.

Key words: Ubiquitous Computing, Human-Computer Interaction, Computer-Supported Cooperative Work, Computer-Supported Collaborative Learning.

Acknowledgements

THE TITLE PAGE OF THIS DOCUMENT calls it a PhD dissertation "by Khaled Bachour," but that is a clear over-simplification of the reality. There is little if anything in this document that I can honestly and accurately attribute to myself alone, so I would like to take a moment to correct this blatantly reductionist title page.

First and foremost I wish to extend my sincere gratitude to my supervisors Pierre Dillenbourg and Frédéric Kaplan. Pierre, conceiver of the original idea behind the *Reflect* table, provided endless academic, material and moral support throughout the duration of my studies all the while maintaining a delicate balance between the authoritative wisdom and spirited guidance that I needed to succeed and to do so joyfully. Frédéric complimented my hesitant caution with a boundless spirit of creative optimism that pushed me to make my work a lot more exciting than I would have allowed it to on my own.

I would like to extend my thanks to the members of my review committee: Liam Bannon, Daniel-Gatica Perez, Kristina Höök, and Jeffrey Huang. Thank you for your insightful comments and for pushing me with your questions to reexamine the work I did from a fresh new angle.

A great part of this work would not have been possible without the help of many people who are not part of the CRAFT team. The table itself would not have existed without the indespensible expertise of Christof Faller and Fritz Menzer in accoustics, Branka Zei Pollerman in voice analysis, Martino D'Esposito in industrial design and René Beuchat in processor architecture, and their mastery of their respective domains. Thanks also to Pierre Jacot, Sonia Orellana, Carine Peretti, Olivier Siegenthaler, Olivier Stauffer and all members and partners of the CEP who provided the necessary support and invaluable feedback on the use of the *Reflect* table in the domain of communications training.

A large chunk of my gratitude goes to the members of CRAFT throughout my years there, especially those who were directly involved with my PhD at some point or another and with whom I've had the pleasure of collaborating directly. My thanks to Quentin Bonnard, whose six-month involvement with the *Reflect* table made them some of the more interesting and enjoyable months of my PhD, and to Wolfgang Hokenmeier and Olivier Guedat whose help in developing and maintaining the table was invaluable. Thanks also to Asheesh Gulati, Andrina Brun, Amit Sharma, Anurag Nilesh and Aniruddha Jha who each had a hand in making the table what it is today. I would also like to thank Guillaume Raymondon and Jean-Baptiste Haué who had conducted the preliminary work which my PhD was based on.

To them and to the rest of CRAFT, I wish to say thank you being my family away from home, for always being ready to help, and for making life as a PhD student a lot easier and more enjoyable than I would ever had expected. Guillaume Zufferey who was with me in the trenches from start to finish, as well as Fabrice Hong and Jessica Dehler, thanks for providing all that a friend could need: help, advice, food and beer. The same goes to my friends and fellow PhD students Hamed Alavi and Andrea Mazzei, and the rest of the CRAFT crew: Patrick Jermann, Son Do Lenh, Marc-Antoine Nüssli, Sebastien Cuendet, Nan Li, Himanshu Verma, Julia Fink, Flaviu Roman, Ingrid Le Duc, Jean-Louis Ricci, Nadine Steiner and Paul Oberson, as well Gaelle Molinari, Mirweis Sangin, Mauro Cherubini, and Nichola Nova, all of whom made CRAFT, at one point or another, the wonderful experience it was to me. Thanks also to David Brechet for always being ready to

give a hand when needed, and of course, to Florence Colomb the backbone of CRAFT and the secret ingredient to the pleasantly cheerful and superbly fluid functioning of our group.

Last but not least, to my mother and father, my brothers and sisters, and my friends in Lebanon and Switzerland and elsewhere in the world, thank you for being who you are and for being, by chance or by choice, part of my life. I could not have done this without your unlimited love and support. I hope I made you proud.

Contents

1	Intro	ductio	n	1					
2	Liter	Literature Review							
	2.1	Ubiqu	litous Computing and Roomware	3					
		2.1.1	Context-awareness	5					
		2.1.2	Privacy	6					
		2.1.3	Territoriality	6					
		2.1.4	Economic concerns	7					
		2.1.5	Redefining places	7					
	2.2	Comp	outer-Supported Collaborative Work and Learning	8					
		2.2.1	Imitation bias	8					
		2.2.2	Awareness	9					
		2.2.3	Grounding and persistence	10					
		2.2.4	The "interactions" paradigm of CSCL	11					
		2.2.5	Participation in collaborative learning	11					
		2.2.6	Participation balance	12					
		2.2.7	Group mirrors	12					
		2.2.8	Regulation	13					
	2.3	Auton	nated Meeting Analysis	13					
	2.4	Voice.	Prosody and Emotion	15					
	2.5	Relate	d Works	15					
		2.5.1	Second Messenger	15					
		2.5.2	Conversation Clock and Conversation Votes	16					
		2.5.3	The metaphoric group mirror	18					
		2.5.4	Meeting Mediator	18					
		255	GroupMeter	19					
	2.6	Summ	ary	20					
з	Refle	Paflact Decign Process							
0	31	The Pi	rinciples of <i>Reflect</i>	21					
	3.2	The D	lesion Invariants of Reflect	23					
	0.2	321	Information to visualize	23					
		322	Mirroring	23					
		373	Unobtrusivo input	23					
		3.2.5	Shared visible output	20					
		325	Low-resolution display	24					
		326	Minimal interactivity	24					
	33	Rofloct	s a Somi_Ambient Display	25					
	3.0	The E	as a semi-simplem Display	25 25					
	5.4	2 <u>/</u> 1	The Virtual Noice-Sencitive Table	25 25					
		347	The Physical Noise-Sensitive Table	25 26					
		313	The First Reflect	20 27					
		J. 1 .J		41					

		3.4.4 The Second <i>Reflect</i>
	3.5	Visualizations
	3.6	Summary
	3.7	Research Ouestions 32
	0.1	
4	Stud	y 1: Participation Balance 35
	4.1	Motivation for Participation Balance
	4.2	User Study on Reflect
	4.3	Experimental Method
		4.3.1 Description of the experiment 37
		4.3.2 Experimental conditions
		4.3.2 Experimental procedure 30
		4.3.5 Experimental procedure
	4 4	4.5.4 Data conection
	4.4	Kesuits
		$4.4.1 Visibility \text{ and } unobtrustoeness \dots $
		4.4.2 General effect on balancing participation
		4.4.3 Effect on over and underparticipators
		4.4.4 Effect on individual awareness
		4.4.5 Effect on topic balance
		4.4.6 Qualitative findings
	4.5	Discussion and Limitations
		4.5.1 Validation of hypotheses
		4.5.2 Limitations of the study
	4.6	Summary
		ý
5	Stud	y 2: Vocal Engagement 49
	5.1	Prosody in Voice
	5.2	Prosody and Engagement
	5.3	Voice Analysis in <i>Reflect</i>
		5.3.1 Measuring and merging features
		5.3.2 Segmenting the conversation stream
		5.3.3 Visualizing engagement 53
		5.3.4 Updated system architecture 54
	54	Experiment 55
	55	Regulte 56
	5.5	55.1 Conoral observations on participation balance and ongagement 56
		5.5.1 General differences in aroused earoes conditions
		5.5.2 General unterence in arousal across conditions
		5.5.3 Gender differences in benavioral changes
		5.5.4 Iemporal evolution of arousal
		5.5.5 Summary
6	Dof	act Outside the Laboratory (5
0		Eucloside the Laboratory 65
	6.1	Exploring the Scope of Use
		$6.1.1 \text{Formal communication training} \dots \dots \dots \dots \dots \dots \dots \dots \dots $
	(^	6.1.2 Uther uses 67
	6.2	Kejlect Kedesigned
		6.2.1 Implementation of <i>AudioButtons</i>
		6.2.2 Using <i>AudioButtons</i> to interact with <i>Reflect</i>
		6.2.3 Updated architecture
	6.3	<i>Reflect</i> at CEP
		6.3.1 Participant feedback
		6.3.2 Trainer feedback
		6.3.3 Lessons learned

	6.4	Summary	76					
7	' General Discussion and Conclusions							
	7.1	Summary	77					
	7.2	General Discussion	78					
		7.2.1 The effect of <i>Reflect</i> on user behavior	78					
		7.2.2 Gender differences in human-computer interaction	78					
		7.2.3 Group mirrors and norms	79					
		7.2.4 Semi-ambient displays	79					
		7.2.5 Prosodic analysis of group meetings	79					
		7.2.6 Voice analysis beyond communication training	80					
	7.3	Limitations	80					
		7.3.1 Generalizability of experimental results	80					
		7.3.2 Duration of the CEP study	80					
		7.3.3 <i>Reflect</i> across culture	81					
	7.4	Final words	81					
Bibliography								
A	Post-experiment Questionnaire							
B	B CEP User Questionnaire							
Lis	List of Figures							
Cu	Curriculum Vitæ							

Chapter 1 Introduction

"I can avoid being seen if I wish, but to disappear entirely, that is a rare gift."

J. R. R. TOLKIEN, from The Lord of the Rings: The Fellowship of the Ring

"Am I talking too much?" is a rather simple question that arises quite often in conversations. Some people do talk too much. Others talk too little. In fact, most people would probably have no trouble at all thinking of a person in their school, workplace, or their social circles that falls in either category. While in general this may not be an issue that can or even needs to be addressed, there are situations where imbalance in conversational speech may be problematic. Collaborative learning is one example we address in this thesis where imbalance in a conversation can be detrimental both to individuals in the group and to the group as a whole. Another issue that arises in conversations is related to *how* people speak. Members of a group who sound bored and uninterested as they speak would likely hurt the motivation of those they are working with. This thesis discusses the use of computers in an attempt to mitigate some of the issues associated with face-to-face collaboration.

The adoption of computing technology, however, has been rather limited in the domain of faceto-face human conversation, and it has remained extraneous to this activity. When it takes a more central role, it is often perceived as intrusive and distracting. This is because computers have, until recently, been developed to be in the center of any activity that involves them. Workstations were designed to dominate our scope of attention while we accomplish such tasks as word processing and internet browsing, or more entertaining activities such as playing games or watching video. Thus adding a computer to a human-human conversation cannot go unnoticed, and it is bound to cause a significant change in the way the conversation takes place. On the other hand, this paradigm has shifted in recent years, and the abundance of technology and the ever decreasing cost of hardware have made it possible to develop computing systems that take a secondary role in the lives of their users. Computers began to hide in the background of the user's activity, thereby allowing them to play a useful role in situations where the user's attention is focused on more important aspects of the activity. A modern automobile for example contains several computer chips that perform different tasks that their user need not be directly aware of. Thus they take a backseat to the more important task the user is focusing on which is driving the car.

Much like driving a car, participating in a face-to-face conversation is a complex task that requires our full attention, for conversations are not simply an exchange of verbal utterances. They are a highly intricate and synchronized dance with both interlocutors constantly perceiving and reacting to each others' utterances, body movements, and facial expressions. It is extremely difficult, if not impossible, for one person to fully engage in a face-to-face conversation while also focusing on some other task. Thus, the traditional computer had no real place in human face-to-face conversations. However, with the advent of the new paradigm of hidden computers, which we shall discuss in more detail in the next chapter, we are now able to imagine computer

devices that may find a place in the delicate world of human conversation. This is the domain of this work, and over the course of this dissertation, we shall present the motivation, the methods, the evaluation and the conclusions of our research in trying to influence face-to-face conversations with a "disappearing computer."

The remainder of this thesis is organized as follows:

- We begin by presenting a review of the relevant literature in Chapter 2 where we provide the proper scientific context in which this work was made. We also describe some similar work that has been done by others.
- Chapter 3 describes the system developed for this work. We explain the design principles, the different steps during the iterative design process of the *Reflect* table, as well as the architecture of the resulting system. We then proceed with our main research questions for this dissertation.
- In Chapter 4 we describe the first user study conducted that evaluated the ability of the *Reflect* table to promote balanced participation among its users. We explain the results of the study and draw the relevant lessons, particularly in terms of the conditions under which the table succeeds in its objective.
- The second user study is explained in Chapter 5 and involves an evaluation of the ability of the table to alter the level of vocal engagement among its users. This chapter also describes in detail the voice analysis system used in the table and grounds it in literature on voice as well as professional practice.
- We then take the table outside the laboratory in Chapter 6, where we explain the challenges faced by this transition, especially in terms of the changes to the original design of the table. We also discuss some of the lessons learned from the use of the table in the real world.
- Finally, we conclude with a summary and a general discussion on the contributions of this work, its limitations and its implications for future research in all of the relevant domains.

Chapter 2

Literature Review

"Computer Science is no more about computers than astronomy is about telescopes."

Edsger W. DIJKSTRA

Our work lies at the crossroads of several research domains. The questions we raise fall within the realm of Ubiquitous Computing and particularly in what relates to Roomware and Ambient Displays. Its applications and domains of use belong to the field of Computer-Supported Collaborative Learning and Computer-Supported Cooperative Work. It benefits from notions of conversation and meeting analysis as well as prosodic analysis of human voice. In this chapter we give a brief overview of these research domains inasmuch as they form the basis of our work. We also describe related works where other researchers attempted to answer questions that overlap with our own.

2.1 Ubiquitous Computing and Roomware

When ordering from a restaurant menu, observing an entrance-only sign at a door or singing karaoke at a bar, we are unknowing users of a very ancient technology: the alphabet. We do not consiously think of ourselves as "reading" while performing any of these activities. In fact, the alphabet, as a man-made technology, is so ubiquitous that we do not notice it is there anymore. This is Mark Weiser's 1991 vision of computing in the 21^{st} century [Weiser 91], which he calls "Ubiquitous Computing" (Ubicomp). The term ubiquitous is not used in the sense that we can take our laptop or smart phone anywhere we want, but in the sense that we are no longer aware of its presence and use it unconsciously, the same way we use the alphabet.

Weiser describes three basic types of devices with embedded computing functionalities that vary in size: tabs are pocket-sized, pads are more like sheets of paper, and boards are wall-scale bulletin board devices. Each of these types of device would carry its own particular function, and they are all connected via wired and wireless networks. Today, these devices have become a reality: tabs are becoming more common in the form of powerful smart phones, interactive multi-touch surfaces such as boards and tables are appearing in hi-tech meeting rooms, and only a few months ago, a new generation of pad devices have been introduced into the commercial market first by Apple, then by other manufacturers of consumer electronics. However, these tabs, pads, and boards do not entirely fulfill Weiser's vision of ubicomp, as they are still inherently computers with different shapes and sizes, and we still treat them as such. We use them to check our email, browse the web, play games, and present information in meetings. During this time though, another trend of ubicomp devices began to appear.



Figure 2.1: Apple's iPad (left), Google's Nexus One Android phone (center) and Microsoft's Surface table (right) correspond to Weiser's pads, tabs and boards respectively, but they do not fulfill the ubicomp vision Weiser proposed.

In 1996, Mark et al. introduced the term "roomware" [Mark 96] that took Weiser's notion of ubicomp to specifically incorporate computing capabilities into the components of the room itself, and Streitz et al. later defined roomware as "computer-augmented things resulting from the integration of room elements (e.g., walls, doors, furniture like tables, chairs, etc.) with computer-based information devices" [Streitz 98]. Roomware has been developping as its own field of research described as an "umbrella" framework for four fields: ubicomp, computer-supported cooperative work, augmented reality, and architecture. Since the term roomware was introduced, many examples of such devices have been developed that were as diverse in form as they were in function. Walls [Geißler 98, Haller 10], tables [Dietz 01, Bathiche 10], chairs [Mota 03], lamps [Do-Lenh 09, Alavi 09], clocks [Brown 07] and other devices were augmented with computing capabilities with varying purposes. Figure 2.2 shows a coffee mug augmented with heat sensors allowing it to warn its user when its content is too hot [Gellersen 99].



Figure 2.2: The MediaCup is a regular coffee mug augmented with the ability to warn its user when its content is too hot [Gellersen 99].



Figure 2.3: Four examples of roomware devices proposed by Streitz et al.: ConnecTables, Comm-Chair, InteracTable, and DynaWall [Streitz 01].

This type of augmentation allows the user to indeed forget about the technology and think of the cup itself as having an additional function, rather than thinking of it as simply a computer in a cup. In other words, unlike smart phones and pad computers, these devices are no longer perceived as computers of different shapes, but are their own new family of devices that bring us closer to Weiser's 1991 vision of ubicomp.

The ubicomp paradigm also inspired a new form of computing introduced first by Weiser and Brown as "calm computing" [Weiser 96] that focuses on utilizing users' peripheral vision as an additional channel for providing background information. Calm computing deals with tasks that require some level of user attention, but that must remain in the background in order to allow the user to focus more on their actual task. This was later taken up by researchers who demonstrated the potential of "ambient displays" which they defined as new way to interface between people and digital information through the use of environmental cues such as sound, light and movement, as opposed to the traditional screen-based direct display of information [Wisneski 98]. They introduce several types of ambient displays such as a lamp that projects light into the ceiling through water; the lamp creates ripples in the water to reflect the amount of network activity detected over the network. The result is that the light on the ceiling shows an intensity of rippling that corresponds to network activity, and that is perceived by a user without them necessarily directing their attention to the ceiling.

The advent of ubicomp and roomware, as with any new paradigm, brought with it new challenges as well as new opportunities especially in our understanding of how augmenting a physical space with technology affects how we interact with the space and with each other.

2.1.1 Context-awareness

With traditional desktop computing, there are few and well-known channels for device input: usually, a mouse and a keyboard. This creates a certain predictability to the interaction as the user knows what the computer "knows" and expects the computer to react based on the specific cues from its input channels, such as clicking the mouse or typing a key. The term "context-aware" computing was introduced to describe systems that are aware of their surroundings and not just of

what the user explicitly inputs. Context-aware systems gather information from their surrounding and use that to provide relevant services or information to the user [Dey 00]. The behavior of a context-aware system is therefore less predictable than traditional context-free systems that rely directly on the user for input. However, with traditional desktop computers, the ability of a system to be aware of its context is fairly limited and usually consists of awareness of time, location, weather, etc...

Roomware devices, on the other hand, have varied modes of input ranging from touch-screens that often simulate the mouse and keyboard, to less direct input methods such as cameras, microphones, as well as heat, motion and other types of sensors. Existing systems currently limit their context-awareness to the physical context. However, for the user, the context goes beyond the physical environment, and includes the social and cultural contexts, particularly when these devices are used by multiple persons simultaneously. The relevant context then can reach fairly complex levels and can include for example the number of people present in a room and the relationships and social interactions between them. Context itself becomes a complex concept with many more dimensions than a single system is able to consider. A truly context-aware system is thus unfeasible in practice when context is defined to include all aspects of the environment the system is in, the physical as well as the social and cultural.

2.1.2 Privacy

Langheinrich describes some of the privacy issues arising from ubicomp technologies [Langheinrich 10], particularly due to some technical capabilities that were not common to traditional computing, such as ability to detect the presence or identities of people in a room. These technologies also interact with users over less-restricted areas of space and periods of time, as opposed to a desktop computer that only interacts with the user when the user is sitting directly in front of it. Ubicomp systems are sometimes also vague on what information is captured from the user, how it is used, why it is needed, and whether or not it is stored.

Bellotti et al. proposed a framework for privacy in ubicomp technologies that addresses these issues [Bellotti 93]. They examine the four main concerns about user data mentioned above and describe how systems can mitigate privacy concerns by giving users feedback, i.e. informing the users about how their data is captured, used, stored and shared, and control, i.e. giving users the ability to determine these factors. For example, a system that has a microphone could reduce privacy concerns by announcing to the user when it is recording (feedback) or allowing the user to actively start and stop recording (control). A device that tracks the user's location would need to inform the user on whether or not the user's identity is attached to the information about their location (feedback) and allow them to determine who has access to this information (control).

Another more general issue with privacy is the concern that it is no longer enough to provide users with a level of privacy that they themselves are comfortable with, as studies on online social networks have shown that the average user's attitude towards privacy is not conservative enough to protect their data from malicious attackers [Gross 05]. It thus becomes a responsibility of the developer of technologies with privacy concerns to ensure that their user's privacy expectations are not only met, but also exceeded to the point of protecting these users from potential attacks that they may not even be aware of.

2.1.3 Territoriality

Territoriality is a notion that was rarely relevant in traditional one-user-one-machine systems. If a person is sitting in front of their desktop computer, that computer becomes within their territoriy, and no other person is expected to grab the mouse or keyboard without permission or in some cases to even look at the screen. With roomware, especially multi-user tabletop or wall systems, territoriality became an important factor and has been the subject of several research studies. Scott et al. describe three types of territories on shared spaces shown in Figure 2.4: group territories that



Figure 2.4: Different arrangements of personal, group, and storage territories in different contexts. Reproduced from [Scott 03]. Leftmost figure shows only one user, so there is no need for a group territory.

includes active elements that are common to all parties, personal territories that are reserved for use by one participant, and storage territories that contain common elements that are not currently in use. [Scott 06]. They also describe design implications for the notion of territories in shared spaces such as providing appropriate functionality in relevant territory regions [Scott 04].

2.1.4 Economic concerns

Davies and Gellerson cite, among the challenges of deploying ubicomp systems, economic obstacles [Davies 02]. The cost-to-value ratio of current roomware and ubicomp technologies is still prohibitively high. While one can easily imagine paying a certin sum for a single-purpose software application, it is harder to accept paying ten-fold that sum for a single-purpose hardware set-up. With consumers getting used to obtaining more and more functionality in smaller and cheaper devices, a piece of furniture dedicated to a single function does not seem so convincing. Current multi-purpose roomware technologies are few, and those that exist such as Microsoft Surface, are not yet within the budget of the average household. This of course will be less of a challenge as hardware becomes cheaper and more abundant, and new roomware devices are developed that are more and more useful for the end consumer.

2.1.5 Redefining places

Ciolfi and Bannon note that roomware augmentation does not only provide new modes of interaction and new possibilities for activities, but also "impacts the culturally influenced qualities of an environment or even changes them to some extent" [Ciolfi 05]. In their work they describe a two-room system for discovering a museum by exploring objects of interest, investigating and reflecting on informative material, and expressing their own opinions. They describe how they design rooms to fit these activities. The Study Room was old-fashioned with a homey intimate feel that allows the visitor to relax and take their time exploring, investigating and reflecting on the material. The visitors would interact with different devices embedded in pieces of furniture such as an interactive desk, a radio and a storage trunk. The Room of Opinion, in contrast was simple, black and white dimly lit room where visitors are not distracted by the outside world. The recording equipment is simple and unintimidating allowing the users to focus on recording their thoughts. Roomware thus does not simply introduce interaction to a physical space, but can also redefine the nature of the place the user is entering.

Two domains of research that are particularly relevant for our work have also been particularly influenced by the advent of roomware. These are discussed in the next section.



Figure 2.5: The Study Room (left) and the Room of Opinion (right) are examples of roomware that changes the way we approach the place [Ciolfi 05].

2.2 Computer-Supported Collaborative Work and Learning

Our work is concerned with two main fields of research, both of which have been significantly influenced by the roomware paradigm. The first is Computer-Supported Cooperative Work (CSCW), a term first introduced by Iren Greif and Paul Cashman in 1984 as a title of a workshop that brought together experts in different fields who shared a common interest in understanding the role of technology in how people work [Grudin 94]. While the term groupware was introduced to refer to technologies that augment, mediate, or otherwise facilitate group work, CSCW was the field of research involved with understanding how these technologies are used and their effect on collaborating and cooperating teams.

The second field, Computer-Supported Collaborative Learning (CSCL), is similar to CSCW, but focuses on collaboration and its specific effects on learning. In this section we describe some notions of CSCW and CSCL that are relevant for our work and how they relate to the roomware paradigm described in the previous section. We describe first some notions of CSCW that also apply to CSCL, and then we move to some notions that are more specific to CSCL.

2.2.1 Imitation bias

One aspect of CSCW research was concerned with computer-mediated communcation (CMC), i.e. using computers to facilitate communication between geographically distributed groups. When looking at the history of research in CMC, one might observe a thread of evolution that started with the development of groupware for geographically distributed groups that attempts to imitate as closely as possible real life communication. Research, however, showed that increasing media richness to more closely imitate face-to-face communication did not always improve effectiveness of the communication medium [Hollan 92]. This "imitation bias" also had another flaw in that it gave the impression that CMC can be, at best, as effective as face-to-face. This also was not the case as research started focusing on the advantages CMC had over face-to-face communication [Dillenbourg 08]. For example, chat systems provided a history of a conversation that is not available in face-to-face conversations. CMC was also more suited for automated analysis of interaction for a deeper understanding of the processes taking place during collaboration, as it was easier to generate communication logs and analyze them when this communication took place through a computer. Realtime automated analysis is particularly interesting as it permits not only post-hoc understanding of the collaboration process but also realtime adaptation of the interaction processes by participants or an external observer such as a teacher or a team leader [Dillenbourg 07]. This is elaborated on in Section 2.2.7.



Figure 2.6: Simple example of social awareness using an icon to indicate that the user is busy, and action awareness indicating that the user is currently typing. These are commonplace in today's chat applications.

The roomware paradigm broke free from the traditional model of computing of human-machine interaction via mouse, keyboard and monitor. This paved the way for face-to-face collaboration with computer support, but without the latter interfering with the natural human-human interaction by blocking direct eye-to-eye contact, for example with vertical computer displays [Prante 04]. Thus the third step of the evolution process was reached and research began exploring how we can introduce some of the advantages CMC had to offer back to the real world by augmenting face-to-face interaction with roomware technologies. Our current work falls precisely in that domain.

2.2.2 Awareness

Workspace awareness has been an interesting challenge for CSCW when dealing with distributed groups. While definitions for what constitutes awareness vary, researchers widely agree on the importance of the visibility of certain properties of group members and their interactions [Carroll 03, Dourish 92, Gutwin 95]. Different frameworks were proposed to define the varying types of awareness that are relevant to collaboration. Carroll et al. proposed three types of awareness: *social awareness* refers to knowing who is present during an interaction as well as their current state, *action awareness* refers to what the others are doing, and *activity awareness* refers to the global activity and how it is going. For distributed teams, even the simplest notions of awareness are sometimes difficult to achieve. Many types of groupware developed to address certain aspects of this issue: from simple chat tools that inform each user of the status of the others as seen in Figure 2.6, to more complex systems that allow one group member to know what part of the document the other group members are currently working on, Figure 2.7.



Figure 2.7: Gutwin and Greenberg proposed a Radar view (right, and top-left of left image) that informs each user what part of a collaboratively constructed concept map the other user is working on. [Gutwin 04].

In face-to-face collaboration, many types of awareness that groupware attempts to establish are naturally and constantly present through speech, gestures, eye-contact, peripheral vision, etc... [Greenberg 96]. However, this is not to say that groupware for co-located groups need not

address the question of awareness. There are many characteristics of a group's activity that are not directly observable, or perhaps not emphasized enough. Alavi et al. proposed for example an awareness tool for co-located groups in the form of a lamp that informs the teacher of a classroom as well as other students on the progress of each team in an exercise session [Alavi 09]. This information can be relevant for the those present in the room, but it is not directly observable. In fact, by making explicit information that may or may not be observed implicitly, the tool pushes participants to notice this information and reflect on its content, perhaps to take action when appropriate, for instance by one team seeking advice from another that seems to have finished an exercise the first team is stuck on. In addition, when the information is in fact directly observable, these awareness tools help offload the charge of remembering this information, in this case by the class teacher, and focus on performing one's task, only referring to this information when it is needed.

2.2.3 Grounding and persistence

Clark and Brennan define *grounding* as "the collective process by which participants try to reach a mutual belief" and describe it as essential to successful communication [Clark 91]. When involved in any collaborative task, participants need, to some extent, to continually maintain common ground on which to base their collaboration. The requirements for common ground vary based on the purpose of the collaboration, and the costs of grounding vary based on the medium of communication.

Common ground is usually constructed by communication, and in order for a certain piece of information to become part of the common ground, it must first be uttered by a participant and then received and understood by all others. This means that when speaking, a collaborator must not only ensure that they utter the phrase they want to transmit but must also ensure that the other party has received it properly. In face-to-face communication or any communication medium that includes video or audio, a collaborator can acknowledge the receipt of their interlocuter's message through backchannel feedback, either verbally by uttering short expressions such us "uh-un" and "okay", or nonverbally with gestures such as nodding their head. Although this does not ensure perfect understanding of the message, it does achieve some common ground that the message has been received and communication can proceed. The cost related to this type of grounding is extremely low. However, in communication media such as chat or email, it is more difficult to make this kind of backchannel acknowledgement, and a higher cost is needed to maintain proper grounding.

Clark and Brennan provide a set of eight dimensions for communication media that influence the cost of grounding. Among these, reviewability describes the ability for one participant to review the past statements of another, and it is capable of reducing the cost of grounding. This is not possible in oral or face-to-face communication, but it is possible in written communication such as chat or email. Dillenbourg and Traum later modify this dimension, preferring the term persistence to separate the fact that the messages remain visible from the fact that they were viewed in the first place, as is implied by the term *reviewability* [Dillenbourg 06]. Indeed, it is often the case that a message is transmitted, but never received, and a *persistent* message is one that can be viewed even after it is first transmitted. In addition, the term *reviewability* was chosen to describe the persistence of information in the form of messages to facilitate grounding; however information that needs to be part of the common ground need not be a product of communication. Grounding can be achieved on information that is presented to all participants from an external source such as a problem description or course materials displayed on a shared screen. The persistence of such information is helpful in reducing the cost of its grounding. Thus external representations of information that needs to be part of the participants' mutual belief not only reduces the cognitive load associated with remembering its content but also functions as a persistent referent that helps participants maintain common ground on its content [Kirsh 10]. For example, it would be easier for two students to collaborate on solving a problem when an external representation of the problem



Figure 2.8: The "interactions" paradigm [Dillenbourg 96] suggests an alternate path to studying the outcomes of collaborative learning.

description is available to them in some shared form, as the cost associated with building a common ground on what the problem is that they are meant to solve is significantly reduced.

2.2.4 The "interactions" paradigm of CSCL

Research on collaborative learning has evolved over the past few decades from studying collaboration in order to determine whether or not it is better than individual learning, to observing collaboration with the intent of determining when it is more beneficial than individual learning, to research aimed at manipulating collaborative processes in ways that foster better learning outcomes. Three paradigms in collaborative learning research were thus explored: the "effects" paradigm, the "conditions" paradigm, and the "interactions" paradigm [Dillenbourg 96]. In the "effects" paradigm, researchers tried to discover whether or not collaboration improves learning outcomes. In this case, there was only one independent variable, and two possible outcomes: collaboration either improves learning outcomes, or it doesn't. The result was a large body of contradictory evidence that led to the notion that collaborative learning *can* be more effective than individual learning, but only under certain conditions [Slavin 83]. What these conditions were was yet to be discovered.

This led to a shift in research into what was called the "conditions" paradigm where researchers in CSCL have been exploring collaborative learning contexts in an attempt to identify those that lead to better learning gains and develop tools that further improve learning outcomes. This proved to be a daunting task as the parameters were many and interacted with each other in complex ways. In addition to group size, age, gender, race, etc... of the participants, more complex variables such as group heterogeneity, individual member expertise, the features of the learning task itself and a good deal of interaction among all these variables made isolating them and studying them independently a near-impossible task.

This led to the what is known as the "interactions" paradigm that proposed again a realignment of focus in CSCL research: rather than attempting to discover conditions under which collaboration is beneficial, one could attempt to discover which types of interaction occurring *within* collaboration lead to better learning outcomes and try to elicit these types of interactions. As seen in Figure 2.8, the paradigm breaks down the complex question *under what learning conditions is collaborative learning beneficial (a)*? into two separate questions: *what types of interactions lead to better learning outcomes (b)*? and *how can these types of interactions be elicited within specific learning contexts (c)*?

2.2.5 Participation in collaborative learning

Researchers in collaborative learning have indeed observed that certain types of interactions are predictive of learning. In particular, students who engaged in elaborated explanation [Webb 91], argumentation [Baker 99], mutual regulation [Blaye 88] conflict resolution [Willem Doise 76] as

well as seeking and providing help [Webb 83] exhibited higher learning gains. We note that these types of interactions share a common theme: they are all based on active participation in the form of verbalization. Verbalization itself becomes a necessary, though not sufficient, predictor of a large class of interactions that in turn are predictive of higher learning outcomes.

However, in the context of collaborative learning, one cannot make the assumption that the more *an individual* speaks, the more the *group* learns. After all, given the generally exclusive nature of conversational turn-taking [Sacks 95], the more one member of a group speaks, the less the others will. Therefore, when looking at learning gains for the *group*, one must look beyond the notion that more verbalization leads to better learning as it is not possible to simultaneously increase the participation levels of all participants in a single group.

2.2.6 Participation balance

Cohen [Cohen 94] describes some criteria for group productivity, without which group learners might benefit less than individual learners. Among these, *lack of equity in participation* is presented as an obstacle to effective learning in a group. Salomon and Globerson also describe the debilitating effects of unbalanced participation [Salomon 89]. They describe two types of effects: the "free-rider" effect, in which an overparticipating member could cause other members to expend less effort on the common task, and the "sucker" effect in which underparticipating members could lead the more active members to lose motivation in the task and thus avoid being taken advantage of. In either case, group productivity decreases.

Cohen also suggests that the difference in participation is not necessarily related to participants' abilities or their expertise, but rather to their perceived status which can come from any number of stimuli including age, gender, social status or race of the participant. In some cases, perceived popularity or attractiveness of individuals can lead to more active participation on their part [Cohen 94, M. Webster Jr. 83]. Moreover, it was shown that the amount of one group member's participation in itself can lead to that member being perceived as having a higher status, thereby leading to even more unbalanced participation [Dembo 87].

Participation balance in collaborative decision-making

When decisions are made in group meetings, there is often a substantial risk that one or more participants who hold critical information are unable to effectively share this information [DiMicco 04]. Proper information sharing is thus a crucial aspect of effective decision making. In reality, however, the variety and number of participants who do in fact contribute to the decision-making process is often less than is deemed appropriate by post-hoc analysis [Huber 90]. As a result, decisions are made with some relevant and potentially critical information missing, leading to suboptimal results. This has been shown consistently in research on information pooling tasks [Stasser 03, Winquist 98, Greitemeyer 03], and is particularly interesting when critical information is only available to an informed minority that fails to share this information due to the conversation focusing on information known to all. This could be avoided if group members were encouraged to participate in a more balanced manner, permitting all members to contribute, and pushing the informed minority to share information even when it goes against the tide of the discussion. However, balanced participation certainly does not guarantee that the information is better shared, as members who would otherwise remain silent, might not use their participation to provide meaningful information. It is in that sense a necessary but not sufficient condition.

2.2.7 Group mirrors

A family of tools to guide collaborating participants towards more desirable behavior, such as balanced participation, are called *group mirrors* [Jermann 01]. Jermann et al. describe three types



Figure 2.9: A group mirror informing participants about the ratio of amount of communication to action taking place during the collaborative task. The representation of this group mirror includes embedded color-coded normative information on what constitutes a good ratio. [Jermann 04]

of computer support for collaborative learning. These vary depending on their level of active involvement in the regulation process. Coaching systems observe and interpret the collaborative setting and provide advice to the learners. Less active are metacognitive tools, that summarize to the users, via a set of key indicators, the state of the interactions taking place without giving advice on how to interpret or act on these indicators. Finally, mirroring tools simply reflect to the users their basic actions by informing them what each member of the group has done. By increasing their awareness of what they are doing, mirroring tools help members maintain a common representation of what is taking place in the collaborative process. The system we propose here is of the mirroring type. It displays to the users a basic representation of the actions they have taken without offering advice or interpretation on the state of the interaction.

Figure 2.9 is an example of such group mirrors. In this example, Billy and Christina are tasked with tuning a grid of traffic lights to minimize road congestion. The system counts how many times they communicate and how many times they take action, and displays the ratio of these in the form of a gauge. This group mirror helps the participants maintain awareness on how they are doing and repair or avoid situations where they take too many actions without first discussing them together.

2.2.8 Regulation

One of the benefits of group mirrors is that, in the presence of guiding norms on what constitutes "good" behavior, increasing awareness of what a group is doing can push the members of the group towards self-regulation [Jermann 04]. Figure 2.10 shows how this regulation takes place as a feedback loop that starts from the state of an interaction, which the system observes, collecting and aggregating data about it. The user then compares the resulting aggregation with an interaction standard, possibly causing the user to change their action which results in a new interaction state.

2.3 Automated Meeting Analysis

The field of conversation analysis formally began to take shape in the mid-sixties when Harvey Sacks started making references to it in parts of his sociology courses at UCLA [Sacks 95], but the use of computers to automate the analysis of conversation is much more recent.



Figure 2.10: The architecture of interaction regulation in the presence of a group mirror proposed by Jermann [Jermann 04] based on Carver and Scheier's general architecture for regulation [Carver 01].

Today, research in automated meeting analysis has advanced significantly and covers both low-level and high-level aspects of group meetings. One basic method of analysis, called speaker diarization, aims at breaking up a meeting into the individual turns of participating speakers. This is done in various ways, such as using probabilistic clustering methods to group samples of audio into single speaker clusters based on some voice features such as pitch and pitch variance [Huang 07]. Others use multiple microphone inputs in order to localize the source of the signal and use the source locations to determine speakers [Anguera 07]. Some researchers have also worked with multi-modal input, using both audio and video signals, to determine speakers in a meeting [Otsuka 08].

Other types of meeting segmentation have been attempted with researchers having some success in segmenting an audio stream into individual dialog acts (units of speech that constitute a single communicative act such as a declarative sentence, a question, etc...) [Ang 05]. Others attempted higher level segmentation and were able to break up a conversation into different "scenes" based on the frequency of turn taking events [Basu 02]. Classification of actions taken by members of a group was also approached using multimodal analysis of the meeting [McCowan 05].

In addition to analyzing the meeting as the object of interest, researchers have also focused on analyzing meetings with the aim understanding the roles of participants in those meetings. Using multi-modal cues, Hung et al. tried to determine the dominant person in conversations, basing their approach on speaker diarization [Hung 08]. Others attempted to classify participants according to their *functional roles*, a taxonomy of actor roles such as Orienter, Giver and Seeker in small groups [Zancanaro 06].

Finally, researchers have also approached high level analysis and attempted to discover complex social features of meetings such as interaction groups [Brdiczka 05], group interest level [Gatica-Perez 05], and influence between group members [Rienks 06]. This domain, referred to as Social Signal Processing [Vinciarelli 09], is an emerging field that aims at examining human interaction and human behavior from a social perspective by analyzing subtle cues such as facial expressions, body movements, and other non-verbal signals.

2.4 Voice, Prosody and Emotion

Nonverbal components of human voice account for a large part of how we understand speech [Scherer 80], and researchers in human-computer interaction hoping to build machines that can understand human speech have begun in recent years focusing their efforts on analyzing prosody. Prosody is the name given to a collection of voice parameters that determine *how* a person is speaking such as rhythm, intonation, stress, etc... Prosodic analysis has been used in automated analysis of human speech, sometimes in conjunction with automated speech recognition in order to improve understanding of spoken words. Classification of dialog acts has been shown to be improved when prosodic cues taken into account in addition to the verbal content of speech [Mast 96, Stolcke 00]. Other researchers have used prosody alone, i.e. without the help of verbal content, to segment speech into individual utterances [Ferrer 03].

Prosody has also studied in the context of emotion analysis, where researchers have determined that certain vocal expressions of emotion have discernable prosodic patterns [Banse 96,Scherer 03]. This has led to a large body of research on recognizing human emotions [Cowie 01] for various purposes and with varying degrees of success. One application of emotion recognition comes in the form of "emotion-sensitive" user interfaces that interact with a user differently based on their current mood [Polzin 00]. However, this type of interaction based on affect requires high accuracy in terms of emotion detection.

A different approach for affective interaction comes from Sundström et al. who propose a usercentered approach that does not attempt to automatically detect the user's emotion, which may be perceived by some as intrusive. Instead, the approach relies on the user to actively manifest an emotional state using physical movements without necessarily defining what particular emotion is being manifested [Sundström 05]. The user thus engages with the system manifesting the emotion they wish to display, and by physically manifesting that emotion they may end up reinforcing it in their actual affective state.

2.5 Related Works

Researchers in the field of Human-Computer Interaction have already done some work on influencing group conversation with mirroring displays. We present here a review of some of this work.

2.5.1 Second Messenger

DiMicco et al. have explored the effect of group mirror visualizations on speaker behavior in collaborating groups [DiMicco 07b, DiMicco 07a]. They have studied both the effects of having information displayed in realtime as the conversation takes place and of having this information displayed between meetings as a replay tool. In both the first and second versions of the system, the group mirror was projected on some shared surface, and sound was captured using head-mounted microphones. They built and tested different versions of a group mirror tool, which they called *Second Messenger*, and described some differences in effect between them. Their first tool, seen in Figure 2.11, showed a histogram of the group members' levels of participation in the task, projected on a wall, along with indicators that show what corresponds to over and underparticipation. Experiments on this tool showed that overparticipators reduced their participation when the tool was used, but underparticipators did not increase theirs, even though the underparticipators of the control group did.

The second version of their system, seen in Figure 2.12, included two components: a realtime visualization that took the shape of four circles whose size was determined by each group member's level of participation. The second component was a replay visualization displayed to participants in between two different tasks. It showed on the left the same circles showing levels of participation, in addition to a detailed summary of the speakers' turns.



Figure 2.11: The first version of the mirroring tool used by Dimicco et al. [DiMicco 05] was a histogram projected on the wall.

The replay tool had a significant effect on speaker behavior after it was displayed. Overparticipators spoke less and underparticipators spoke more. This desired effect was not completely achieved when only the realtime tool was used. By displaying information in real time, *Second Messenger* pushed overparticipators to reduce their levels of participation but the effect was not as strong for underparticipators.

Second Messenger showed promising results for mirroring displays as it indicated that these tools can in fact influence group behavior by promoting self-regulation.



Figure 2.12: The second version of the mirroring tool used by Dimicco et al. [DiMicco 05] included a realtime component (left) projected on the table and a replay visualization (right) shown to the participants between tasks.

2.5.2 Conversation Clock and Conversation Votes

Other researchers have also studied the effects of these visualizations. Bergstrom and Karahalios implemented two systems, the *Conversation Clock* [Bergstrom 07a, Bergstrom 07c] and *Conversation*

2.5. RELATED WORKS

Votes [Bergstrom 07b]. In both systems, a visualization representing the current conversation is projected onto some shared surface, and sound was captured using lapel microphones attached to the users' clothes.



Figure 2.13: The *Conversation Clock* displays a realtime snapshot of the conversation history on the surface of the table [Bergstrom 07a].

The *Clock* seen in Figure 2.13 shows which member of the group spoke at each time and allows the users to get a snapshot of the conversation history every time they look at the surface. This is done by creating a circle of colored bars that point towards the center. The more a user speaks, the more bars of their color appear on the outer edge of the circle, and the louder they speak the longer these bars are. Whenever the perimeter of a circle is filled, the circle moves towards the center and another circle is created in its place. The result is visual history of the conversation that highlights speaker participation levels.



Figure 2.14: *Conversation Votes* shows a summary of speaker participation in addition to whether or not the other participants agree with each utterance [Bergstrom 07b].

Conversation Votes, Figure 2.14 goes further and allows members of the group to anonymously "vote" indicating to the table whether or not they agree with what is being said. This information is visualized onto the table along with the speaking patterns of the users. This is done in a manner similar to the *Clock*, but using a straight line instead of a circle. Bars are color-coded for users, and votes of support a particular utterance has received are shown as small circles on the sides of the bar representing the utterance. The voting system only permitted positive votes as users of a previous experiment felt uncomfortable voting negatively on another user's contribution, even when it was anonymous.

Results of studies on these visualizations also showed behavioral changes among participants. Overparticipators reduced the length of their turns, and underparticipators increased the number of turns taken. The visualizations also took some attention away from the conversation, although participants reported that there was no loss in the quality of their interaction. Qualitative analysis

showed that participants were most interested in displayed information that was about their own contribution.

2.5.3 The metaphoric group mirror

Streng et al. made a comparison between two types of group mirrors that display the quality of argumentation in a group discussion [Streng 09]. One showed relevant information in the form of a simple diagram, whereas the other displayed the same information in the form of a scenic view where the quality of the interaction is translated into the weather conditions in the scene. Figure 2.15 shows these two representations. The quality of the argumentation was not evaluated automatically; an expert was judging the group's performance and silently signaling changes to the display.

The researchers showed that, in addition to 70% of the participants preferring the metaphoric group mirror over the diagram group mirror, the former was more effective in correcting deficient behavior both in terms of the amount of correction and the speed of that correction. They predict that, while the generalizability of the results depend on many factors not the least of which is the quality of the metaphoric mirror, the value of a metaphoric representation compared to a simple diagrammatic one cannot be ignored.



Figure 2.15: The metaphoric group mirror, seen in four different weather conditions (left), and the diagram group mirror (right) are projected on the wall during the meeting [Streng 09].

2.5.4 Meeting Mediator

Meeting Mediator (MM), developed by Kim et al., also presented participants of group meeting with a realtime visualization describing their conversation [Kim 08]. However, rather than display it on a shared space, MM displayed the visualization on handheld devices that participants had in their possession. MM collected, in addition to high-level speech features, body movements and distance to other users, among other properties. Seen in Figure 2.16, MM displays a mirror of the group dynamics indicating how much each member is speaking, how balanced the conversation is, and how much interactivity is exhibited among the members of the group. This is accomplished using a colored circle, whose color indicates degree of interactivity and position indicating balance, and lines connecting the circle to four nodes representing the users, with the thickness of the edge between the circle and each node corresponding to that user's amount of speech.

Of particular interest to this kind of display, as opposed to the two previous examples we described where the information is projected on a shared space, is that MM can be used for groups that are not collocated, because both the capture and display of information are done independently for each user and controlled by a remote server.

Experiments on MM showed that the presence of the device reduced the difference between dominant and non-dominant speakers by making everyone enthusiastic and energetic.



Figure 2.16: The *Meeting Mediator* displayed mirroring feedback on the small screens of individual users' handheld devices.



Figure 2.17: Leshed et al. used different types of visualizations to give feedback to participants in an online chatting conversation [Leshed 10]. On the right, an alternative visualization was used in place of the simple bars (bottom left)

2.5.5 GroupMeter

Groups mirrors for conversations have not been limited to face-to-face conversations. Leshed et al. incorporated *GroupMeter* in an online chatting tool to provide its participants with feedback on their involvement in the conversation [Leshed 10]. Given that it is much easier to extract information related to the content of communication when using a chat tool than in face-to-face communication, *GroupMeter* was capable of providing higher level feedback to participants. The system, seen in Figure 2.17, displays information about the number of words spoken, the number of words that have emotional significance, the number of self-references, and the number of times a user agrees or disagrees with the others. In one version of the system, some of these values are displayed directly beneath the chat window. In the other version, a graphical representation using a metaphor of a school of fish was used, with the size of the fish indicating the amount of speech and the closeness of the fish to the center indicated group cohesion as evaluated from self-references and amount of agreement among team members.

The results of user studies on this system indicated that in the absence of normative guidance, participants were not sure whether the large values for some of the information displayed was good or bad. For instance, some thought that referring to oneself was a good thing and others thought the opposite, leading to self-regulation being inconsistent among members of the group. It was also shown that participants preferred the metaphor visualization over the simple bars.

2.6 Summary

In this chapter we introduced the main domains of research our work falls in as well as the relevant concepts within these domains. We also introduced the different types of group mirror displays for meetings developed in the recent years by other researchers. These group mirrors varied in their design, their implementations and in the research questions they raised. In the next chapter, we present our own system, the *Reflect* table, as well as the research questions we will address in this thesis.

The *Reflect* table differs from these other displays primarily by being a self-contained roomware device that incorporates its own microphones and display. Our research questions also differ in that we go beyond studying whether or not the table has an effect, and attempt to understand how this effect comes about and under what conditions. Finally, we also extend our research questions beyond the scope of laboratory evaluations of a new technology by exploring the potential for this type of device in the real world.

Chapter 3

Reflect: Design Process

"Technological advance is an inherently iterative process. One does not simply take sand from the beach and produce a Dataprobe. We use crude tools to fashion better tools, and then our better tools to fashion more precise tools, and so on. Each minor refinement is a step in the process, and all of the steps must be taken."

CHAIRMAN SHEN-JI YANG, "Looking God in the Eye", from Sid Meier's Alpha Centauri

Reflect is an interactive table designed to provide realtime feedback to members of a group about the dynamics of their face-to-face collaboration, with the aim of promoting certain kinds of behavior in conversations. As it is impossible to produce, in one attempt, a complex system that is correct in both its objectives, its design and its implementation [Lindgaard 94], the *Reflect* table has undergone a long process of evolution that saw changes to its hardware components, its core functionalites and the technologies that implement them as well as its objectives and some of its design principles. In this chapter we look at the history of *Reflect* and trace back the steps in the iterative design process that led to its current state, as well as the reasons behind the decisions taken in that process.

3.1 The Principles of *Reflect*

The *Reflect* table was originally designed as a tool that measures "noise per team member." Its main functional objectives were three-fold: to capture, to analyze and to visualize the conversation taking place. Its research objective was to study the effect of this feedback on individual behavior and on group regulation. Each of these objectives imposed certain design requirements which we describe here:

- Capture: The table required some audio input capable not only of distinguishing the voices of group members from background noise, but also of distinguishing the voices of individuals from others in the same group. The obvious conclusion was that microphones were to be used; however it was the configuration of microphones that was not easily decided. Headmounted or lapel microphones, directional table-mounted microphones and microphone arrays were among the options considered.
- 2. Analyze: Audio streams do not generally lend themselves easily to meaningful visual representation. The table thus needed to incorporate some analytical functionality that collects a stream of multi-channel audio and extract relevant data that can be visualized intuitively. There was no intention of analyzing the semantic content of group members' speech, but rather our interest was in the conversational structure of the group meeting. Turn-taking patterns, turn duration and participation levels were the original focus of our interest. Prosodic features of speaker voices were not analyzed until the later stages of the design lifecycle.

- 3. *Visualize*: Of main concern was the granularity of how the table visualizes information, with current display technologies offering the capability of incorporating a high-resolution screen into the surface of the table. This would allow very detailed representation of the conversation to be displayed. Alternatively, low-resolution displays have more limited expressivity, but make up for it in the simplicity of the visualization and are often less intrusive.
- 4. *Study*: Finally, we wanted to study the effects of the table by experimentally comparing group behavior around the table against behavior around a regular meeting table. We were therefore interested in reducing to a minimum the changes in how the augmented table is used compared to its traditional counterpart. By limiting these changes to the augmentation itself, we would be able to attribute any subsequent changes to the users' behaviors to the table display and not some other factor.



Figure 3.1: The proposed theoretical architecture of the system.

3.2 The Design Invariants of *Reflect*

With the above principles in mind, we embarked on an iterative process of design that led to the development of four versions of *Reflect*. Although the table evolved considerably over its different design cycles, many of what would later become the fundamental design invariants of the table were already established prior to the first version.

3.2.1 Information to visualize

Early on in the design process of the table, it was decided that the table would capture, analyze and process low-level information about the dynamics of the group. The original goal of the table was to promote balance in collaboration, so the primary focus was on individual group members' levels of participation. Thus the information displayed was first limited to the quantity of speech each participant produced during the discussion. The last version of the table, however, enriched this information by displaying the level of engagement in the task based not only on the amount of participation but also on prosodic speech patterns, as is described in Section 3.4.4.

3.2.2 Mirroring

The aim of *Reflect* is to function as a mirroring tool for collaborative groups. As described in Chapter 2, the term mirroring tool refers to the informative, rather than normative, nature of the system [Jermann 01]. Mirrors do not tell their users what they are doing right and what they are doing wrong, in the same way that bathroom mirrors do not tell a user if their hair looks good or not. They simply show them a reflection of their current state, and leave it for the users themselves to decide what, if anything, needs to be changed. In the same manner, *Reflect* is not meant to judge the quality of the interaction, nor is it meant to actively pursue a certain kind of collaborative behavior on the part of its users. Its role in that respect is to inform the users of the current state of the conversation, and it is up to the users to decide what needs to be done.

The rationale behind this is that conversation, in essence, in a complex social phenomenon that is not disposed to algorithmic analysis. It is therefore neither desirable nor feasible with current computational models to produce a system capable of evaluating, without prior knowledge about the context of the conversation, how well participants are doing. For example, although participation balance is beneficial in many contexts as described in Chapter 2, there are instances where one speaker is expected or even required to participate more than the others, such as meetings where members have different roles or different fields of expertise.

Our system would thus remain neutral in terms of its judgment of the situation, and its role would be strictly informative. We cannot deny however that by explicitly making available certain information we are potentially inducing an implicit norm among at least some members of the group about how or if this information should monitored and therefore controlled. In fact, by not providing any explicit normative guidance, this may result in inconsistent norms being developed by different members of the group, each having their own interpretation of what the visualization is trying to say [Leshed 10]. In addition, even if the information displayed is inherently neutral, the actual visualization used to represent this information can induce a specific norm, as will be described in Section 3.5 with the territorial metaphor.

3.2.3 Unobtrusive input

A key condition we wanted the augmented table to satisfy was that it should be used in as natural a manner as possible. In other words, users would be able to sit around the table and immediately begin their meeting without having to perform any additional preparation to ensure the proper functioning of the table.

As such, we found that head-mounted and lapel microphones to be too intrusive as they required extraneous steps to set-up before and after each meeting. In addition, attaching microphones to the users' bodies would be an additional variable that might alter user behavior in ways that are irrelevent to the goals of our research. A single microphone solution that distinguishes users based on their voice patterns was also ruled out as the technology was not yet ready to identify multiple speakers in a natural context where significant overlap in speech occurs. Moreover, current reliable voice recognition solutions often require training which also goes against the natural use of the table; people do not generally need to train their table before the first use.

Two solutions were thus explored. The first system used one microphone per-user placed at the level of the table. This was used in the first two versions of the table and described in more detail in sections 3.4.1 and 3.4.2. The second system consisted of a beam-forming microphone array placed at the center of the table. It was used in the two latter versions and is described in Section 3.4.3.

3.2.4 Shared visible output

The fourth design invariant was that the display of the table should be easily accessible and shared by all. We wanted the display to remain within the field of peripheral vision of the participants so that they do not need to actively seek the information displayed. Instead, members of the group would be able to perceive the information at all times with minimal additional physical and cognitive effort.

The shared nature of display was meant to reinforce the notion that the display is a mirror for the group, not for the individuals within the group. In addition, seeing information displayed in front of everyone makes it difficult to ignore when this information shows that a change in behavior may be needed.

This led to the conclusion that the information would be displayed on the surface of the table itself, but not in a way that makes the surface unusable as a regular table surface.

3.2.5 Low-resolution display

When a group of people decide to sit around a table for a discussion, there is often something in particular they need to be doing or talking about. A table that is meant to help them in that task should not then take their attention away from their real goals. This was the main reason behind excluding the possibility of showing detailed and precise information about the discussion. Such a precise display would require group members to spend some time and cognitive effort analyzing and interpreting what the table is showing, and this time and effort would be taken away from the actual task. Thus, in order to avoid further increasing the cognitive load members of the group need to cope during their collaboration, the display of the table would remain as simple as possible while giving out only limited information.

3.2.6 Minimal interactivity

The same rationale for limiting the resolution of the display also applies to the way in which participants can interact with the table. Allowing users to interact directly with the table, via buttons or other controls, would increase the risk of distracting the users from their own task. A user who is manipulating some functionality of the table through its interface would likely be less attentive to the conversation taking place. It was therefore decided to reduce the interactive components of the table to a bare minimum, adding interactivity only as needed.
3.3 Reflect as a Semi-Ambient Display

We introduced *ambient displays* in the previous chapter as a way to utilize the peripheral vision of a user by displaying information in the form of environmental cues that the user is aware of, but not necessarily attentive to. The *Reflect* table also lies in the background of the users' task so that they are not directly monitoring it; however, the position of this display is in the very center of the users' workspace, the surface of the table, enhancing its visibility. This trade-off in the display, between being unobtrusive and in the background of the task but still visible and central to the relevant physical space, is what we define as being *semi-ambient*.

3.4 The Four Versions of Reflect

In the course of four years, four different versions of the table were developed, each providing insight and lessons for the design of the next. Only the two latter versions underwent a thorough evaluation, whereas the first two were tested with brief exploratory studies. In addition to specific hardware described for each version of the table, all tables used a standard PC running Microsoft Windows, in earlier versions, and Ubuntu Linux, in the last version, as well as a multichannel soundcard connected to pre-amplifiers for each microphone. All versions of the table were controlled by software written in the Java programming language.



Figure 3.2: The first prototype: the Virtual Noise-Sensitive Table.

3.4.1 The Virtual Noise-Sensitive Table

The name *Reflect* was not given to the augmented table until later stages of development. In the meantime, the earlier prototypes were given the more descriptive name: *Noise-Sensitive Table* (NST). The first of these NSTs was a regular wooden table with several holes around the edges where large dynamic microphones were fitted. Each microphone was placed at a position where one user would be seated. A patch of coated white paper was fixed on the center region of the table functioning both as a projection screen and a makeshift white board. Overhead, a metallic crossbar



Figure 3.3: The second prototype: the Physical Noise-Sensitive Table.

held a projector that beamed down a visualization of the conversation onto the white board. The low-resolution display was simulated by projecting information using an 8x8 matrix of small colored circles. The speaker was determined by thresholding the input levels on the individual microphones. This however, only works well when speaker volumes are predictable.

Seen in Figure 3.2, this first version was developed as a Masters project by Guillaume Raymondon and was used as part of a course on Computer-Supported Cooperative Work. The table was evaluated by students of that course with a short user study. The work was supervised by Jean-Baptiste Haué.

The results of the study informed the future designs of the table on the following aspects:

- The use of individual microphones for each member of the group required some tuning, and while it did provide decent detection of speaker when properly tuned, it did violate the principle of natural use of the table. Tuning had to be done too often to make the table usable in a natural manner.
- The positioning of the microphones was also a problem. The microphones were placed over the surface of the table in front of participants. This made them highly obtrusive since they prevented users from placing documents, laptops or other items in the space directly in front of them.
- When used in a well-lit room, the projected display was not clearly visible. This required the table to be placed in a dimly lit room causing a radical change in the environment in which a regular table would be used, and would have adversely influenced the validity of any study conducted on the table.

3.4.2 The Physical Noise-Sensitive Table

The second version of the NST addressed the issues encountered in the first prototypes and led to new insights on how to further imporve the design. In addition to the structural change, replacing the old wooden surface of the virtual NST with a new body, the projected visualization was replaced with an 8x16 matrix of Light-Emitting Diodes (LEDs) embedded into the center of the

table. The LEDs were installed on eight individual printed circuit boards (PCBs) with 16 LEDs on each. These boards, controllable via Universal Serial Bus (USB) connections, were custom-built for the table by René Beauchat of the Processor Architecture Laboratory at the Swiss Federal Institute of Technology. The display was covered with a frosted glass surface that blurred the visualization and further reinforced the low resolution of the displayed information. The same microphones as the previous version were used as no better solution had been found at that point. However, in order to avoid the problem of obstructing the use of the table with the microphones, these were placed under the surface of the table pointing upwards towards the speakers. The result was a hidden and unobtrusive set of microphones that freed up the surface of the table for use by the members of the group. In addition, the LED-based display was more luminous than its projected counterpart and was thus clearly visible even in well-lit rooms.



Figure 3.4: One of eight printed circuit boards containing sixteen multi-color light-emitting diodes used in the second, and future versions of the table.

While the second NST successfully addressed the main issues of its predecessor, it had issues of its own. Of primary concern was that the threshold-based input, which required tuning when the microphones were over the surface, was nearly impossible to calibrate when microphones were no longer capable of directly capturing the users voices. Microphones placed under the surface of the table were much less discriminative of different speakers.

The physical NST was thus closer to our vision of an augmented table with unobtrusive input and embedded ambient display, but it simply did not function as required. It was therefore time to re-examine the system for audio input and find a solution that satisfies both the principles of the table design and its functional requirements. Our needs were met by another laboratory at our institute that specializes in acoustics and audio processing, and a collaboration between our groups led to the next version of the system, the *Reflect* table.

3.4.3 The First Reflect

The third version of the augmented table, which at that point acquired the name *Reflect*, was the first fully functional prototype. It inherited its display from its predecessor, the physical NST, in the form of an 8x16 matrix of multi-color LEDs. Its physical structure was redesigned into a metallic skeleton with a sturdy frosted glass pane covering the entire surface of the table. The new structure, the work of industrial designer Martino d'Esposito, was meant to be easily reproducible as it was made almost entirely of standard issue aluminum beams. The skeleton of the table also included a compartment large enough to include all necessary hardware needed for the table to function. The electronic components were easily accessible when needed by simply lifting the glass pane on top



Figure 3.5: The first version of Reflect, successor of the Noise-Sensitive Table.

of the table. The main change the *Reflect* table underwent with respect to its predecessors was its method for input.

Beam-forming microphones

The input configuration of the first version of *Reflect* was the most significantly altered component with respect to its predecessors. The individual microphones that required cumbersome tuning and calibration were replaced by an elegant three-microphone array capable of reliably determining the direction from which sound is coming. A microphone array is a multi-microphone configuration coupled with audio processing software that is capable of performing functions single traditional microphones are incapable of doing. The array used in *Reflect* is referred to as a beam-forming array, as it is capable of creating several beams in the region around it and filters sounds into different channels depending on which beam they arrive from [Faller 10]. This solution, developed



Figure 3.6: A triangular configuration of three small boundary microphones is capable of determing where the sound is coming from.



Figure 3.7: The basic system architecture of the *Reflect* table.

by Christoph Faller at the Audiovisual Communications Laboratory of the Swiss Federal Institute of Technology in Lausanne, provided a perfect solution for our microphone worries. In place of the bulky dynamic microphones used in previous versions of the table, three compact boundary microphones, fitted in a small triangular configuration at the center of the table, were enough to reliably determine which person around the table was speaking at each instant. Boundary microphones, seen in Figure 3.6 are small microphones that benefit from their closeness to a flat surface in order to amplify the audio signal they receive [Davis 06].

This input configuration provided the necessary capabilities in terms of speaker detection while maintaining an unobtrusive presence of microphones. The small, almost flat microphones placed at the center of the table took up very little space and freed up the areas in front each user. They were discreet enough to often be mistaken for decoration, and many of those who encountered the table did not even know there were microphones on the table until these were pointed out. This likely made the microphone array easy to ignore when the table was being used.

System Architecture

The first *Reflect* was a working prototype that allowed us to determine the speaker in a reliable manner. Its architecture, seen in Figure 3.7, was also the basis on which all studies and additional functionalities, which will be described in Chapters 4, 5 and 6, were built. The table and the subsequent version consisted of a microphone array for input and a LED matrix display for output. It maintained a model of its users that contained relevant information such as location and participation levels, and later prosodic data. The beam-forming module analyzed the raw audio data stream produced by the microphones and generated a speaker at each instance. The renderer module retrieved information from the user model and produced a visualization that it then sent to the LED controller which in turn displayed it to the users. Two additional modules will be introduced to this architecture in Chapters 5 and 6. In Chapter 4, we describe the first complete user study we conducted that evaluated the effect of this version of *Reflect* on user behavior.

3.4.4 The Second Reflect

The user study conducted on the first version of *Reflect* and described in Chapter 4 showed promising results in terms of the effect this table can have on user awareness and their behavior. The next step was two-fold: to expand the capabilities of the table and to widen the scope of its use. Indeed, while the new version of the table differed the least from its predecessor in terms of hardware, it underwent the biggest changes in its functionality and usability since the conception of the first NST. The changes to the capabilities of the table involved primarily a new mode of input analysis, and the scope of its use was expanded to include situations where the table is used as an integral part of an activity, rather than a background peripheral.



Figure 3.8: Nearly identical in form to the first version of *Reflect*, the second version differs considerably from its predecessors in its functionalities and the intended scope of use.

Voice Analysis

One of the clearest limitations of the original table design is its inability to distinguish different kinds of speech. It indiscriminately recognizes any kind of noise heard around it as speech and treats it in exactly the same manner. This understandably led to the often-asked question: shouldn't the table be more concerned with the quality of a speaker's contribution rather than the quantity of noise they make? We thus began to explore different methods for analyzing voice with the objective of finding prosodic attributes of voice that are both meaningful and computable. This resulted in integrating a voice analysis system into *Reflect* that is capable of automatically determining the vocal arousal of the speaker, which itself is an indicator of their perceived emotional implication. This system is described in more detail in Chapter 5.

Scope of Use

The original intended scope of the *Reflect* table was limited to what we refer to as casual collaborative learning, i.e. situations in which learners come to work together in unstructured, unsupervised groups. However, as the table started getting exposed to more and more people from different domains, it became clear that there are contexts other than collaborative learning where the table could be of use. It became even more so when voice analysis was integrated into the system. Banks, human resource personnel and, most prominently, communication training centers showed interest in using *Reflect* technology. However, some of the original choices and principles in the design of *Reflect* were less adapted to the new contexts of use and thus needed to be re-examined. The changes *Reflect* underwent during this process were not instantaneous and involved several iterations until they reached a state that was satisfactory to ourselves and its new users. Among these changes, described in more detail along with the process that led to them in Chapter 6, the most significant was a button interface for directly interacting with the application running on the table. This change, necessary for situation in which the table is used more directly as a tool for a specific task, added a functionality that had been deliberately left out because of the *Minimal Interactivity* principle described in Section 3.2.6 of this chapter.

3.5 Visualizations

All four versions of the table used a low-resolution display of moderately spaced pixels, implemented either as LEDs or as projected dots. While limited, this display offered a wide range of possibilities on how to display information. We describe here the different modes of visualization, shown in Figure 3.9, that were implemented on the table.



Figure 3.9: Six-person visualizations: territorial (left), column (center) and conversation trail (right).

• *Territorial.* The first visualization, dubbed the territorial visualization, displayed information in the form of colored territories of light in front of each user that expand and contract based on certain attributes of that user's involvement in the conversation. When the table displays the amount of participation of each group member, this information is portrayed in the sizes of the territories. Members of the group that speak a lot, have large territories, and those the speak little have small territories in front of them. The borders between territories shift over time based on the relative values represented by the adjacent territories. The display could be parameterized to show either the entire history of the conversation, or a fixed duration window.

The use of this territorial metaphor was found to be appropriate as it had the potential to create a natural inclination towards more balanced conversation in certain contexts. This disposition to balance would come from human perception of territories and the social protocols associated with them. For example, it is not considered socially appropriate to use portions of a shared space that are directly in front of another group member [Scott 04]. Thus, a mismatch between the distribution of territories as visualized by the table and the group members' notion of how these territories should be distributed could lead to some members to try to rectify it.

The territorial visualization was the only visualization implemented until the advent of the third version of the table, i.e. the first *Reflect* table.

- *Column.* In the column visualization, information is displayed in the form of a histogram, with each member of the group represented by one or two columns of LEDs (depending on the number of users). This visualization was particularly useful to display spatially neutral information, for example information that is not related to one particular member of a group, but to the group as a whole. It is also easier for users to attribute absolute values to columns than to territories because, unlike territories whose sizes depended in part on the sizes of neighboring territories, column sizes are determined independently and based only on the value they represent. It is thus easier to compare values represented by each column than is the case with territories.
- *Conversation Trail.* The third implemented visualization did not have the general purpose nature of the two others, but rather was restricted to a single type of displayed information. The conversation trail displays a ball of bright pixels that continually follows the position of the current speaker, and moves from one speaker to the next as speaking turns proceed. As it moves, the ball leaves a soft trail of light behind it, which over time, draws a graph of the conversation, highlighting members of the group between which frequent exchanges take place. Research has shown that certain roles and combinations of roles for members of a group often result in certain shapes of conversational graphs [Pléty 96].

3.6 Summary

Reflect is a table for group conversations designed to display visualization the provides participants with realtime feedback about their behavior. The table is meant to augment the meeting without interfering and as such remains in the background of the interaction. It evolved from a simple projected display into a specially designed hardware with a beam-forming microphone array. The next three chapters describe three different evaluations the table underwent and the lessons learned from these evaluations. Figure 3.10 summarizes the four versions of the table and their properties.

	Version 1	Version 2	Version 3	Version 4
Display	Projected	LEDs	LEDs	LEDs
Number of mics	Four	Four	Three	Three
Type of mics	Dynamic	Dynamic	Boundary	Boundary
Position of mics	Over surface	Under surface	Surface center	Surface center
Detects speaker	With calibration	No	Yes	Yes
Voice analysis	No	No	No	Yes
Buttons interface	No	No	No	Yes
Operating System	Windows XP	Windows XP	Windows XP	Ubuntu Linux

Figure 3.10: Summary of the different properties of the four versions of the table.

3.7 Research Questions

Our work in this dissertation builds on all the concepts and notions defined in the previous chapter, and uses the system described in this chapter to address questions on the role of computers in influencing face-to-face collaboration.

In terms of regulation, we address the following question:

1. Does Reflect influence user behavior and promote self-regulation?

- 2. If self-regulation does take place, how does it come about and under what conditions? What are the limitations of group mirrors like *Reflect* in this respect?
- 3. Can we influence human face-to-face collaboration without distracting participants from their primary task?

We then push the boundaries of the technology we developed beyond the confines of the laboratory and examine the scope of its usage in the real world. We ask the following questions:

- 4. Is there a role for *Reflect* beyond self-regulation in face-to-face collaboration?
- 5. What are the design implications for deploying such a device in the real world?

These questions will be addressed over the remainder of this thesis in two laboratory studies and one real-world experience.

Chapter 4

Study 1: Participation Balance

"Most conversations are simply monologues delivered in the presence of witnesses."

MARAGARAT MILLER

A user study was conducted on the first version of the *Reflect* table that was described in Section 3.4.3. The aim of the study was to evaluate the table on two criteria: whether or not it is able to improve group members' awareness about the conversation they are taking part in, and whether or not it is capable of pushing members of the group to alter their behavior.

4.1 Motivation for Participation Balance

In Chapter 2, we described the benefits of balanced participation both in the context of collaborative learning and group decision-making. We described how unbalanced participation can lead to lower learning outcomes for some group members, loss of group motivation, and suboptimal decisions.

To illustrate some of the issues of unbalanced participation, we present a small study conducted with eight subjects divided into two groups of four. We gave the subjects a task in which they were asked to rank, individually at first and then in group, a list of 15 objects in order of their importance for survival in the desert. They were given 10 minutes to complete the task individually. They were then asked to discuss the problem for 30 minutes and come up with a single ranking that they all agree upon. This type of task, known as a choice shift task, is used to determine the influence each member of the group had on the group's final decision, by comparing the group decision with the initial individual decisions of each member. The choice shift for each member is the distance between their initial decision and the final group decision. The most influential member is the one who has the smallest choice shift. In this task, the choice shift was computed as the sum of the differences between the rank given by the individual and that of the group for each object. It ranges from 0 (identical to group decision) to 112 (opposite to group decision).

We measured the individual members' participation in terms of their total talking time during the group discussion phase. We compared that to the individual choice shift. In both groups, one member clearly dominated the discussion as can be seen in Figure 4.1. It is important to note that in both cases, the individual rankings made by the dominating speakers before the start of the discussion were, according to experts, relatively poor when compared to some of the original decisions made by members of their group. This indicates that the dominant speakers did not have more expertise on the topic of discussion than other members. Interestingly, in both situations most participants, including both of the dominating members, were not aware that the conversations they had were not balanced. Moreover, when asked, they were not able to determine which member did in fact dominate the meeting.



Figure 4.1: Participation of group members in the choice shift task.

We drew two conclusions from the study. The first is a confirmation that difference in participation is not necessarily attributed to difference in level of expertise, in that the more expert peer would participate more. The second, surprisingly, is that it is not always obvious for members of a group who it was that spoke more than the others, even when one speaker dominated the conversation significantly.

4.2 User Study on Reflect

Our first user study on the *Reflect* table focused on the question of participation balance. We thus went on to explore if, by displaying information about participation levels to members of a group, the table is able to (1) increase their awareness of their respective level of participation, and (2) aid members of the group in having more balanced discussions.

In addition, we wished to evaluate the physical design of the table itself, mainly in terms of the unobtrusiveness and the visibility of the table described in Sections 3.2.3 and 3.2.4 as design

invariants of the interactive table.

Formally, we wanted to examine the following hypotheses.

From a design perspective:

- H1: The display of the table is visible and is looked at by the participants during the discussion.
- H2: The display of the table is discreet, unobtrusive and does not distract from the actual task at hand.

A balance between these two properties of the display is needed to make the table suitable for real-world use.

From a behavioral perspective, we examined if:

- H3: Individuals are more aware of their own and their partners' levels of participation when using *Reflect* to display those levels. By validating this hypothesis, we would be able to conclude that the information displayed on the table is seen and assimilated into the user's mental model of the conversation taking place.
- H4: Groups that are shown their levels of participation on *Reflect* are more balanced than those that are not. By validating this hypothesis, we would conclude that having this information displayed on the table promotes participation balance and helps the participants reduce over or underparticipation.

4.3 Experimental Method

We describe here the details of the user study in which we evaluated our hypotheses.

4.3.1 Description of the experiment

Groups of four subjects were randomly selected from a pool of bachelor students that had volunteered for the experiments. The study included 18 groups (72 subjects - 44 male, 28 female). All-male, all-female and mixed groups were used. Subjects were paid 50 Swiss Francs (around 45 US Dollars) for their two-hour involvement in the experiment. The groups were asked to solve a murder mystery task offered to us by Stasser and Stewart [Stasser 92].

The task materials were translated into French and adapted for groups of four. In this task, each subject was given a copy of investigation logs that included maps, interviews and a snippet of a news article. They were asked to accuse one of three suspects of having committed the murder. Each individual version of the investigation logs contained certain important pieces of information that were not available in others. This information was in the form of additional lines of dialogue that were woven into the remainder of the text, making it impossible for the reader to determine what information is available to others and what isn't. This ensured that all subjects were required to participate in the discussion in order to gather all the necessary information. This type of task, referred to as a hidden profile task, is often used in experiments involving group decision-making and information pooling [Stasser 92]. An excerpt of the task material is seen in Figure 4.2.

4.3.2 Experimental conditions

We used two experimental conditions that were identical except for the content of the information displayed on the surface of the table. In the first condition, the subjects were shown their levels of participation i.e. how much time each student spent talking. This condition will be referred to as the *speaker condition*. In the second, they were shown the focus of the discussion, i.e. how

Lt. NI.:	You play golf with Mr. Malone on Saturday morning. Right?
R. R. :	Yes, I do. We have a regular foursome.
Lt. M.:	Can you tell me anything about his relationship with
	Mr. Guion?
R. R. :	They were always good friends until these last few weeks.
	They had some sort of business disagreement. Mickey
	wouldn't say a whole lot about it, though. They've
	had problems in the past, but it's never been this bad.
Lt. M. :	What time did Mr. Malone arrive at the golf course
	last Saturday?
R. R. :	Around 7 as usual.
Lt. M.:	Ok, I appreciate your help.

Figure 4.2: Excerpt from the task material where the inspector Lt. Moody interrogates Rick Rooney, one of the victim's golf partners, about one of the suspects, Mickey Malone. The two italicized lines were only included in one version of the investigation logs making the key fact about the suspects time of arrival at golf available only to one participant.

much time was spent discussing the case of each of the three suspects in the murder mystery. This condition will be referred to as the *topic condition*.

We note here that we are not particularly interested in observing the effect of a topic visualization on the behavior of groups. Displaying information about topic balance serves the purpose of having a situation against which we can compare the effect of the speaker visualization. To that effect, the topic condition is a control condition and could have been replaced with a condition whereby no visualization is displayed at all. However, we chose the topic visualization instead in order to counter the effects of novelty and potential distraction that the speaker visualization would have had compared to a condition where nothing at all is displayed on the table.

In both conditions, the columns visualization was used. In fact, the choice of visualization was motivated by the need for a single visualization that can be used for both conditions. Although the territorial display may have been more suitable for displaying speaker levels, it is not at all suited for displaying the time spent on each topic since, unlike the speakers, the different topics do not have a meaningful spatial position that would justify the location of their corresponding territories. This was not a problem for the column visualization as columns were spatially neutral. By labeling the columns with white stickers posted on both ends of the table, we were able to attribute any kind of information to what each column represents. In the topic condition, the names of the individual suspects were used to label the columns, and in the speaker condition the first names of the subjects were used. In both conditions, name tags were placed at the corners of the table in order to familiarize the users with their parteners' names. This was essential in the speaker condition so that users are able to identify each of their parteners' columns. Both conditions were thus made as similar as possible to one another, with the exception of the actual information displayed on the surface of the table.

Participation levels were detected automatically by the table using the beam-forming microphone array described in Section 3.4.3. The subject of discussion was determined using the "Wizard of Oz" technique i.e. with a human listening to the conversation as it took place and remotely signaling the topic of discussion to the table.

A third neutral condition, in which no information is displayed on the table, was not included in the design of the study as it would have been quite costly, and the benefits of having such a condition were not compelling enough.



Figure 4.3: The visualizations used in the first study. A four-person version of the column visualization was used in the speaker condition, with each column corresponding to the speaking time of one participant. A similar three-column visualization showed the group members how much time was spent on each of the three suspects. In both conditions, the columns were labeled at both ends of the table with participants' and suspects' names.

4.3.3 Experimental procedure

The subjects were first asked to sign a consent form informing them that the purpose of the experiment is to evaluate novel collaborative tools and that the experiment was not intended to measure their own skills or abilities. They were also made aware that they will be filmed, and their conversation would be recorded via the microphones in the table. The subjects were then asked to read the investigation logs individually for 30 minutes, during which the table was used as a simple timer that kept the subjects informed of the time remaining. This was accomplished by using the LEDs of the table to display a progress bar that starts lighting up on one side of the table and gradually reaches the end when the 30 minutes are up. The subjects were allowed to annotate their copies of the logs and were told that they would keep the copies with them during the discussion. At that point, the subjects were not yet informed that their copies of the investigation logs contained information that was not available to others.

They were then given 60 minutes to reach consensus on a suspect. In order to start the discussion, the subjects were asked to come up with possible means, motive and opportunity for committing the crime for each suspect. They were informed that, in order to accuse a suspect, they must be convinced that all of these three elements pointed against him and that the other two suspects were missing at least one of the elements. The subjects were then made aware that they may possess unique information that is not available to others. In addition, they were told that they were not permitted to give their copy of the investigation logs to another participant and that each participant was only allowed to read from his or her own copy. This was to avoid a common strategy we observed in our pre-experiments where subjects, upon realizing that the unshared information that is not shared, thereby defeating the purpose of the experimental setup, which is to create a discussion. Finally, the visualizations were explained to the subjects, but no mention was



Figure 4.4: Answers to the questions "Did you look at the table?", "Did the display on the table bother you?" and "Did the display on the table distract you?" across conditions.

made of the theoretical benefit of a balanced discussion either in terms of levels of participation or subject focus.

4.3.4 Data collection

During their discussion, the subjects were filmed and their voices were recorded using the built-in microphones of the table. Logs of participation levels and of the time spent discussing each suspect were generated and saved. At the end of each experiment, the subjects were asked to fill in a post-experiment questionnaire, which contained 19 questions mostly about the experience they had during the experiment and included four open questions. The questionnaire, included in Appendix A, also asked the users to estimate the amount of time each group member spoke as well as the amount of time they spent discussing each suspect.

4.4 **Results**

Two groups were excluded from the analysis of logs because of human and system errors that led to the loss of recordings and logs for those groups, but not the questionnaires, which were included in analysis related purely to questionnaire answers.

4.4.1 Visibility and unobtrusiveness

We address here the issue of whether or not the table is indeed *visible* and *unobtrusive* as proposed in the first two hypotheses.

The post-experiment questionnaire included some questions meant to get a sense of how subjects perceived the table. Some of the questions and their answers will shed some light on this issue. When asked "Did you look at the table?" the vast majority (88%)of the subjects in both conditions (96% in the speaker condition) said they looked at the table either "sometimes" or "often" as seen in Figure 4.4.

In terms of the obtrusiveness of the table, 86% of participants said they were not bothered by the table and 60% said they were not distracted by it. These answers vary across conditions as



Figure 4.5: Boxplot showing difference between balance in participation across the two conditions for subjects who claimed to believe participation balance is important.

shown in Figure 4.4. Note that in the speaker condition, which is the condition of primary interest to the study, only 25% reported being distracted by the display.

Fifteen percent reported feeling "uncomfortable with seeing their participation levels displayed for all to see." Finally, when asked if they would like to use such a table for other meetings, 66% answered "yes" in the speaker condition whereas only 25% answered "yes" in the topic condition.

We can thus conclude that the table design seemed to satisfy the *visibility* criterion, in that its visualization was looked at most of the time. The subjects also seemed comfortable with the table showing their levels of participation, enough to want to use it in the future. Few reported being bothered by it, but a quarter of the users were distracted. These results indicate the table is also *unobtrusive* to a large extent, but there is nonetheless room for improvement.

4.4.2 General effect on balancing participation

For measuring the effect of the table on balancing participation levels, we compared how balanced groups were in the speaker condition versus the topic condition. We measured balance for each subject as the difference between perfectly balanced participation (i.e. taking up 25% of the total speaking time of the group) and that subject's participation level.

We started by comparing means of individual user balance across conditions using an independent samples t-test. We found no significant effect between how balanced users were in the speaker condition and the topic condition ($m_s = 7.29, m_t = 8.1, t[62] = -0.59, p > 0.1$).

We then took a closer look at the result and noted the following. In the post-experiment questionnaire, the subjects were asked the question: "Do you think it is important for members of the group to participate in a more-or-less balanced manner?" We looked again at the effect of the table on the group members' ability to balance their participation, excluding participants in both conditions that answered "no" to this question (36% of the participants in the study). As we mentioned earlier, *Reflect* is not designed as a tool for enforcing group balance, but rather for supporting it by improving participant awareness. The intention to participate in a balanced manner must thus come from the users themselves, and when this intention is absent, any balancing behavior the user exhibits would likely be coincidental.



Figure 4.6: Change in participation levels of extreme participators in both conditions. Using the speaker visualization, overparticipators reduce their level of participations and underparticipators increase theirs. In the topic visualization both extreme participators move in the direction of further imbalance.

With the remaining participants (46 subjects), i.e. those who claimed that balance in participation is important, we compared the means of their participation levels across two conditions and obtained a statistically significant difference ($m_s = 5.0, m_t = 8.5, t[38] = 2.18, p < 0.05$). In other words, participants who had their participation levels shown to them during the task were statistically more balanced than those who had information about topic focus displayed. This result can be seen in Figure 4.5.

4.4.3 Effect on over and underparticipators

We studied the effect of the different visualizations on a specific subgroup of participants, namely the extreme participators: those who overparticipated and those who underparticipated. We were interested in seeing how, over time, these extreme participators modify their behavior. The objective here is to see if spending time around the table would eventually lead to change in behavior. For that, we divided the 60 minute logs into two equal parts of 30 minutes each. We computed the relative participation of each participant during each of the 30 minutes. We then determined those participants who were extreme participators during the first half-hour, and examined how their participation level changes during the second half-hour.

In line with the method used by DiMicco et al. to determine extreme participators [DiMicco 07b], we defined overparticipators as those who spoke more than the mean participation level (25%) plus the standard deviation of participation levels among all participants. A similar definition was used for underparticipators. We ended up with ten overparticipators and ten underparticipators, divided equally across the conditions.

We noted that, on average, during the first half hour overparticipators in the speaker condition spoke less than overparticipators in the topic condition, though the effect was not significant. More interestingly, in the second half hour, overparticipators in the speaker condition spoke less than they did during the first half hour while in the topic condition, they spoke even more. When comparing the second half-hour participation levels of overparticipators across conditions, we



Figure 4.7: Error levels while estimating speaker levels and time spent on different suspects across conditions.

found a significant correlation between participation levels and the condition ($m_s = 37.1, m_t = 47.6, t[8] = -3.97, p < 0.01$). The effect is similar when looking at underparticipators. During the first half hour, underparticipators spoke more in the speaker condition than they did in the topic condition, and in the second half hour, they increased their participation in the speaker condition and reduced it even more in the topic condition. However, when comparing the second half hour participation levels across conditions, the difference is not significant ($m_s = 11.5, m_t = 6.1, t[8] = 1.304, p > 0.1$). These results, illustrated in Figure 4.6, are similar to the findings of DiMicco et al. [DiMicco 07b].

Though some of these results do not show a statistically significant effect, which is possibly related to the small number of extreme participators, they do show a trend indicating that the table has the desired effect on participation levels.

4.4.4 Effect on individual awareness

We measured the effect the table has on the subjects' ability to estimate both speaker levels for all participants as well as time spent on each topic of discussion (i.e. the suspects). We wanted to evaluate how much users are aware of the information displayed on the surface of the table. The subjects where thus asked, as part of the post-experiment questionnaire, to estimate for each member of the group, including themselves, the relative level of participation (as a percentage of total participation). Note that the visualization on the table was switched off just *before* the participants were informed that the task is over, and the questionnaire was handed out about a minute afterwards. We computed the estimation error of each participant as the sum of differences between their estimate of how much each subject spoke and the actual percentage of time that subject spoke.

For all estimations made, the participants were significantly better at estimating the information in the condition where that information was displayed to them. In other words, when estimating speaker levels, the average error made by the users was significantly lower in the speaker condition than in the topic condition ($m_s = 4.0, m_t = 5.8, t[62] = -3.3, p < 0.01$), and when estimating time



Figure 4.8: Average shift from perfect balance in topic discussion across conditions.

spent on each suspect, the average error was significantly lower in the topic condition than in the speaker condition ($m_s = 5.8, m_t = 4.3, t = 2.4, p < 0.05$). These results are summarized in Figure 4.7.

4.4.5 Effect on topic balance

In addition to the effect of the table on group balance in terms of participation levels, we also investigated the effect on balance in topic discussion for the topic condition, even if this was not the intended purpose of the table. There are of course some conceptual differences between topic balance and participation balance. Unlike participation levels where each member of the group is primarily responsible for his or her own level of participation, no single member is responsible for how much time is spent on each topic. In addition, changes in topic occur much less frequently than changes in speaker, especially near the beginning of the discussion. When the group begins discussing one suspect, they tend to stick to that suspect for a long time before moving to a next one. Finally, the nature of the task does not necessitate that suspects are discussed equally. Some details of the murder mystery require more in-depth discussion than others.

That said, we report that no significant difference was found terms of topic balance across the conditions ($m_s = 5.7, m_t = 6.1, t[49] = -0.24, p > 0.1$). The time spent on individual suspects in the experiments varied greatly among groups. Not surprisingly, a large number of participants (70%) felt that it is not "important to spend more-or-less the same amount of time discussing the case of each suspect." In the case of participation levels, we were able to put aside subjects who felt that speaker balance is unimportant. However, we cannot do so here since, as stated before, topic balance is not determined by individual users, but by the group as a whole.

4.4.6 Qualitative findings

In order to better understand the effect of the table on our subjects, we present here a brief summary of some qualitative analyses done with one of the groups that took part in our experiment. We discuss here a case study of a group who solved the murder mystery task in the speaker condition.

We chose this example because it illustrates both a clear regulatory effect the table had on some members, as well as a clear lack of effect it had on others.

For our analysis, we considered the subjects' answers to two of the open questions in the post-experiment questionnaire:

- 1. Can you indicate one or more occasions where the visual display influenced your behavior?
- 2. Can you indicate one or more occasions where the visual display had a negative impact on the collaboration?



Figure 4.9: *Rate of participation* of members of one group is the amount of speech produced by each member over a certain amount of time. Four points of interests are labeled. The state of the table on these points of interest can be seen in Fig. 4.10.

Some interesting observations can be made about this group discussion.

1. Participant C responded to the second question by saying that when she noticed that her LEDs weren't lit, she got "frustrated." We can clearly see in Fig. 4.9 that the *rate of participation* for this student began much lower than that of participants B and D, but eventually, and for the remainder of the discussion, Participant C began speaking almost as much as participants B and D.

Although frustration is not a desirable emotion we wish our table to invoke in its users, the end result of self-regulation is beneficial.

2. A clearer example of deliberate self-regulation was observed in Participant D who explicitly noted in her answer to the open questions that she "tried not to surpass the speaking time of [Participant B]" and that sometimes she "refrained from talking to avoid having a lot more lights than the others." This is also visible in the graph where we see that Participant D started off participating slightly more than the others. At one point, she reduced her participation level and eventually maintained it at the same rate as Participant B.



Figure 4.10: The state of the table at the four points indicated in Fig. 4.9. 10 minutes into the discussion in (a) the participation is clearly unbalanced. Participant D begins reducing her level of participation. In (b), Participants B and D begin to approach each other while Participant C still lags behind with less than half the total speaking time. In (c), the point in Fig. 4.9 where we see Participant D begin to increase her participation again, the table shows Participants B and D with equal participation. Participant C is still increasing her rate of participation at this point. Near the end of the experiment, in (d), Participants B and D have almost equal participation levels, while C remains slightly behind. Participant A never shows concern for his low participation level.

3. In contrast, we clearly see the total lack of balancing effect the table had on Participant A who kept his participation at an absolute minimum. This participant said, in response to questions in the questionnaire, that he *rarely* looked at the table and that he did not feel it is important for members of the group to participate equally.

Note that the three other participants reported that they looked at the table either *sometimes* or *often*, and all three felt that it was important for members of the group to participate equally.

This case study, while far from sufficient, provides insight into the potential regulatory effect this table can have on group discussion. It emphasizes the informative and not normative role the table has in this kind of setting, i.e. if a user is not interested in participating in a balanced manner, the table will have little or no effect on their behavior.

4.5 Discussion and Limitations

The results of the experiment allow us to draw some conclusions about the effect of a device such as *Reflect* on group behavior. We summarize the main findings here.

4.5.1 Validation of hypotheses

The design of the table appears to have achieved the visibility of the shared display (the first hypothesis) given that the vast majority of participants reported having looked at the table at least sometimes during the experiment. The display was nonetheless distracting (the second hypothesis) to about a quarter of the participants, which indicates that there may be a loss of productivity related to the presence of the display. However, as we did not include a measure performance in our evaluation of the groups, we cannot determine whether or not distraction is sufficiently severe so as to detract from the benefits of the behavioral changes the table induces.

Our third hypothesis is validated: users are more aware of their participation levels when using the table in the speaker condition. The significant difference we found when comparing errors in estimating participation levels indicates that the use of the table increased user awareness of these levels. This, of course, does not imply that the users directly used the display of table to learn these levels. It is also possible that by simply knowing that this information was displayed, the users became more conscious of how much they and others were participating. On the other hand, with 88% of the users reporting that they looked at least sometimes on the table (96% in the speaker condition), it seems safe to make the claim that the information displayed on the table did indeed increase awareness on participation levels among the members of the group.

The fourth hypothesis is only partially validated: users who were shown their participation levels are more balanced than those who are not. Though this turned out to be true in general, it is only statistically significant when considering users who claimed to believe it is important to participate in a balanced manner. Given the informative, rather than normative, nature of the table, this is not surprising. The table does not raise a red flag when a participant speaks too much or too little, thus prompting them to balance their behavior. If a user speaks too much and believes it is acceptable to speak too much for whatever reason, being made aware of their overparticipation will not push them to reduce their levels of speech.

Our results also showed a significant difference among the second half-hour balance between overparticipators across conditions. Underparticipators also increased their participation in the speaker condition and decreased it further in the topic condition, though the difference was not statistically significant. In both cases however, the trend is clear: extreme participators are pushed in the right direction by having the participation levels displayed. However, given the small number of extreme participators, this result is only partially conclusive, and further investigation is needed to establish whether the effect is truly present or not.

4.5.2 Limitations of the study

As a first study, this experiment tried to understand the effect *Reflect* has on small groups. Due to the laboratory nature of this study, the subjects used the table for short periods of time, and once only. They were working with people they did not know beforehand and would likely never meet afterwards. This limits our ability to generalize the results to possible real-world uses of the table. For example, if a group of four people who work together on a daily basis, have regular meetings on such a table, what will the effect be? Will they eventually lose interest in the feedback provided by the table and start ignoring it? Or will they learn to build a sense of trust with the table as an objective observer and rely on it for guidance? These questions cannot be answered by our one-hour experiments.

4.6 Summary

This chapter presented a the first study conducted on the *Reflect* table. The table measured and displayed speaker participation levels, and we evaluated the effect this had on subjects. Four hypotheses were tested and following is a summary of the results:

- H1. The table is visible: *Validated*. The vast majority of participants report looking at the table sometimes or often.
- H2. The table is unobtrusive: *Partially validated*. Most participants found the table not to be distracting.
- H3. The table increases awareness: *Validated*. Awareness about the information displayed was significantly higher in the respective conditions.
- H4. The table balances conversations: *Partially validated*. Only participants who judge participation balance to be of value balanced their participation. This effect was stronger for overparticipators than for underparticipators.

Chapter 5

Study 2: Vocal Engagement

"I was taken by His voice and His gestures, not by the substance of His speech."

GIBRAN KHALIL GIBRAN, from Jesus, the Son of Man

In the previous chapter we evaluated the ability of the *Reflect* table to influence *how much* members of a group speak during a meeting by displaying participation levels. We explore here whether the table is able to influence the behavior of its users on another dimension: *how* they speak. In particular, we look at how prosody influences our perception of the speaker, how an automated voice analysis module was developed and implemented into *Reflect*, and the resulting influence this had on user behavior.

5.1 **Prosody in Voice**

Observers have, for a very long time, recognized the importance of nonverbal components of speech when it comes to how speech is understood and how the speaker is perceived. Psychologists have determined that these nonverbal components, which include hand gestures, eye contact and vocal prosody, contribute significantly to the meaning of what is being said, to the point where they can contradict the verbal content of the speech act [Scherer 80], sometimes inadvertently [Ekman 69].

In addition to altering the meaning of the verbal part of speech, nonverbal cues contain information about the speaker such as their perceived emotions [Scherer 77] and some personality traits [Scherer 78]. In fact, it has been shown that this nonverbal information is sometimes more important than the actual verbal content of speech in predicting the outcome of verbal communication [Pentland 08]. For example, researchers have analyzed the prosodic features of the speech of U.S. presidential candidates during televised debates over eight elections and found that the successful candidate could be predicted solely on nonverbal vocal features [Stanford Jr. 02]. It has also been shown that when the goal of a certain speech is persuasion, pure information is not as effective as information conveyed with the appropriate emotion in the appropriate degree [Scherer 94].

5.2 Prosody and Engagement

In the context of face-to-face collaboration, we look at prosody as a means to measure the perceived engagement of a group member in the task, i.e., how much each member of the group speaks and how much their voice is perceived by the other team members as being actively involved in the collaborative task. We judged this attribute of human voice to be of significance to the collaboration process, as it is important for the motivation of members of a group to feel that those they are working with are engaged in the task at hand. In addition, by working with voice analysis experts, we found this high-level attribute of voice to be computationally tractable. The result of our collaboration is a prosodic model of perceived engagement that was integrated into the *Reflect* table.

Our prosodic model of perceived engagement is based on the model that describes emotion as a two-dimensional construct with the dimensions being *valence* and *arousal* [Russell 80]. Valence refers to whether an emotion is that of pleasure or displeasure. Low-valence emotions include anger and depression; high valence emotions include happiness and comfort. Arousal describes how awake a person is. Thus excitement and distress are high arousal emotions whereas depression and contentment have low-arousal. Figure 5.1 shows how some emotions where positioned in the arousal-valence space.



Figure 5.1: A distribution of certain emotions in the arousal-valence space [Russell 80].

Our model of engagement is based on identifying group members who are both actively participating by speaking, and who show a high level of *arousal* in their voice. In other words, engagement is defined as a function of participation and arousal. This function is described more concretely in Section 5.3.3.

5.3 Voice Analysis in *Reflect*

Our model of vocal arousal is based on a model developed by Branka-Zei Pollerman of Vox Institute¹ and is supported by several studies in emotion recognition [Cowie 01, Scherer 03]. It involves four prosodic features of voice:

- **Pitch**: is defined as the perceived fundamental frequency of a sound. It is measured in Hertz (hz). Several methods exist to automatically compute pitch including auto-correlation algorithms and Fast-Fourrier Transforms.
- **Pitch Variance**: describes how pitch varies over time. It is measured in Hertz per Second (hz/s) and can be computed directly from a temporal representation of pitch.

¹http://www.vox-institute.ch

- **Intensity**: is the perceived loudness of sound and is measured in decibels (dB). It is computed as the amplitude of the wave describing the sound.
- **Rhythm**: is the rate of speech measured in syllables per second (syl/sec). A method for automatically estimating rhythm with sufficient accuracy was recently developed and is based on counting peaks in intensity that correspond to voiced speech [de Jong 09].

5.3.1 Measuring and merging features

In order to measure these features automatically, which was essential for their use in a realtime system, we used a software package for phonetics called Praat [Boersma 01]. Praat allowed scripting of audio manipulation processes in order to extract prosodic features from sound files. A Praat script was thus written that integrated existing scripts measuring different features and eliminated the need for manual intervention in the computation process. The result was a single script that analyzes an audio file and returns the values of its four different features of interest.

We were however interested in generating a single measure of engagement and display that to the user, rather than four individual features that might overload the user's attention. In addition, the average user is not expected to know what values of these features corresponds to high or low arousal. More importantly, this judgment of whether a certain value is too high or two low varies across gender, especially in the case of pitch. We thus developed a method to convert feature values into a more meaningful representation by comparing them to a set of predetermined reference values. For each feature, two sets, one for male and one for female speakers, of reference values representing the mean value an engaged speaker is expected to manifest according to voice experts, as well as the standard deviation across a sample population of French speakers in order to properly evaluate how far a given value is from the reference mean. For that we use a variation on the standard score, *z*, which measures the number of standard deviations a specific value is from the mean and is generally computed as:

$$z = \frac{value_{abs} - mean}{std} \tag{5.1}$$

where *value*_{*abs*} is the actual value of the feature as computed by the Praat script, and *mean* and *std* are the reference mean and standard deviation respectively.

For better readability of the scores, making them more meaningful to non-technical users, we converted each feature into a numeric value that equals 100 when that feature matches its reference mean, and increases/decreases by some α number of points for every standard deviation of variation from the mean. This is given by the following formula:

$$value_{norm} = 100 + \alpha.z \tag{5.2}$$

where $value_{norm}$ is the target standardized value for an individual feature, and where α was chosen to be 20. Finally, all features are averaged together in order to produce a single standardized measure of vocal arousal. A weighted average could have been used to give more weight to one feature over the others, for instance giving less weight to pitch and pitch variance given that they relate to the same physical property of voice. We refrained from doing so as we did not have a clear hypothesis that would have informed us on the value of these weights, and we thus maintained a uniform average.

The resulting procedure takes a single audio file containing a voice sample as input along with the gender of its speaker and produces a single standardized value representing the level of engagement or vocal arousal manifested in the sample as shown in Figure 5.2.

5.3.2 Segmenting the conversation stream

The procedure described in the previous section requires segmented data in the form of short samples of audio, however the *Reflect* table provides a constant stream of audio that needs to be



Arousal as cumulative of four features

Figure 5.2: This screenshot of a voice analysis tool developed to quickly give feedback on a speaker's voice shows a standardized value of arousal that is based on standardized values for the four features. The central horizontal line corresponds to a value equal to the reference, with each grey bar corresponding to one standard deviation difference from the reference mean.

processed almost as fast as it is produced in order to ensure proper responsiveness of the display. This meant that the table needed to collect enough audio data for the computation of features to be possible, while maintaining an appropriate level of responsiveness. If the samples were too short (less than one second) then short bursts of energy or quick changes in pitch would have generated vocal arousal patterns that vary greatly from one second to another. In addition, the accuracy of speech rate increased with the length of the sample and very short samples would have resulted in unreliable number of syllables per second. On the other hand, very long samples (more than a few seconds) would lead to a very slow response from the table that might render the display irrelevant if the speakers perceive it as too slow. The resulting trade-off between samples that were long enough to process but short enough to process promptly led us to experiment with different alternatives.

We first observed that five-second voice samples were sufficiently long to ensure reliable extraction of features, however this turned out to be slightly long for the users. This was particularly a problem when a user made an short interjection in the conversation that lasted less than five seconds, resulting in that utterance being ignored by the table as it did not have enough data to proceed. These short utterances of less than five seconds were quite frequent, and required a change in sample duration.



Figure 5.3: To increase reliability of feature extraction, the system uses five-second windowed samples; however, three-second windows are allowed at the beginning and the end of an utterance to increase responsiveness.

Rather than simply reduce the duration of the samples, bringing us closer to poor feature extraction, we opted for a method that uses a dynamic sample length illustrated in Figure 5.3. This proceeded as follows:

- 1. Collect audio samples that correspond to a single user in a buffer.
- 2. When the duration of audio in a buffer reaches three seconds, process that buffer and return value of vocal arousal.
- 3. Continue to collect audio samples, updating the vocal arousal every second, until the buffer contains a maximum of five seconds of user speech.
- 4. While the user continues to speak add the new samples to the buffer while dropping the oldest samples maintaining a five-second buffer, updating the vocal arousal every second.
- 5. When the user stops talking, continue dropping old samples until the buffer reaches the minimum three-second sample, after which the buffer is emptied and updating the vocal arousal stops until new data is received.

The result is a system that converts the audio stream of the table into a stream of prosodic features representing three to five second samples, produced once every second while a single user is speaking. The final step in this process is visualizing the resulting stream of vocal arousal on the surface of the table.

5.3.3 Visualizing engagement

Recall from Section 3.5 that several visualizations were implemented for the table, of which two were general purpose representations of some specific value. These, referred to as the *territorial* and the *column* visualizations, displayed values in the form of territories around each user and a bar chart respectively. In order to use these visualizations, we needed to determine what specific value to display and how to extract that value from the stream of vocal arousal.

Displaying current engagement

The simplest approach was to display each arousal value as it is received. Thus the table would become very reactive, showing participants a realtime representation of their vocal arousal as they speak. This representation, while interesting for some applications as will be discussed in Chapter 6, was not suited for our own purposes. In Section 3.2.5, we defined as one of the design invariants of the *Reflect* table the need to reduce the cognitive load required by the users in order to extract information from the table visualization. Displaying this realtime information about current levels of vocal arousal meant that the users need to constantly monitor the display to avoid missing some of the information displayed. Persistence of information, as described in Chapter 2, was needed both to reduce cognitive load and to ensure proper grounding of the shared information among participants in the conversation [Dillenbourg 06].

Displaying long-term engagement

The alternative was to maintain a longer-term representation of each speaker's vocal arousal. This global arousal would represent a speaker's overall level of arousal perceived in their voice. Since engagement is defined to include a representation of speaker's participation in addition to their vocal arousal, we introduce the notion of decay to achieve this more realistic representation of engagement.



Figure 5.4: The most recent sample will have a weight of 1, with each preceding sample having 10% less weight in the average that produces the global arousal.

• *Global Arousal.* We compute global arousal as a weighted average of the entire history of the speakers' vocal arousal values with most weight given to the most recent values. Figure 5.4 shows the weights used in this average. This is accomplished by computing a new global arousal, *global_n*, when a new arousal value, *ar_n* is received according to the following formula:

$$global_n = \beta * ar_n + (1 - \beta) * global_{n-1}$$
(5.3)

with β representing the rate of decrease of the averaging weight, in our case 0.1.

• *Decay.* The computation of the global arousal value for each participant only takes into account new arousal values, which in turn are only produced when a given user is speaking. As such, a user with a given level of global arousal who no longer participates in the conversation maintains that level of arousal which is counter intuitive to our notion of perceived engagement. A user that does not participate for long periods of time cannot be perceived as being engaged in a conversation. We therefore introduced the notion of decay, that is the slow decrease in global arousal value for a silent participant. The rate of decay was determined by trial and error with the aim of finding a rate that was not noticeable for someone monitoring the value of global arousal, but which became significant when the participant is silent for considerable period of time. This rate would thus be relative to the context: how long the meeting is, how often participants are expected to speak, etc... The rate we chose for our evaluation of the table was a 1% decrease in global arousal for every five seconds of silence on the part of the user.

The combination of *global arousal* and *decay* constituted our model for perceived engagement, and this is the value that was incorporated into the table during our evaluation.

5.3.4 Updated system architecture

A new module was thus added to the *Reflect* architecture to implement arousal detection. This model consisted in a processing thread that listens to the audio stream coupled with the stream of current speaker IDs determined by the beam-forming algorithm. When appropriate, it computes

5.4. EXPERIMENT



Figure 5.5: The new architecture of the *Reflect* table: a Prosody Analyzer module listens directly to the microphone input and updates the user model with arousal values based on the detected speaker's gender.

an arousal value based on the gender of the relevant user and updates that user model with the new value. The updated system architecture is seen in Figure 5.5

5.4 Experiment

In order to evaluate the effect of this new version of the *Reflect* table, we conducted a second experiment with the aim of observing changes in user behavior the table would have. For this new experiment, we wished to validate the following hypothesis:

H5: When shown their level of engagement, subjects will show a higher level of arousal during the task.

Note that, unlike the previous experiment, we did not predict that the table would lead to "more balanced" engagement, but rather to increased engagement. This is due to two reasons:

- 1. Having observed the previous experiment, we did not note situations where engagement as measured in vocal arousal reached extremely high levels and needed to be reduced. In fact, the vocal arousal of participants was generally less than desired. We thus expected that subjects would tend to increase their arousal when it is shown.
- 2. The notion of balanced engagement is also not tractable. Unlike participation balance which can be objectively evaluated as equal sharing of speaking time, balance in arousal is subjective,

and it would be difficult for us as experimenters to determine a priori what the subjects should perceive as an appropriate level of arousal.

Experimental procedure

We designed this second experiment so that its results are comparable to those of the first study, i.e., we applied the same experimental procedure, in a setting as similar as possible to the first study, and using the same task. We thus recruited 36 additional subjects, selected randomly from a pool of volunteers and paid 50 Swiss Francs (around 45 US dollars). These were divided into 9 groups of 4 and underwent the same experimental procedure as the first 72 subjects, this time using the vocal arousal visualization described above. The same post-experiment questionnaire was used with two additional questions related to arousal. This effectively added a third experimental condition, which we refer to as the *arousal condition*, to the two conditions of the first experiment (speaker and topic), the results of which will be compared in the next section.

The visualization used in this new condition is similar to that of the speaker condition of the first experiment seen in Figure 4.3, with columns representing the global arousal of the subjects. Unlike the speaker and topic conditions of the first experiment, the columns in this new condition were half lit at the start of the experiment representing a moderate level of engagement and went either up or down depending on how the subject behaved during the discussion.

In addition to the similarity of the experimental procedure, all three conditions were recorded in the same way using the three built-in microphones of the table. This made it possible to apply any audio analysis to the recordings and obtain results that would allow us to compare the three conditions.

5.5 Results

We applied the vocal arousal measurement and stream segmentation method described in Section 5.3 to the recordings of all three conditions. We logged, for each subject, the arousal values associated with their participation. We computed the average arousal level for each user and compared these levels across conditions.

5.5.1 General observations on participation balance and engagement

Looking at some questionnaire answers we observed certain differences on how subjects viewed the issues of participation balance and engagement. We explore these differences here as they will become relevant in the discussion about the quantitative results of this study.

Implicit norms

We referred in Chapter 3 to the informative rather than normative nature of the *Reflect* table, and in the previous chapter we noted that the in order for the table to have an effect, a certain norm such as the importance of participation balance needs to be present among the users of the table. In the arousal condition, we asked participants if they believed it is important for members of the group to show a high level of engagement; 85% of subjects answered positively. Compared to the 64% of subjects in the previous experiment who answered positively to the importance of balanced participation in terms of quantity of speech, this high percentage indicates a more salient implicit norm regarding engagement. However, it could also be that engagement is valued more than balance because it is less clearly defined, leading each participant to develop their own interpretation of what engagement is. For instance, our definition of engagement may not fit perfectly with one or more subjects' understanding of what engagement is.



Figure 5.6: The average arousal level of subjects across conditions. The speaker and arousal condition had marginally significant higher arousal than the topic condition. No difference was observed between arousal values of speaker and arousal conditions.

Competition vs. collaboration

In the open questions of the post-experiment questionnaire, subjects were asked if and how the table influenced their behavior and about negative effects it may have had. In the arousal condition, 9 subjects (25%) refer to the visualization creating competition, either by explicitly stating so ("*It creates a situation of competition.*") or by referring to their column as their "score" or "number of points" ("Gives the desire to increase one's 'score"'). No subject in the two other conditions made any such reference to competition in their questionnaire answers.

This was not an anticipated effect, but it is not surprising. Indeed, while reducing or increasing one's own participation may involve an individual's effort, ensuring participation balance is, in a global sense, a group effort. One group member may be able to increase another's participation level by asking them a question and may reduce it by interrupting them or by asking others for input. It is thus difficult to attribute one group member's balanced participation to their own efforts alone, and we thus see participation balance as a collaborative effort. This is even more the case in the topic condition where no single group member is responsible for the topic of discussion. On the other hand, in the arousal condition each member is solely responsible for the degree of arousal they show, and, although one can imagine a case where a person asks another to speak louder or softer, it does not in reality occur very often. In addition, given the implicit norm we discovered in the previous section, it seems common for people to perceive higher arousal as better than lower. Therefore, subjects in the arousal condition found themselves in a situation where the table was displaying in a shared space their level of engagement, of which they are solely responsible and which they judge to be positively correlated with a desired behavior.



Figure 5.7: The average arousal level of subjects across conditions, split by gender. Female subjects showed significantly higher arousal in the speaker condition, whereas males showed significantly higher arousal in the arousal condition.

5.5.2 General difference in arousal across conditions

While our hypothesis predicted a higher arousal level among subjects when information about arousal was displayed on the surface of the table, this did not turn out to be the case. Indeed the average arousal of subjects in the arousal condition was not significantly higher than the topic condition, ($m_a = 100, m_t = 96, t[62] = 2.78, p = 0.11$). However, there was a significant difference when arousal was compared in the speaker condition with the topic condition ($m_s = 102, m_t = 96, t[60] = 2.24, p < 0.05$).

This led us to believe that the speaker condition, originally designed to improve participation balance, was also causing an increase in vocal arousal when compared to our control condition, i.e. when information about topic discussion is displayed, despite the fact that our arousal condition, designed to increase vocal arousal, did not. We wished to understand the reasons behind this unexpected behavior. Thus, we went on a deeper exploration of the data that led us to an even more unexpected result.

5.5.3 Gender differences in behavioral changes

We decided to see if the different visualizations used on the *Reflect* table have varying effect on subjects based on gender. The result was the subjects of different genders had different arousal levels in the different conditions. Seen in Figure 5.7, when compared to the topic condition, males showed marginally significant higher vocal arousal in the arousal condition ($m_a = 102, m_t = 95, t[27] = 2.018, p = 0.052$), whereas females showed marginally significant higher arousal in the speaker condition ($m_s = 107, m_t = 97, t[23] = 1.720, p = 0.063$). In other words, female subjects showed higher vocal arousal when shown their participation levels, whereas male subjects showed higher arousal when the table displayed their level of engagement.

Given that the experiment was not designed with gender difference in mind, it is difficult to explain this result with a great deal of confidence. For instance, it could very well be that the



Figure 5.8: A correlation was found between the amount of speech and the arousal exhibited by individual subjects. This correlation was stronger in the case of female subjects than their male counterparts.

presence of a hidden independent variable which is highly correlated with gender and that has not been controlled in this experiment, such as field of study of the participating students, is responsible for this result. However, this was not possible to verify with the data we had available to us. Therefore, we proceeded with an assumption that the observed difference is indeed due to gender, and we tried to find a reasonable set of hypotheses that explains the data and that is consistent with the literature on gender differences.

The competitive male

Studies have demonstrated the competitive nature of males compared to females, both in terms of how they seek competitive situations [Niederle 07] and in terms of their performance gains during these tasks [Gneezy 03]. In short, males were found to seek competitive tasks more often than women, and when given competitive incentive a performance gain is achieved by men that is not present for women. Thus if we accept the competitive nature of the arousal condition as argued for in section 5.5.1, we would not be surprised that the table had a stronger effect on male subjects than it did on female subjects. We cannot be very confident in an explanation derived from post-hoc analysis of our data when the experiment itself was not designed to explore the notion of competition. However, this hypothesis seems to be consistent with both the results and the literature, and as such it remains a likely hypothesis in need of further study to validate.

Correlation between arousal and amount of speech

We first hypothesized that there may be a correlation between a speaker's amount of speech and their arousal. This correlation was found to exist as is shown in Figure 5.8, and was found to be stronger for female subjects (r[40] = 0.591, p < 0.001) than for males (r[56] = 0.387, p < 0.005).

This correlation may be explained by two factors. The first is that a person who has a lot to say is more likely to be engaged in the task and hence show higher arousal. The second is that interruptions are often won or lost based on the contending speakers' arousal, i.e. a speaker showing higher arousal would be more likely to successfully interrupt another, and less likely to be interrupted, leading to a higher amount of speech. In addition, research has shown that in mixed gender groups, which constitute the vast majority of our groups, males are more likely to interrupt females than the other way around [Zimmermann 75, Smith-Lovin 89]. This is consistent, given that the correlation is much stronger for female subjects, with the explanation that the observed correlation is due to higher arousal speakers being less likely to be interrupted.

This correlation and the yet unexplained significant increase of female arousal in the speaker condition led us to examine the amount of speech for female subjects in the speaker condition. We found that females in the speaker condition spoke significantly more than their male counterparts, and more than females in other conditions, as seen in Figure 5.9.



Figure 5.9: Females in the speaker condition spoke more than females in other conditions and males in the same condition.

In addition to the possible presence of a hidden variable, we explored the possibility that differences in amounts of speech were due to group composition, such that subjects spoke more or less depending on how many other members of their group were of the same gender. However, no significant difference was found in amounts of speech for subjects based on group composition, either for males (F[3, 54] = 0.708, p = 0.55) or females (F[3, 38] = 0.929, p = 0.44). Another explanation could have been related to how female subjects react to the visualization, in that it pushes them to increase their participation. This does not seem to be the case as out of the 11 female subjects in the speaker condition, 5 reported to have deliberately decreased their participation. Our conclusion is that the increase in female participation in the speaker condition is not due to the table visualization, but rather to some uncontrolled variable or perhaps due to several of the 11 female subjects in the sample having simply been more talkative, or more committed to a task for which they were paid. This would also explain, given the correlation between amount of speech and arousal, the high arousal perceived among female subjects of the speaker condition.

5.5.4 Temporal evolution of arousal

After exploring the arousal as a single per-subject measure that we compare across conditions, we now look at arousal as a per-group measure and note how that changes over time. We thus
considered all arousal values within a group as representing the engagement of that group, rather than seperate these values for individuals within it. Each group had at least 960 and at most 1700 timestamped arousal values (on average 1340). We segmented the stream of arousal values into consecutive chunks such that there were 100 chunks which were then averaged to produce 100 arousal values each representing around 45 seconds of group activity. Finally, a moving average filter is applied in order to reduce the effect of noise.



Figure 5.10: The black curve represents the arousal of the group averaged out over all participants across all three conditions, the light colored curves represent arousal within conditions.

General trend

Figure 5.10 shows the curves resulting from vertically averaging all arousal values per group for all participants, and within each condition. The first observation is the common trend that participants in all conditions tend to increase in engagement as the task progresses. A general linear model analysis of the data showed the presence of an effect of time on arousal (F[1] = 22.4, p < 0.001) and no interaction effect between time and condition (F[1] = 2.5, p > 0.1) indicating that arousal varies consistently over time regardless of condition. We attribute the increase in arousal partly to the nature of the task that tends to be more engaging as more details about the murder are revealed, as is usually the case with mystery storylines that get more exciting as they progress to a climactic revelation of the murder's identity. Subjects may also get more and more accustomed to each other and find it easier to engage with their group as time goes by. The lack of difference in the slope of the trend line across conditions may indicate that the effect the table may have on subject arousal, as described in Section 5.5.2, does not vary over time, but rather globally pushes subjects to increase their arousal.

Our second observation was that there are fluctuations in engagement that vary across conditions; however, the last ten minutes of the task experience a high level of engagement across all conditions. It is interesting to also note a common drop in engagement just before the final rise. This pattern in arousal, which is more visible in the aggregated curve where the other within-condition fluctuations disappear, is probably linked to the fact that subjects are informed at the 50^{th} minute that they have 10 minutes remaining. This may have added some pressure especially for groups



that had not yet reached consensus.

Figure 5.11: Evolution of arousal over time for individual groups in the arousal condition.

Individual group trend

Looking at individual groups in Figure 5.11, we note that engagement fluctuates in a rather consistent manner with alternating periods of high and low arousal, which last between 10 and 20 minutes each. Figure 5.12 shows this fluctuation for a single group with certain sections highlighted that we describe here to illustrate what these rises and falls in arousal correspond to. The leftmost section corresponds to an argument between two members about the possibility that someone commits murder without premeditation. This section is transcribed in Figure 5.13. The second segment highlighted in Figure 5.12 and transcribed in Figure 5.14 corresponds to an episode in which one subject reads a section of text to his confused partners. The final segment involves loud laughter and joking among members of the group.

This observation on fluctuation of arousal falls within the realm of conversation analysis, rather than the domain of this dissertation, which is the use of technology to augment conversations. It is, however, in our opinion a very interesting observation and we will return to it in Chapter 7 where we discuss implications for future research and design.

Limitations of the study

The study we conducted showed us the effect on members of a group of a display showing individual members' levels of engagement. The results obtained indicated a difference in levels of arousal between male and female subjects of the study, even in conditions where levels of engagement were not displayed. These differences were not anticipated and thus were not planned for in the design of the experiment. We attempted to explain these findings by exploring the data more deeply; however, a new study needs to be conducted and planned with gender difference in mind, with relevant questions in the post-experiment questionnaire. The conclusions made in terms of gender differences must thus be treated with the appropriate caution until such a time as they are validated with a more targeted study.

5.5.5 Summary

The new version of *Reflect* measured vocal arousal of the speaker's emotion and displayed it on the surface as a perceived level of engagement. Results of our study showed significant differences in



Figure 5.12: Three sections of one group's arousal pattern display different levels of arousal for the group.

```
A: Yes, but it's like now we're arguing because.
B: We don't agree on the suspect.
A: We don't agree.
C: Yeah but...
A: And then I take my phone and I throw it in your face.
B: It's possible.
C: But you're not gonna hit me on the head.
A: Why not?
C: I don't know, we're human afterall.
A: Also...
C: And then, This is not how at an adult age we handle things.
A: Wait, I hope you're joking when you speak like that. Look at what's happening, all the crimes everyday all the... wars everywhere.
```

Figure 5.13: Excerpt of transcipt of segment (A) highlighted in figure 5.12.

behavior among certain groups of subjects, namely males in the prosody condition and females in the speaker condition. It also showed us certain trends in the evolution of arousal among members of the group over time. Regardless of the condition, subjects were generally more engaged as time went by, and their engagement generally reached its maximum near the end. Patterns of alternating periods of low and high arousal were observed among individual groups. A: Billy said, so he arrived at eight o'clock, [reading] It was laying in front of the garage door [pause] the side door where I get the mower out. I remember moving it to the side so I could get the mower out. B: Why? A: To get the lawn mower out. C: I don't have that. What did he say? A: He said... B: Take out... A: He said... B: Take out... A: The lawn mower B: Yeah but what did he move? A: The crowbar. [pause] B: Who? A: Billy.

Figure 5.14: Excerpt of transcipt of segment (B) highlighted in figure 5.12.

Chapter 6

Reflect Outside the Laboratory

"The trouble with her is that she lacks the power of conversation but not the power of speech."

GEORGE BERNARD SHAW

In the two previous chapters we described two laboratory experiments that evaluated different aspects of the *Reflect* table by comparing user behavior when confronted with different visualizations during a collaborative task. These experiments are not enough to gain an understanding of the real potential of the table both in terms of regulating group behavior and in the scope of its use. Thus, the table needed to leave its research setting and undergo real-world testing. For that purpose, certain changes needed to be made to its functionalities that would allow its usage outside the strictly controlled laboratory experiments. This chapter describes the experience of using the *Reflect* table outside the laboratory as well as the modifications it underwent in preparation and as a result of that transition.

6.1 Exploring the Scope of Use

Thus far, the use of the table, in its conception, its implementation and its evaluation, was limited to a specific application: augmenting casual collaboration among small groups, particularly in a learning context. However, the table and its visualizations were not designed such that they were particularly specialized for this context of use, and were thus generic enough to be ported to other domains.

In collaboration with Helyos Partners, a Geneva-based consultant group¹, we began exploring domains that could benefit from such a technology. The table met interest among professionals in some specific domains: namely personal communications skills training and human resource management. Communications training professionals found that their courses could be enriched by using such a table to give direct feedback to participants in many activities involved in these courses, such as role-playing activities. Human resource managers showed interest in a table that can help give real-time feedback to all participants of job interviews as well as performance evaluation meetings. For instance, one issue that arises often in performance evaluation is that the subordinate involved does not get ample opportunity to express themselves, and the *Reflect* table could help bring a balance to this type of situation.

¹http://www.helyospartners.com/

6.1.1 Formal communication training

We targeted at first the domain of communications training, and a prototype of the table was sent to Washington D.C. and presented at the 2009 International Conference and Exposition of the American Society for Training and Development². Our experience in that exposition informed us that professionals in this domain are interested in the technology and its potential but uncertain about investing in a costly and rather bulky piece of hardware whose value has not yet been proven. One of the issues encountered was that most of the trainers needed to be mobile because they often offered their courses at their clients' premises and not their own. A heavy table such as *Reflect* was simply not something they can carry with them.



Figure 6.1: The portable version of *Reflect* (top left) has arrays of LEDs arranged radially and uses the same microphone configuration as the table version (top right). Eight LED arrays, each containing eight multi-color LEDs (bottom right), are detachable (bottom left) for added portability.

We thus began exploring two tracks for making the table more valuable to this professional field. The first track consisted of building a cheaper more lightweight version of the *Reflect* table that would solve the trainers' mobility problem. We built a prototype of a portable circular array of Light-Emitting Diodes (LED) with the three microphones required for direction detection. This prototype, seen in Figure 6.1, would function as a USB peripheral that the trainer would connect to their laptop running special software. The prototype, developed by Wolfgang Hokenmaier, is now on hold for lack of sufficient resources to complete its development.

The second track involved collaborating with a local institution, the Centre d'Education Permanente pour la Fonction Publique (CEP), with the aim of enhancing the functionality of the table

²http://www.astd.org

to better suit the needs of communications training professionals. The CEP is a center that offers personal skills training courses which include communication skills. The center agreed to install a prototype of the table at their premises in exchange for our ability to collect and analyze data about its use, including questionnaires for participants in their courses and interviews with their trainers. Our objective was to get insight on how the table would be used in such a context and use that insight to modify it as needed.

6.1.2 Other uses

Among those interested in the *Reflect* table was a bank in Geneva, Switzerland that has asked for a table to be installed in their Human Resources department. However, the timing of this request made it such that feedback on the use of the table in this context would not be given in time for inclusion in this dissertation. The table will thus be delivered to the bank in hopes that a study on its impact will be conducted after this work is complete.

In addition, a prototype of the table has been installed in one of the publicly available small meeting rooms of the Rolex Learning Center (RLC) of the Swiss Federal Institute of Technology in Lausanne. This table is made available to use by the entire student and faculty body of the institute as well as the University of Lausanne. As the table has been installed during the summer break, it has not had a lot of users for us to begin studying its use. This will change in the coming period as the summer break ends and the students of both universities return to the campus.

Unlike the bank and RLC, the CEP has had ample opportunity to use a *Reflect* table that has been in their offices in Lausanne, Switzerland for over a year. This chapter focuses primarily on how the table was used in this center, and how it was modified to meet the needs of its users.

6.2 *Reflect* Redesigned

The use of *Reflect* outside the controlled lab environment required some changes to its initial design. However, we tried to stay true to the original design philosophy for the table while adapting to the needs of the new situation.

The decision to make the table available to a communication training center meant that the table needed to have some level of configurability in order to provide the trainers with more freedom to choose how the table is used in their courses. At that time, the only way to interact with the table, apart from speaking around it, was via remote administration protocols that required the table to be connected to a network. This led to two main concerns:

- There was a clear privacy concern with having a meeting table with microphones that can be accessed remotely, particularly when its actual users (for e.g. participants in training courses) do not necessarily have reason to trust the table and its developers.
- Controlling the table via remote access network protocols is not a trivial task for the average user, and as the table was meant to be used without our supervision, a more direct method of manipulating the table was needed.

These two constraints led to the development of a new interaction model for the table whereby a user can directly interact with the display, rather than the table being completely autonomous as was the case in its original design. More precisely, we wished to add some basic functionalities to the table that allow for such interaction without network access. We first implemented two fundamental operations for manipulating the table:

• **Clear** the memory of the table when the visualization needs to be reset, such as at the beginning of each new use.

• **Change** the visualization displayed on the table, in other words, navigate within a list of visualizations such as those described in Section 3.5.

A version of the table with these implemented changes was delivered to the CEP and was presented to their training staff, who were given several months to familiarize themselves with the table and its capabilities. Its use during this time was limited and consisted of demos and pilot tests. Following this phase, we invited interested trainers for a meeting where we described our objectives for the table and asked them for feedback on what they plan to do with it. During this meeting, the trainers discussed some ideas about possible features that can be added to the table.

One idea discussed was the possibility of creating detailed on-demand visualizations based on the logged data of a particular meeting. These would be generated after the table is used during debriefing sessions, where the trainer would explain in more detail aspects of the session that are relevant to the lessons of that course. This could be coupled with a video recording of the session so that the trainer can give a more detailed analysis. We found this change to be very interesting and we had already begun work on such a system, but it was not possible to complete implementation in the short time we had remaining for the study.

Other ideas discussed were equally interesting and much easier to implement. Among these two were promptly implemented, and a new version of the table software was installed in time for the trainers to use the table in their courses. The implemented changes were:

- A **pause** operation that would allow a the trainer to speak to participants in a course about what the table is displaying without altering the visualization with his or her voice.
- A **blank visualization** was also added in order to allow for exercises where participants engage in a role-playing activity without realtime feedback on the table, but still have the table monitor the conversation and display relevant information *after* the activity is over.

It is important to note a conceptual difference between the operations introduced to the table by ourselves and those requested by the CEP trainers, in terms of the way users interact with the table. The *clear* and *change* mode operations did not fundamentally alter the interaction paradigm of the table as these were generally operations that would be done at the beginning and end of a meeting, maintaining the initial semi-ambient property of the table, i.e. being in the background but remaining visible, *during* the interaction. However, the *pause* and the *blank* visualization were functions that took away the semi-ambient nature of the table by either (a) making it more central to the activity involving its use, as would be the case when a trainer pauses the table thus interrupting the conversation to bring the participants' attention to it, and (b) pushing it further into the background of the activity, as is the case with the blank visualization, where it becomes a tool for providing post-conversation feedback rather than realtime in-conversation feedback.

Several alternatives were discussed in order to determine how these additional functionalities would be implemented on a system whose only modes of input were three microphones and a power button. As we wished to keep the changes to the table to a minimum, the result was the development of what we referred to as *AudioButtons*.

6.2.1 Implementation of AudioButtons

The *AudioButton* functionality of the *Reflect* table was inspired by our need to interact directly with the table coupled with our desire to avoid adding additional hardware components such as buttons or other peripherals to the table. We thus decided to use the only external component of the table, namely the microphones, as an input device for issuing commands to the table. However, we wished to avoid using speech commands that were likely to trigger false positives if speakers around the table make an utterance that the table detects as a command. Instead we opted for using the microphones directly as buttons by exploiting the fact that the microphones, being so close to each other, detect sounds almost identically regardless of the nature or location of the sound source. This property allows us to easily detect when one of the microphones is tapped,



Figure 6.2: The three audio streams generated by the microphones when (left) one microphone is tapped and (right) the surface of the table is tapped.

as this generates a loud burst of audio energy on a single microphone that is not replicated on the other two. As shown in Figure 6.2, when a burst of audio energy is detected it is fairly easy to determine, by comparing the three microphone signals, whether it is due to one microphone being tapped or whether a loud noise was generated on or around the table. A processing thread thus continually monitors the three audio streams generated by the microphones and notifies the system in the event it detects a loud burst of energy in one audio stream that does not correspond to any such activity on the other two streams. This functionality was implemented by Quentin Bonnard, and the newly augmented *Reflect* microphones became known as *AudioButtons*.



Figure 6.3: The three audio buttons are used to pause, reset, and navigate through the different visualizations on the table.

6.2.2 Using AudioButtons to interact with Reflect

The *Reflect* table had three microphones and thus had three *AudioButtons*. The four functionalities described in Section 6.2 were implemented using the three *AudioButtons* as seen in Figure 6.3 by assinging one button to the Reset and Pause functions and the two other buttons to the navigating forward and backward within a list of available visualizations, including a blank visualization.

The single *AudioButton* causes the table to pause if tapped and resets the table if tapped several times consecutively. The latter behavior was meant to ensure a reset is not executed if the button is tapped accidentally, or if the *AudioButton* listener generates a false positive. When the table is paused, its display blinks slowly indicating that it is no longer listening to the conversation. The visualizations made available in the table were:

- 1. A territorial visualization showing amount of speech.
- 2. A column visualization showing amount of speech.
- 3. The conversation trail visualization.
- 4. A territorial visualization showing realtime arousal.
- 5. A blank visualization.

Note that we restricted the use of the *AudioButtons* to its simplest form: interacting by tapping one button, giving us three possible commands. Given the way the *AudioButton* was implemented, one can easily imagine increasing the number of possible commands for example by tapping two buttons simultaneously, or by tapping several buttons consecutively. However, we found the single *AudioButton* commands to be difficult enough to fully grasp for some people as it was not always easy to remember which *AudioButton* served which function.

6.2.3 Updated architecture

The resulting system architecture seen in Figure 6.4 consisted of augmenting the previous architecture with a new module called AudioButton Listener that monitors the raw audio data produced by the microphones and performs the following functions:

- Pauses the system by ordering the microphone controller to suspend the audio stream.
- Changes the visualization by notifying the renderer.
- Resets the User model by restarting the application.

6.3 *Reflect* at CEP

The CEP, or Centre d'Education Permanente de la Fonction Publique, is a public institution in the Vaud canton of Switzerland that brings together professionals from different domains that provide training courses on a wide variety of topics such as stress management, information technology, and political administration. Among the courses offered, there is a significant number that deal with communication skills across different dimensions: bilateral vs. group communication, collaborative vs. competitive communication, peer-peer vs. supervisor-subordinate meetings, verbal content vs. vocal form. Examples of these courses include: arguing with sensitivity against another speaker's protests, how to lead two-way interviews, and cooperation and communication in small groups.

A *Reflect* table was installed in one of the teaching rooms of the CEP. An introductory session was held where interested trainers were invited, and a brief explanation of the table and its functionalities was provided. It was during this introductory session that the trainers requested the features described in Section 6.2. The table was then left in the CEP for several months, during which trainers were allowed to use it in any courses they felt appropriate. During these courses, we asked the trainers to hand out short questionnaires seen in Appendix B to their course participants. At the end of this testing period, we interviewed some of the trainers in order to learn about how they used the table, what effect it had on their courses and how it can be improved.

6.3. REFLECT AT CEP



Figure 6.4: The AudioButton Listener module interacts with the remainder of the system by listening to the microphone input and issuing the necessary commands to the renderer, the microphone controller, and the user model.

6.3.1 Participant feedback

We had no way to enforce or control the distribution and completion of questionnaires at the end of each use of the table; as a result, participants in some courses did not complete questionnaires. The questionnaires we did receive included answers of 34 participants in four sessions. One session involved a course on how to modify one's voice to maximize one's persuasive influence; it used the table visualization that displayed arousal. The three other sessions, run by the same trainer, used the territorial visualization that shows participation levels in a course on how to manage one's subordinates. During the latter three sessions, not all participants used the table directly; some sat around and observed other participants involved in a role-playing activity on the *Reflect* table. In all, 21 participants in the course use the table directly while 13 simply observed. We report here the questionnaire answers and specify when the answers of the observing participants were not included.

The participants were almost equally divided by gender (16 female, 18 male) and had an average age of 42, with two participants not providing their age.

Figure 6.5 indicates that, like the subjects in our laboratory studies, most participants in the training courses did look at the table at least sometimes. Among the four people reporting rarely looking at the table, two were observers.

Three questions related to the participant's direct involvement with the table and so we only report questionnaire answers from those who actually used the table. These questions and their answers are shown in Figure 6.6. Question 2 (Did you feel that the table was correctly reflecting your participation?) reinforced our belief that the speaker detection algorithm and prosody analysis,



Figure 6.5: Answers to Question 1.

which had thus far not been formally tested, were sufficiently accurate, as 20 of 21 participants found the display to correctly reflect their behavior, including the 8 participants that used the table in the vocal arousal mode. As indicated by Question 3 (Did you feel that the table helped you maintain control over your participation?), more than half the participants did not find the table to be particularly helpful in terms of maintaining control over their participation. The participants instead, as indicated by their answers to open questions, found the table to be more informative than useful as a tool to achieve some particular end. One of those who did feel the table had an effect reported being encouraged to participate, while three others said the table caused them to reduce participation, though the majority did not specify the actual influence the table had. In some instances, this may have been due to the lack of clear guidance on what constitutes "correct" behavior which is often not applicable in a particular activity. We will go into more detail on this aspect of the relationship between the table display and the trainer's instructions for the activity in section 6.3.2. Finally, the answers to Question 5 (Were you comfortable using the table during your training?) indicate very few participants to feel uncomfortable with information about their participation to be displayed on the surface of the table.

The three remaining items, showing in Figure 6.7 were general questions about the table, so all participants were included including those who simply observed. A large number of participants felt that the table added value to their training, and an equal number would have liked to use the table in future training courses. On the other hand, more than half did not wish to have such a table in a meeting at their regular jobs.

The answers to open questions were quite diverse. Some took the opportunity to compliment the design of the table indicating that they liked the lights, the glass, and the simplicity of the display. Others reported liking the table because it was objective and allowed for discovery, but some found it to be merely indicative and not particularly suited for some professional contexts. One participant was worried about the capacity of the table for surveillance.



Figure 6.6: Answers to questions 2,3 and 5 that relate to the direct use of the table.



Figure 6.7: Answers to questions 4,6 and 7 that were general questions concerning the table.

6.3.2 Trainer feedback

A semi-structured interview that lasted between thirty minutes and one hour was conducted with three of the trainers who used the table in their courses, each offering a different type of course.

- 1. The first trainer held several sessions of a course on the use of nonverbal communication, in which one module focused on vocal prosody and its effect on persuasion. One of the sessions is included in the questionnaire feedback described above.
- The second trainer held four sessions for a course on managing subordinates which included training for conducting one-on-one interviews. Three of these sessions were included in the questionnaire feedback.
- 3. The third trainer used the table once in a course on group problem solving strategies. No questionnaire was given out during this session.

Feedback from Trainer 1

The first trainer gave a course that focused on nonverbal communication, in which she introduced different aspects of voice and their role in our perception of the speaker. At the end of the course, the participants engaged in a role-playing activity in which one participant was assigned the task of attempting to persuade another through an appeal to emotion. The arousal visualization was used so that participants are able to monitor in real time how much emotional arousal was perceived in their voice as they attempt to persuade their partners.

The trainer had several remarks on the use of the table in her course. There was a lot of enthusiasm among the participants on the prospect of using a table that can analyze their voices. This was most visible when the table responded quickly by displaying the appropriate amount of arousal, whereas participants seemed frustrated when there was no reaction from the table as was the case when an utterance was too short. Participants found it interesting to compare each others' territories in order to determine who was the "best" speaker. One participant remarked how she finally realized why her friends tell her how inexpressive she was and how she always sounds sad, after seeing her small territory on the table.

On the design of the table, the trainer remarked that accessing the different visualizations with the *AudioButtons* was not easy. She likened finding the right visualization on the table with searching for an audio segment on an old cassette recorder. She also found the display of the table to be lacking in luminosity, making it less visible than she would have preferred, especially for people who were observing the interaction from afar.

She finally remarked that the value of the table lies in the objectivity of its feedback despite the fact that it deals with a domain that has often been constricted to the realm of subjective judgement, namely the perception of voice.

Feedback from Trainer 2

The second trainer used the table during an role-playing activity as part of a course on managing one's subordinates. The activity involved two-person meetings where one took the role of the supervisor while the other played the subordinate. The territorial visualization for amount of speech was used. The participants conducted three different types of meetings: one for the subordinate to talk about their employment, one for the supervisor to evaluate the subordinate, and one for the two to fix the subordinate's objectives for the next year. The trainer gave instructions to the participants on how speaking time would ideally be distributed between the two roles in each of the scenarios. The trainer noted that the participants were not always capable of maintaining the proper level of participation and did not use the table to try to regulate. Instead, the trainer would interrupt at times when the participation balance was very skewed. He would pause the table, show the participants their levels and attempt to discover with them why they are speaking so much or so little. He expressed the importance of the role of the trainer when using the table, which is to intervene when necessary and not expect the table do the training by itself.

On the other hand, the trainer found that the actual role the table accomplishes does not sufficiently justify its use especially when this use requires significant overhead. For example, in

one session he had to ask all of his participants to move to another room to conduct the role-playing activity just to use the table, and he found it hard to justify this with the value the table provided. The amount of speech of each participant, he said, is only one of potentially hundreds of possible measures of a participants' role in communication, and having special hardware specifically designed for that was not easy to justify. He thus found it necessary to decrease participants' expectations on the table by specifying that it is part of a research project and that it is still under experimentation.

He also mentioned that the main value of the table is not with the type of information it displays but with the objectivity in which it is displayed, making this table far more useful in situations where participants might deny their excessive or insufficient participation. In his case, participants were virtually always satisfied with the trainer's assessment of their participation, and did not need further validation from an objective source.

Feedback from Trainer 3

The third trainer gave a course on problem solving in groups. He held 20 minute group problem solving activities where he instructed his participants to conduct their meeting following a fixed five-step procedure: narrowing down the problem, brainstorming the possible causes, brainstorming possible solutions, selecting the best solution, and planning the appropriate action. The trainer used the territorial visualization for amounts of speech but did not give any specific instructions on how the table needs to be used during the task or how the territories need to be distributed to ensure effectiveness of the group.

The trainer observed that while the table helps people understand their role in a conversation and pushes them to be more active, it was mostly ignored by his participants. He blames this fact on two issues, the first being a problem with the table itself being far too discreet. The second is that the task he had assigned to his participants was already very engaging and they were unable to focus on dealing with the content of the task (solving the problem) and the structure of the task (the five-step process) and at the same time monitor the table visualization. In addition, he remarked that the participants had little or no guidance in terms of how they were to share speaking time; no implicit or explicit norm was present. This led one participant, at the end of her task to ask if her disproportionately large territory on the table was good or bad.

The trainer finally said that in future uses of the table, he would use more appropriate tasks such as debates or information pooling tasks where participation levels are important to monitor and control.

6.3.3 Lessons learned

The qualitative feedback from both the trainers and participants in training courses using this table provided us with some important insight on the role of the *Reflect* table in this context. We describe here the main lessons learned from this experience.

On the utility of objective feedback

Several trainers and subjects noted the importance of the objectivity in the table feedback, some referring to that as its most important asset. When it comes to human-human communication a truly objective judge is sometimes hard to come by, and even when this judge is present it is easy for one party to publicly or internally consider the judge's assessment to be biased or unfair. Thus, it is of particular interest to explore domains of use where such objective feedback would be highly appreciated. Consider for example emotionally charged debates or tough negotiations between two self-interested parties. In these cases, it is difficult to find a human observer that can be trusted by both parties to reliably determine even something as factual as the amount of speech of the parties involved.

On the simplicity and the obtrusiveness of the table

Both the input interface and the display of the table were deliberately designed to be simple and unobtrusive for reasons described in Sections 3.2.3 and 3.2.6. The participants, who did not wish to be overloaded with information during their tasks, appreciated the simplicity of the table output. Some trainers, however, found the *AudioButtons* were too simple a tool to navigate the table and would have preferred a more complex control scheme. In terms of obtrusiveness of the table, we learned that our semi-ambient display showed a level of discretion that was not suitable for use in training courses. In fact, given that in this context the use of the table is more deliberate and its role is more precise, a brighter more intrusive display would have been more appropriate. Trainers wanted their participants to be actively reminded of the information displayed on the table, and this was not the case with the subtle display may be suitable for one context such as our original domain of casual collaboration but may not appropriate for another use, such as formal communication training. Thus this trade-off needs to be tweaked specifically for each domain of use.

On the importance of guiding norms

Finally we got further insight on the importance of guiding norms when using the a mirroring display. Many participants did not know what it was they were expected to do with the knowledge of their participation levels. Without an underlying norm on what constitutes "good" participation, the table seems like nothing more than a playful gadget. In this experiment, when norms were missing, participants tended to either ignore the table or to seek guidance from the trainer. We believe that the presence of explicit or implicit norms is an important factor in determining usefulness of this table.

6.4 Summary

This chapter introduced a new model for interacting with the *Reflect* table using *AudioButtons*. We also reported on the use of the table outside of the laboratory setting by describing an experience where trainers in communications skills conducted role-playing activities for their course participants around the table. This experience informed us on different aspects of the *Reflect* table that had not been throroughly addressed before.

Chapter 7

General Discussion and Conclusions

"The Road goes ever on and on Out from the door where it began. Now far ahead the Road has gone, Let others follow it who can!"

J. R. R. TOLKIEN, from *The Hobbit*

We summarize here the contributions of this work and discuss its limitations and its implications on the relevant domains.

7.1 Summary

Several studies were conducted on the use of the *Reflect* table under different conditions. Here is a brief list of the main findings:

- 1. User of the table are more likely to balance their levels of participation when this information is displayed on a common surface, especially when they believe participation balance is important.
- 2. Overparticipators are more likely to reduce their participation levels in order to achieve balance than underparticipators to increase their participation.
- 3. Users are better at estimating theirs and their partners' participation levels when this information is displayed on the table.
- 4. Balance in topic discussion was not improved, nor was it deemed important.
- 5. Displaying levels of engagement to users only influenced male subjects causing them to exhibit higher vocal arousal.
- 6. Users' amounts of speech were correlated with their vocal arousal, and this effect was stronger for female users.
- 7. An unexplained increase in amount of speech and vocal arousal for female users was seen when their participation levels were displayed on the surface of the table.
- 8. A pattern of fluctuation in arousal was observed among members of each group.
- 9. Users found information displayed on the surface of the table to be visible and generally not distracting.

In addition to these experimental results, the use of the table as part of training courses for personal communication skills highlighted three main issues:

- 1. Users attached great value to the table as a genuinely objective judge.
- 2. In the absence of normative guidance, the table has limited utility.
- 3. The context in which the table is to be used must have fundamental implications on the trade-off between the visibility and unobtrusiveness of the display.

7.2 General Discussion

Many lessons were learned from the experience of designing, building and evaluating the *Reflect* table both inside and out of the laboratory. We reflect here on some of these lessons.

7.2.1 The effect of *Reflect* on user behavior

Our experiments reinforced previous findings in group mirror research described in Section 2.5; displaying information about participation levels alters user behavior during collaborative meetings. *Reflect* achieves this while maintaining minimal changes to natural user behavior by replacing microphones attached to the user with a beam-forming microphone array and projected displays with a matrix of LEDs under the usable surface of the table.

The different circumstances and different ways in which this occurs also shed some light on how this behavioral change comes about. For example, when displaying participation levels, users who believed participation balance is important were more likely to be balanced than those who did not. This indicated that simply displaying information to users does not necessarily alter their behavior.

When the level of engagement was displayed on the table, the outcome was that male users increased their engagement but female users did not. This led us to examine the difference in terms of regulating participation balance and levels of engagement, where we remarked that displaying engagement may have promoted competition among the participants rather than regulation. Another difference between the two situations was that it was easy to define and label over and under participators, but it is not evident what an over-engaged speaker is. Even if we were to define such notion as over-engagement, it is unlikely that any of the speakers in our study would satisfy that category given the generally lukewarm attitude subjects had in dealing with the task, where the vast majority were at most moderately engaged. This meant that regulating engagement in this task was equivalent to increasing it, and it may be the case that a group mirror display is less capable of increasing engagement in the same way that it was less capable of increasing participation.

Finally, it may also be that deliberately controlling one's engagement is more difficult than controlling their participation levels. One reason for that would be that participation level is something the user can directly observe and control, whereas engagement was a complex metric based on several variables unknown to the users and thus harder for them to regulate. This would lead us to the notion that when designing a group mirror for self-regulation, one must target aspects of participation that the users perceive as important to regulate and that the users are capable of regulating.

7.2.2 Gender differences in human-computer interaction

Our second study showed some differences in the way male and female users behave around the *Reflect* table. The studies themselves were not designed to monitor such differences, so the lessons we learned from this were mostly conjectural rather than grounded in experimental evidence. However, we do believe that these differences do exist, and further study is needed to understand

their nature. Our experience also taught us that such differences may be show up in areas where we do not expect them.

Research on gender differences in human-computer interaction has already shown various ways in which male and female users approach technology differently. We believe that with the advent of Ubiquitous Computing, whereby technology is integrated into the very fabric of human life, these differences need to be considered more seriously in the design and implementation of new technology. The increasing role of computers in our daily personal lives, and not only our professional lives, further reinforces this need.

7.2.3 Group mirrors and norms

Our experience with the *Reflect* table highlighted the importance of implicit and explicit norms in how group mirror displays influence their users. The use of the table in communication training courses showed us that in the absence of guiding norms from the trainer, users of the table may get confused and fail to interpret the infromation displayed on the table. In the context of our laboratory experiments, different norms led to different types of behavior. Both balanced participation and high engagement were perceived as important, however a group mirror for engagement may have led to competition rather than cooperation. Thus our experience has given us some insight on the role of norms in how group mirrors influence users.

We also mentioned earlier that the presence of a group mirror itself may induce or modify the existing norms. This aspect of the relationship between norms and group mirrors has not been adequately studied in this work, and we think it is an important notion that needs to be addressed in future research. In particular, does displaying information about participation levels influence the users' perception of how important balanced participation is? Is it possible to change the way the information is displayed to alter this perception?

7.2.4 Semi-ambient displays

We defined a display as semi-ambient if it strikes a balance between visibility and unobtrusiveness. We found that the *Reflect* table met that criterion in the sense that it was looked at often, its information was perceived and assimilated to the point of causing behavioral change, and yet remained largely not distracting. This seemed to be appropriate for a group mirror used during a collaborative task. On the other hand, this balance did not seem appropriate for other uses of the table, particularly the use in communication training courses, where it was perceived as too discreet. This makes the effectiveness of a semi-ambient display dependant on the context in which it is used.

7.2.5 Prosodic analysis of group meetings

Study 2 revealed an interesting pattern in the general levels of arousal of the group when considered over the duration of the task. This pattern was not explored further as the objective of this work is not to analyze meetings, but rather to analyze the effect of the *Reflect* table on meetings. However, we found this pattern to be very interesting and needs further study to shed light on other areas of research. It may be relevant for example for automatic segmentation of meetings where each observed change in arousal may signify a change in the state of the meeting. It could also be potentially used to classify meetings based on the pattern of arousal fluctuation they exhibit, as well as identify different speaker roles based on each speaker's influence on the group's arousal fluctuation. Finally, it may help as a tool to navigate audio recordings of a meeting by selecting portions of the meeting based on the degree of arousal exhibited.

These open up new research directions that benefit from the results of our experiment as a starting point. We can also envisage the utility of this kind of observation in the domain of this thesis itself by considering a new kind of visualization that would benefit from meeting segmentation and meeting classification to display information to the participants about the kind of meeting they appear to be having, and where they are in that meeting.

7.2.6 Voice analysis beyond communication training

This work has also contributed to the domains of voice analysis and communication training by pushing voice analysis further towards the realm of precise scientific measurement. In addition to improving the quality of courses by providing trainers with automated analysis tools, this also has other applications such as bringing the experience of speech analysis experts beyond the limits of course training. One can imagine software applications to improve public speaking skills or to help prepare for job interviews without the presence of a human expert to judge the quality of one's voice. Such technology could popularize voice analysis and make it more accessible for the general public.

7.3 Limitations

There are several limitations with the system we designed as well as the method we used to address the questions raised. We discuss some of these limitations here.

7.3.1 Generalizability of experimental results

Our two laboratory studies gave us insight on the role of group mirrors in the form of semiambient displays in regulating group behavior. However, both studies were conducted in a highly controlled setting which leads to a loss of generalizability. Some of the variables that were controlled, such as group size, duration of the discussion, and the nature of the task, would of course vary greatly in real world situations. These parameters might strongly affect the way the table influences user behavior. Many questions are thus left unanswered. What would happen when our table is used repeatedly by the same group of people over longer periods of time? Would people even use the table for collaborative work outside the context of a paid experiment? Which of the two visualizations we experimented with would users prefer to use in a given context?

To address these questions, we have already began the next steps in the evaluation of this technology by placing prototypes of the table in the real world. Users will be able to use the table on their own time, for their own reasons and in their own way. We hope that over time this will generate some real world insight on the usability of this technology.

7.3.2 Duration of the CEP study

A *Reflect* table has been at the Centre d'Education Permanente pour la Fonction Publique (CEP) for about 16 months, at the time of writing this dissertation. During the first 9 months the table was continually being developed to incorporate some of the additional features we described in Section 6.2. In the end, the actual time the table was used by trainers was only a few months, and while this was sufficient for some trainers to get a chance to experiment with its use, it was not sufficient to explore its full potential. As some trainers told us during our interviews, they would have liked to have more time to design exercises that maximize the utility of the table, perhaps after a few modifications to the table itself.

Thus our experience with the table at the CEP only allowed us to scratch the surface of the potential of *Reflect* in this domain. We hope that in the coming months the trainers will be able to integrate the table more closely in their courses, which will give us better insight on the usefulness or lack thereof of the table.

7.3.3 *Reflect* across culture

The *Reflect* table was built and evaluated in French-speaking part of Switzerland, with the vast majority of its participants being deliberately chosen from a relatively consistent cultural background. The table deals with conversation, a fundamentally human activity whose underlying rules and structure are deeply rooted in social culture that vary significantly from one area of the world to another. This means that, while the lessons we learned from our work are not completely restricted to the cultural sphere of French-speaking Switzerland, they are far from generalizable across the world. In addition to the different value we associate to participation balance and self-regulation, other more basic differences exist in the way people speak and perceive speech. For example, prosodic patterns of speech vary, and what may sound like a highly engaged speaker in one culture may be perceived very differently in another.

7.4 Final words

This dissertation presented the outcome of a few years of research during which an interactive table was designed, implemented, and evaluated both in and out of a laboratory setting. This work built on notions from Ubiquitous Computing and Computer-Supported Collaborative Work and Learning. We positioned this work in what we refer to as computer-augmented communication, where the computer does not intervene directly in human-human communication, but rather supports it while remaining in the background. This allows to both maintain natural human communication while at the same time provide some computer support in an unobtrusive manner. We described how such a system can indeed influence human behavior in conversational settings, but we also showed some of its limitations. In addition, we explored other domains of use for our device by taking it out of the laboratory and studying its potential in the world of communication training.

Throughout our work we have avoided the use of the terms *smart* and *intelligent* when referring to the *Reflect* table, not out of modesty, but because of our belief that these labels create a certain set of expectations that, by design, our table does not attempt to meet. Indeed, as we described in Chapter 3, the *Reflect* table is capable of many functionalities, but it avoids performing certain functions associated with intelligent behavior such as understanding the context of its use and interpreting the data it is collecting. The table does not attempt to assess the quality of collaboration or of an individual's contributions, rather it provides its users with information that may aid them in making that judgment themselves. After all, whether or not someone is talking too much is not simply a function of the amount of noise that person is producing, nor even the number of words they say. It is a complex question that depends on an intricate set of parameters that form the physical, social, and cultural context within which that person is speaking. What is their role in the group? What is the group discussing? What are the interpersonal relationships within the group? What did this person do during the last week, both with respect to what is being discussed and in general? Analyzing the context of the interaction is infinitely complex, and the *Reflect* table does not claim to take part in it. It simply provides its users with relevant information and leaves it to them to answer the question: "Am I talking too much?"

Complete bibliography

[Alavi 09]	H. Alavi, P. Dillenbourg & F. Kaplan. <i>Distributed Awareness for Class Orchestra-</i> <i>tion</i> . Learning in the Synergy of Multiple Disciplines, pages 211–225, 2009.
[Ang 05]	J. Ang, Y. Liu & E. Shriberg. <i>Automatic dialog act segmentation and classification in multiparty meetings</i> . In Proc. ICASSP, vol. 1, pages 1061–1064. Citeseer, 2005.
[Anguera 07]	X. Anguera, C. Wooters & J. Hernando. <i>Acoustic beamforming for speaker diariza-</i> <i>tion of meetings</i> . IEEE Transactions on Audio, Speech, and Language Processing, vol. 15, no. 7, pages 2011–2022, 2007.
[Bachour 08]	Khaled Bachour, Frederic Kaplan & Pierre Dillenbourg. <i>Reflect : An Interac- tive Table for Regulating Face-to-Face Collaborative Learning</i> . Lecture Notes in Computer Science. Springer, Berlin / Heidelberg, 2008.
[Bachour 10]	Khaled Bachour, Frederic Kaplan & Pierre Dillenbourg. <i>An Interactive Table for Supporting Participation Balance in Face-to-Face Collaborative Learning</i> . IEEE Transactions on Learning Technologies, vol. 3, pages 203–213, 2010.
[Baker 99]	M. J. Baker. <i>Argumentation and constructive interaction</i> . In Pierre Coirier & Jerry Andriessen, editors, Studies in Writing: Foundations of Argumentative Text Processing, vol. 5, pages 179–202. University of Amsterdam Press, 1999.
[Baker 02]	Michael Baker. <i>Forms of cooperation in dyadic problem-solving</i> . Revue d'intelligence artificielle, vol. 16, no. 4-5, pages 587–620, 2002.
[Banse 96]	R. Banse & K.R. Scherer. <i>Acoustic profiles in vocal emotion expression</i> . Journal of personality and social psychology, vol. 70, no. 3, pages 614–636, 1996.
[Basu 02]	S. Basu. Conversational scene analysis. PhD thesis, Citeseer, 2002.
[Bathiche 10]	S. Bathiche & A. Wilson. Microsoft Surface, 2007, 2010.
[Bellotti 93]	V. Bellotti & A. Sellen. <i>Design for privacy in ubiquitous computing environments</i> . In Proceedings of the third conference on European Conference on Computer- Supported Cooperative Work, page 92. Kluwer Academic Publishers, 1993.
[Bergstrom 07a]	Tony Bergstrom & Karrie Karahalios. <i>Conversation Clock: Visualizing audio patterns in co-located groups</i> . In HICSS, page 78, 2007.
[Bergstrom 07b]	Tony Bergstrom & Karrie Karahalios. <i>Conversation votes: enabling anonymous cues</i> . In CHI '07: CHI '07 extended abstracts on Human factors in computing systems, pages 2279–2284, New York, NY, USA, 2007. ACM.
[Bergstrom 07c]	Tony Bergstrom & Karrie Karahalios. <i>Seeing More: Visualizing Audio Cues</i> . In Interact, 2007.

[Blaye 88]	A. Blaye. <i>Confrontation socio-cognitive et resolution de problemes</i> . PhD thesis, Centre de Recherce en Psychologie Cognitive, Universit'e de Provence, France, 1988.
[Boersma 01]	Paul Boersma. <i>Praat, a system for doing phonetics by computer.</i> Glot International, vol. 5, no. 9/10, pages 341–345, 2001.
[Brdiczka 05]	O. Brdiczka, J. Maisonnasse & P. Reignier. <i>Automatic detection of interaction groups</i> . In Proceedings of the 7th international conference on Multimodal interfaces, page 36. ACM, 2005.
[Brown 07]	Barry A. T. Brown, Alex S. Taylor, Shahram Izadi, Abigail Sellen, Joseph Kaye & Rachel Eardley. <i>Locating Family Values: A Field Trial of the Whereabouts Clock</i> . In John Krumm, Gregory D. Abowd, Aruna Seneviratne & Thomas Strang, editors, Ubicomp, vol. 4717 of <i>Lecture Notes in Computer Science</i> , pages 354–371. Springer, 2007.
[Carroll 03]	J.M. Carroll, D.C. Neale, P.L. Isenhour, M.B. Rosson & D.S. McCrickard. <i>Notifica-</i> <i>tion and awareness: synchronizing task-oriented collaborative activity</i> . International Journal of Human-Computer Studies, vol. 58, no. 5, pages 605–632, 2003.
[Carver 01]	C.S. Carver & M.F. Scheier. <i>On the self-regulation of behavior</i> . Cambridge Univ Pr, 2001.
[Ciolfi 05]	L. Ciolfi & L. Bannon. <i>Space, Place and the Design of Technologically-Enhanced Physical Environments</i> . Spaces, Spatiality and Technology, pages 217–232, 2005.
[Clark 91]	H.H. Clark & S.E. Brennan. <i>Grounding in communication</i> . Perspectives on socially shared cognition, vol. 13, pages 127–149, 1991.
[Cohen 84]	Elizabeth G. Cohen. <i>Instructional groups in the classroom: organization and processes</i> , chapter Talking and working together: status interaction and learning, pages 171–188. Orlando: Academic, 1984.
[Cohen 94]	Elizabeth G. Cohen. <i>Restructuring the classroom: conditions for productive small groups</i> . Review of Educational Research, vol. 64, no. 1, pages 1–35, Spring 1994.
[Cowie 01]	R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz & J.G. Taylor. <i>Emotion recognition in human-computer interaction</i> . IEEE Signal processing magazine, vol. 18, no. 1, pages 32–80, 2001.
[Davies 02]	N. Davies & H.W. Gellersen. <i>Beyond prototypes: Challenges in deploying ubiquitous systems</i> . IEEE pervasive computing, vol. 1, no. 1, pages 26–35, 2002.
[Davis 06]	D. Davis & E. Patronis. Sound system engineering. Focal Press, 2006.
[de Jong 09]	N.H. de Jong & T. Wempe. <i>Praat script to detect syllable nuclei and measure speech rate automatically</i> . Behavior research methods, vol. 41, no. 2, page 385, 2009.
[Dembo 87]	M.H. Dembo & T.J. McAuliffe. <i>Effects of perceived ability and grade status on social interaction and influence in cooperative groups</i> . Journal of Educational Psychology, vol. 79, pages 415–423, 1987.
[Dey 00]	A.K. Dey & G.D. Abowd. <i>Towards a better understanding of context and context-awareness</i> . In CHI 2000 workshop on the what, who, where, when, and how of context-awareness, vol. 4, pages 1–6. Citeseer, 2000.

[Dietz 01]	Paul Dietz & Darren Leigh. <i>DiamondTouch: a multi-user touch technology</i> . In UIST '01: Proceedings of the 14th annual ACM symposium on User interface software and technology, pages 219–226, New York, NY, USA, 2001. ACM.
[Dillenbourg 96]	Pierre Dillenbourg, Mike Baker, Agnes Blaye & Claire O'Malley. <i>The evolution of research on collaborative learning</i> . In Learning in Humans and Machine: Towards an interdisciplinary learning science, pages 189–211. Elsevier, Oxford, 1996.
[Dillenbourg 06]	P. Dillenbourg & D. Traum. <i>Sharing solutions: Persistence and grounding in multimodal collaborative problem solving</i> . Journal of the Learning sciences, vol. 15, no. 1, pages 121–151, 2006.
[Dillenbourg 07]	P. Dillenbourg & F. Fischer. <i>Basics of Computer-Supported Collaborative Learning</i> . Zeitschrift für Berufs- und Wirtschaftspädagogik, vol. 21, pages 111–130, 2007.
[Dillenbourg 08]	P. Dillenbourg. <i>Integrating technologies into educational ecosystems</i> . Distance Education, vol. 29, no. 2, pages 127–140, 2008.
[DiMicco 04]	Joan M. DiMicco, Anna Pandolfo & Walter Bender. <i>Influencing group participation with a shared display</i> . In CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work, pages 614–623, New York, NY, USA, 2004. ACM Press.
[DiMicco 05]	J. M. DiMicco. <i>Changing Small Group Interaction through Visual Reflections of Social Behavior</i> . PhD thesis, Massachusetts Institute of Technology, 2005.
[DiMicco 07a]	Joan Morris DiMicco & Walter Bender. <i>Group Reactions to Visual Feedback Tools</i> . In PERSUASIVE, pages 132–143, 2007.
[DiMicco 07b]	Joan Morris DiMicco, Katherine J. Hollenbach, Anna Pandolfo & Walter Bender. <i>The impact of increased awareness while face-to-face</i> . HumComput. Interact., vol. 22, no. 1, pages 47–96, 2007.
[Do-Lenh 09]	S. Do-Lenh, F. Kaplan, A. Sharma & P. Dillenbourg. <i>Multi-finger interactions with papers on augmented tabletops</i> . In Proceedings of the 3rd International Conference on Tangible and Embedded Interaction, pages 267–274. ACM, 2009.
[Dourish 92]	P. Dourish & V. Bellotti. <i>Awareness and coordination in shared workspaces</i> . In Proceedings of the 1992 ACM conference on Computer-supported cooperative work, page 114. ACM, 1992.
[Ekman 69]	P. Ekman & W.V. Friesen. Nonverbal Leakage and Clues to Deception. 1969.
[Faller 10]	Christof Faller, 2010. http://www.illusonic.com.
[Ferrer 03]	L. Ferrer, E. Shriberg & A. Stolcke. <i>A prosodybased approach to end-of-utterance detection that does not require speech recognition</i> . In ICASSP, Hong Kong. Citeseer, 2003.
[Gatica-Perez 05]	D. Gatica-Perez, I. McCowan, D. Zhang & S. Bengio. <i>Detecting group interest-level in meetings</i> . In Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP). Citeseer, 2005.
[Geißler 98]	J. Geißler. <i>Shuffle, throw or take it! working efficiently with an interactive wall</i> . In CHI, vol. 98, pages 18–23. Citeseer, 1998.
[Gellersen 99]	H.W. Gellersen, M. Beigl & H. Krull. <i>The MediaCup: Awareness technology embedded in an everyday object</i> . In Handheld and Ubiquitous Computing, pages 308–310. Springer, 1999.

[Gneezy 03]	U. Gneezy, M. Niederle & A. Rustichini. <i>Performance in Competitive Environments: Gender Differences</i> *. Quarterly Journal of Economics, vol. 118, no. 3, pages 1049–1074, 2003.
[Greenberg 96]	S. Greenberg, C. Gutwin & A. Cockburn. <i>Awareness through fisheye views in relaxed-WYSIWIS groupware</i> . In Graphics interface, pages 28–38. Citeseer, 1996.
[Greitemeyer 03]	T. Greitemeyer & S. Schulz-Hardt. <i>Preference-consistent evaluation of information in the hidden profile paradigm: Beyond group-level explanations for the dominance of shared information in group decisions.</i> Journal of Personality and Social Psychology, vol. 84, no. 2, pages 322–339, 2003.
[Gross 05]	R. Gross, A. Acquisti & H.J. Heinz III. <i>Information revelation and privacy in online social networks</i> . In Proceedings of the 2005 ACM workshop on Privacy in the electronic society, page 80. ACM, 2005.
[Grudin 94]	J. Grudin. <i>Computer-supported cooperative work: History and focus</i> . Computer, vol. 27, no. 5, pages 19–26, 1994.
[Gutwin 95]	C. Gutwin, G. Stark & S. Greenberg. <i>Support for workspace awareness in educational groupware</i> . In The first international conference on Computer support for collaborative learning, pages 147–156. Citeseer, 1995.
[Gutwin 04]	C. Gutwin & S. Greenberg. <i>The importance of awareness for team cognition in distributed collaboration</i> . Team cognition: Understanding the factors that drive process and performance, vol. 201, 2004.
[Haller 10]	M. Haller, J. Leitner, T. Seifried, J.R. Wallace, S.D. Scott, C. Richter, P. Brandl, A. Gokcezade & S. Hunter. <i>The NICE discussion room: integrating paper and digital media to support co-located group meetings</i> . In Proceedings of the 28th international conference on Human factors in computing systems, pages 609–618. ACM, 2010.
[Hollan 92]	J. Hollan & S. Stornetta. <i>Beyond being there</i> . In Proceedings of the SIGCHI conference on Human factors in computing systems, page 125. ACM, 1992.
[Hoyles 85]	Celia Hoyles. <i>What is the point of group discussion in mathematics?</i> Educational studies in mathematics, vol. 16, pages 205–214, 1985.
[Huang 07]	Y. Huang, O. Vinyals, G. Friedland, C. Müller, N. Mirghafori & C. Wooters. <i>A fast-match approach for robust, faster than real-time speaker diarization</i> . In Proceedings of IEEE ASRU, pages 693–698. Citeseer, 2007.
[Huber 90]	George P. Huber. <i>A theory of the effects of advanced information technologies on orga-</i> <i>nizational design, intelligence, and decision making.</i> The academy of management review, vol. 15, no. 1, pages 47–71, Januray 1990.
[Hung 08]	H. Hung, Y. Huang, G. Friedland & D. Gatica-Perez. <i>Estimating the dominant person in multi-party conversations using speaker diarization strategies</i> . In International Conference on Acoustics, Speech and Signal Processing (ICASSP), Las Vegas. Citeseer, 2008.
[James 93]	D. James & J. Drakich. <i>Understanding gender differences in amount of talk: A critical review of research</i> . Gender and conversational interaction, pages 281–312, 1993.

[Jermann 01]	Patrick Jermann, Amy Soller & Martin Mühlenbrock. <i>From mirroring to guiding: A review of the state of art technology for supporting collaborative learning</i> . In European Perspectives on Computer-Supported Collaborative Learning, pages 324–331, 2001.						
[Jermann 04]	P. Jermann. <i>Computer support for interaction regulation in collaborative problemsolv-</i> <i>ing</i> . Unpublished Ph. D. thesis, University of Geneva, Switzerland, 2004.						
[Kim 08]	T. Kim, A. Chang, L. Holland & A.S. Pentland. <i>Meeting mediator: enhancing group collaborationusing sociometric feedback</i> . In Proceedings of the ACM 2008 conference on Computer supported cooperative work, pages 457–466. ACM, 2008.						
[Kirsh 10]	David Kirsh. <i>Thinking with external representations</i> . AI and Society, pages 1–14, 2010. 10.1007/s00146-010-0272-8.						
[Langheinrich 10]	Marc Langheinrich. <i>Ubiquitous Computing Fundamentals</i> , chapter Privacy in Ubiquitous Computing. Chapman and Hall, 2010.						
[Leshed 10]	Gilly Leshed, Dan Cosley, Jeffrey T. Hancock & Geri Gay. <i>Visualizing language use in team conversations: designing through theory, experiments, and iterations</i> . In CHI EA '10: Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems, pages 4567–4582, New York, NY, USA, 2010. ACM.						
[Lindgaard 94]	Gitte Lindgaard. Usability testing and system evaluation: a guide for designing useful computer systems. Taylor and Francis, 1994.						
[M. Webster Jr. 83]	M. Webster Jr. & J. E. Driskel Jr. <i>Beauty as Status</i> . The American Journal of Sociology, vol. 89, no. 1, pages 140–165, July 1983.						
[Mark 96]	Gloria Mark, Jörg M. Haake & Norbert A. Streitz. <i>Hypermedia structures and the division of labor in meeting room collaboration</i> . In CSCW '96: Proceedings of the 1996 ACM conference on Computer supported cooperative work, pages 170–179, New York, NY, USA, 1996. ACM.						
[Mast 96]	M. Mast, R. Kompe, S. Harbeck, A. Kießling, H. Niemann, E. Nöth, E.G. Schukat-Talamazzini & V. Warnke. <i>Dialog act classification with the help of prosody</i> . In Fourth International Conference on Spoken Language Processing, pages 1732–1735. Citeseer, 1996.						
[McCowan 05]	I. McCowan, D. Gatica-Perez, S. Bengio, G. Lathoud, M. Barnard & D. Zhang. <i>Automatic analysis of multimodal group actions in meetings</i> . IEEE Transactions on Pattern Analysis and Machine Intelligence, pages 305–317, 2005.						
[Mota 03]	S. Mota & R.W. Picard. <i>Automated Posture Analysis for Detecting Learner? s Interest Level.</i> 2003.						
[Niederle 07]	M. Niederle & L. Vesterlund. <i>Do Women Shy Away from Competition? Do Men Compete Too Much?</i> *. The Quarterly Journal of Economics, vol. 122, no. 3, pages 1067–1101, 2007.						
[Nikolovska 06]	Lira Nikolovska. <i>Conversation table,</i> 2006. http://web.mit.edu/lira/www/projects/convtable.htm.						

[Otsuka 08]	K. Otsuka, S. Araki, K. Ishizuka, M. Fujimoto, M. Heinrich & J. Yamato. <i>A real-</i> <i>time multimodal system for analyzing group meetings by combining face pose tracking</i> <i>and speaker diarization</i> . In Proceedings of the 10th international conference on Multimodal interfaces, pages 257–264. ACM, 2008.
[Pentland 08]	Alex (Sandy) Pentland. <i>Honest Signals: How They Shape Our World</i> . The MIT Press, 2008.
[Pléty 96]	Robert Pléty. <i>L'Apprentissage Coopérant</i> . Ethologie et psychologie des communi- cations. Presses Universitaires de Lyon, 1996.
[Pollerman 10]	Branka Zei Pollerman, 2010. http://www.vox-institute.ch/.
[Polzin 00]	T.S. Polzin, A. Waibel <i>et al. Emotion-sensitive human-computer interfaces</i> . In ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion, 2000.
[Prante 04]	T. Prante, N.A. Streitz, P. Tandler, I. Fraunhofer & G. Darmstadt. <i>Roomware: Computers disappear and interaction evolves</i> . Computer, vol. 37, no. 12, pages 47–54, 2004.
[Rienks 06]	R. Rienks, D. Zhang, D. Gatica-Perez & W. Post. <i>Detection and application of influence rankings in small group meetings</i> . In Proceedings of the 8th International Conference on Multimodal interfaces, page 264. ACM, 2006.
[Russell 80]	J.A. Russell. <i>A circumplex model of affect</i> . Journal of personality and social psychology, vol. 39, no. 6, pages 1161–1178, 1980.
[Sacks 95]	H. Sacks. Lectures on Conversation. Blackwell Publishing, 1995.
[Salomon 89]	G. Salomon & T. Globerson. <i>When teams do not function the way they ought to</i> . International Journal of Educational Research, vol. 13, pages 89–99+, 1989.
[Scherer 73]	K. R. Scherer, H. London & J. Wolf. <i>The voice of confidence: Paralinguistic cues and audience evaluation</i> . Journal of Research in Personality, vol. 7, pages 31–44, 1973.
[Scherer 77]	K. R. Scherer & J. Oshinsky. <i>Cue utilization in emotion attribution from auditory stimuli</i> . Motivation and Emotion, vol. 1, pages 331–346, 1977.
[Scherer 78]	K. R. Scherer. <i>Personality inference from voice quality: The loud voice of extroversion</i> . European Journal of Social Psychology, vol. 8, pages 467–487, 1978.
[Scherer 80]	K. R. Scherer. <i>The functions of nonverbal signs in conversation</i> , pages 225–244. NJ: Erlbaum, Hillsdale, 1980.
[Scherer 94]	K. R. Scherer. <i>Interpersonal expectations, social influence, and emotion transfer,</i> pages 316–336. Cambridge University Press, Cambridge and New York, p.d. blanck (ed.). edition, 1994.
[Scherer 03]	K.R. Scherer, T. Johnstone & G. Klasmeyer. <i>Vocal expression of emotion</i> . Handbook of affective sciences, pages 433–456, 2003.
[Scott 03]	S.D. Scott. <i>Territory-based interaction techniques for tabletop collaboration</i> . In UIST 2003 Conference Companion, pages 17–20, 2003.
[Scott 04]	Stacey D. Scott, M. Sheelagh, T. Carpendale & Kori M. Inkpen. <i>Territoriality in collaborative tabletop workspaces</i> . In CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work, pages 294–303, New York, NY, USA, 2004. ACM.

[Scott 06]	S.D. Scott & S. Carpendale. <i>Investigating Tabletop Territoriality in Digital Tabletop Workspaces</i> . 2006.
[Slavin 83]	R. E. Slavin. Cooperative Learning. Longman, New York, 1983.
[Smith-Lovin 89]	L. Smith-Lovin & C. Brody. <i>Interruptions in group discussions: The effects of gender and group composition</i> . American Sociological Review, vol. 54, no. 3, pages 424–435, 1989.
[Stanford Jr. 02]	Gregory W. Stanford Jr. & Timothy J. Gallagher. <i>Spectral Analysis of Candidates'</i> <i>Nonverbal Vocal Communication: Predicting U.S. Presidential Election Outcomes.</i> Social Psychology Quarterly, vol. 65, no. 3, pages 298–308, 2002.
[Stasser 92]	Garold Stasser & Dennis Stewart. <i>Discovery of Hidden Profiles by Decision-Making Groups: Solving a Problem Versus Making a Judgment</i> . Journal of Personality and Social Psychology, vol. 63, no. 3, pages 426–434, September 1992.
[Stasser 03]	G. Stasser & W. Titus. <i>Hidden profiles: A brief history</i> . Psychological Inquiry, vol. 14, no. 3, pages 304–313, 2003.
[Stolcke 00]	A. Stolcke, K. Ries, N. Coccaro, E. Shriberg, R. Bates, D. Jurafsky, P. Taylor, R. Martin, C.V. Ess-Dykema & M. Meteer. <i>Dialogue act modeling for automatic tagging and recognition of conversational speech</i> . Computational linguistics, vol. 26, no. 3, pages 339–373, 2000.
[Streitz 98]	N.A. Streitz, J. Gei ler & T. Holmer. <i>Roomware for cooperative buildings: Integrated design of architectural spaces and information spaces</i> . In Cooperative Buildings. Integrating Information, Organization, and Architecture: First International Workshop, CoBuild'98, Darmstadt, Germany, February 1998. Proceedings, page 4. Springer, 1998.
[Streitz 01]	Norbert A. Streitz, Peter Tandler & Christian Muller-tomfelde. <i>Roomware: Towards the Next Generation of Human-Computer Interaction based on an Integrated Design of Real and Virtual Worlds.</i> pages 553–578. Addison Wesley, 2001.
[Streng 09]	S. Streng, K. Stegmann, H. Hußmann & F. Fischer. <i>Metaphor or diagram?: comparing different representations for group mirrors</i> . In Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7, pages 249–256. ACM, 2009.
[Sundström 05]	P. Sundström, A. Ståhl & K. Höök. <i>A user-centered approach to affective interaction</i> . Affective Computing and Intelligent Interaction, pages 931–938, 2005.
[Vinciarelli 09]	A. Vinciarelli, M. Pantic & H. Bourlard. <i>Social signal processing: Survey of an emerging domain</i> . Image and Vision Computing, vol. 27, no. 12, pages 1743–1759, 2009.
[Webb 83]	Noreen M. Webb & Linda K. Cullian. <i>Group Interaction and Achievement in Small Groups: Stability over Time</i> . American Educational Research Journal, vol. 20, no. 3, pages 411–423, 1983.
[Webb 91]	N. M. Webb. <i>Task-related verbal interaction and mathematics learning in small groups</i> . Journal for Research in Mathematics Education, vol. 22, no. 5, pages 36–389+, 1991.
[Weiser 91]	M. Weiser. <i>The computer for the twenty-first century</i> . Scientific American, vol. 265, no. 3, pages 94–104, 1991.

[Weiser 96]	M. Weiser & J.S. Brown. <i>Designing calm technology</i> . PowerGrid Journal, vol. 1, no. 1, pages 75–85, 1996.
[Willem Doise 76]	Anne-Nelly Perret-Clermont Willem Doise Gabriel Mugny. <i>Social interaction and cognitive development: Further evidence</i> . European Journal of Social Psychology, vol. 6, no. 2, pages 245–247, 1976.
[Winquist 98]	J.R. Winquist & J.R. Larson Jr. <i>Information pooling: When it impacts group decision making</i> . Journal of Personality and Social Psychology, vol. 74, no. 2, pages 371–377, 1998.
[Wisneski 98]	C. Wisneski, H. Ishii, A. Dahley, M. Gorbet, S. Brave, B. Ullmer & P. Yarin. <i>Ambient displays: Turning architectural space into an interface between people and digital information</i> . Cooperative buildings: Integrating information, organization, and architecture, pages 22–32, 1998.
[Zancanaro 06]	M. Zancanaro, B. Lepri & F. Pianesi. <i>Automatic detection of group functional roles in face to face interactions</i> . In Proceedings of the 8th international conference on Multimodal interfaces, page 34. ACM, 2006.
[Zimmermann 75]	D.H. Zimmermann & C. West. <i>Sex roles, interruptions and silences in conversation</i> . Newbury House, Rowley, MA, 1975.

Appendix A Post-experiment Questionnaire

The following questionnaire was given to participants in all the three conditions with some minor changes. Questions 3 and 6 where only added for the third (arousal) condition. Questions 12 and 13 were modified in each condition to match what the table actually displayed during that condition. Only the French text was included in the distributed questionnaire.

			Grou	ıp:	
	Post Experiment (Questionnaire	Parti	cipant:	
* Age:	* Sexe:	* Langue Maternelle França French mother tongue?	aise ?	OUI YES	NON <i>NO</i>
* Nombre d'heures pa Number of hours sp	ssées en ligne chaque jour: ent online everyday :	* Si NON, depuis quand par If NOT, since when do you	liez vo <i>i speal</i>	us le fra k french	nçais ? ?

Merci de répondre aux questions suivantes de manière aussi précise que possible. *Please answer the following questions as accurately as possible.*

 Quelle était la difficulté de la tâche? How difficult was the task?

Très facile					Très	difficile
Very Easy					Very	difficult
1	2	3	4	5	6	7

- 2. En terme de **quantité de parole**, à combien jugez-vous la participation de chaque participant ? In terms of **quantity of speech**, how much would you say each participant participated?
 - a. Participant A _____%
 - b. Participant B _____%
 - c. Participant C _____%
 - d. Participant D _____%
- 3. En terme de l'**implication dans la tache**, comment jugez-vous la participation de chaque participant ?

In terms of involvement in that task, how would you say each participant participated?

		Pas d'implication <i>No involvement</i>					Bea A loi	ucoup d' t of Invo	coup d'implication of Involvement	
a.	Participant A		1	2	3	4	5	6	7	
b.	Participant B		1	2	3	4	5	6	7	
c.	Participant C		1	2	3	4	5	6	7	
d.	Participant D		1	2	3	4	5	6	7	

- 4. Combien de **temps** pensez-vous avoir passé pour discuter chaque suspect ? How much **time** would you say you spent discussing each suspect in the case?
 - a. Eddie _____%
 - b. Billy _____%
 - c. Mickey ______%
- 5. Jugez-vous **important** que **chaque participant parle plus ou moins autant que les autres** lors de la discussion ?

Do you think it's **important** for **each member of the group to speak more or less equally** during the discussion?

OUI - YES NON - NO

6. Jugez-vous **important** que **chaque participant montre un haut niveau d'implication** lors de la discussion ?

Do you think it's important for each member of the group to show a high level of involvement during the discussion?

OUI - YES NON - NO

7. Jugez-vous **important** que **plus ou moins le même temps** soit passé à discuter **le cas de chaque suspect** ?

Do you think it is **important** that **more or less equal time** is spent discussing **the case of each** *suspect*?

OUI - YES NON - NO

8. Avez-vous **regardé** la table ? Did you **look** at the table? JAMAIS NEVER

RAREMENT	PARFOIS	SOUVENT
RARELY	SOMETIMES	OFTEN

- Est-ce que l'affichage sur la table vous a dérangé(e)? Did the display on the table bother you? OUI - YES NON - NO
- 10. Est-ce que **l'affichage** sur la table **vous a distrait(e)**? *Did the display on the table distract you*? OUI - YES NON - NO

11. Avez-vous eu l'impression que la table montrait des informations **en rapport** avec ce que vous faisiez ?

Did you feel that the table showed information that was relevant to what you were doing?

OUI - YES NON - NO

12. Vous êtes-vous senti(e) à l'aise de voir des informations sur votre implication affichées à la vue de tous ?

Did you feel comfortable seeing information about your involvement displayed for all to see?

OUI - YES NON - NO

13. Avez-vous l'impression que les autres membres du groupe se sont sentis à l'aise de voir des informations sur leur implication affichées à la vue de tous ? Did you have the impression that the other members of the group felt comfortable seeing information about their behavior displayed for all to see?

OUI - YES NON - NO

14. Aimeriez-vous avoir une telle **information affichée** pendant des **autres réunions** que vous aurez ?

Would you like to have this kind of **information displayed** during **other meetings** you will be having?

OUI - YES NON - NO

15. Si vous recevriez une telle table, l'utiliserez-vous ? If you are given a table like this, would you use it?

NON	NO
OUI, comme table de salon.	YES , as a coffee table.
OUI, comme table de travail.	YES, as a work table.

16. Etes-vous **satisfait(e)** des décisions prises par votre groupe ? Are you **satisfied** by the decision taken by your group?

OUI - YES NON - NO

17. À votre avis, qui a **vraiment** tué Robert Guion ? In your own opinion, who **really** killed Robert Guion?

Mickey Malone	Billy Prentice	Eddie Sullivan	Quelqu'un d'autre	Ne sais pas
			Someone else	I don't know

18. Pouvez-vous indiquer une ou plusieurs occasions où l'affichage visuel a influencé votre manière de vous comporter ?
 Can you indicate one or more occasions in which the visual display influenced the way you

19. Pouvez-vous indiquer une ou plusieurs occasions où **l'affichage visuel** a eu un **impact négatif** sur la collaboration ? *Can you indicate one or more occasions in which* **the visual display** had a **negative impact** on the collaboration?

20. Que pensez-vous de **la table et de l'affichage** ? What do you think of **the table and the display**?

behaved?

21. Avez-vous d'**autres** commentaires ? Do you have **other** comments ?
Appendix B CEP User Questionnaire

The following questionnaire was given to participants of CEP training courses. Only the French text was included in the distributed questionnaire.

Questionnaire sur l'utilisation de la table Reflect

Questions about the use of the Reflect table



Les informations que vous allez fournir dans ce questionnaire seront traitées de manière anonyme et seront utilisées uniquement par des membres de l'équipe de recherche de l'EPFL.

The information you will provide in this questionnaire will be treated anonymously and will be used soley by the members of the EPFL research team.

Sexe/*Gender*: _____ Age: _____ Date : ___

Heure/*Time* : ____ h ____

Indiquez votre place par une croix sur le schéma de la table ci-contre, en vous orientant avec la disposition des microphones.



Indicate your place with a cross on the adjacent diagram, using the microphone arrangement to orient yourself.

Merci d'entourer une réponse par question : Please circle one answer per question :

- 1. Avez-vous regardé la table ?

 Did you look at the table?

 Jamais
 Rarement

 Parfois
 Souvent

 Never
 Rarely
 Sometimes
- Aviez-vous l'impression que la table reflétait correctement votre participation ? Did you have the impression that the table was correctly reflecting your participation? OUI - YES NON - NO
- Aviez-vous l'impression que la présence de la table vous aidait à maintenir le contrôle sur votre participation ?
 Did you have the impression that the presence of the table helped you maintain control over your participation?
 OUI YES
 NON NO
- Aviez-vous l'impression que la présence de la table apportait un plus à votre formation ? Did you have the impression that the presence of the table added value to your training? OUI - YES NON - NO
- Etiez-vous à l'aise par rapport à l'utilisation de la table lors de cette formation ? Were you comfortable using the table during this training? OUI - YES NON - NO

- 6. Aimeriez-vous utiliser une table similaire lors de prochaines formations ?
 Would you like to use a similar table during future training?
 OUI YES
 NON NO
- 7. Aimeriez-vous utiliser une table similaire lors de vos réunions de travail ?
 Would you like to use a similar table during work meetings?
 OUI YES
 NON NO

Merci de répondre aux questions suivantes : *Please answer the following questions:*

- 8. Qu'avez-vous aimé à propos de la table ? Pas aimé ? What did you like about the table? Didn't like?
- 9. Comment est-ce que la table à influencé votre attitude ? How did the table influence your behavior?
- 10. Avez-vous d'autres commentaires ? Do you have other comments?

List of Figures

2.1	Modern day pad, tab, and board 4
2.2	The <i>MediaCup</i>
2.3	Examples of roomware
2.4	Territories on a tabletop
2.5	The Study Room and the Room of Opinion
2.6	Social and action awareness on a chat window
2.7	Radar view for workspace awareness 9
2.8	The "interactions" paradigm
2.9	A group mirror displaying the ratio of communication to action 13
2.10	The architecture of interaction regulation 14
2.11	First version of Second Messenger 16
2.12	Second version of Second Messenger
2.13	The Conversation Clock 17
2.14	<i>Conversation Votes</i>
2.15	The Metaphoric Group Mirror 18
2.16	The Meeting Mediator
2.17	<i>GroupMeter</i>
3.1	The theoretical architecture of the system
3.2	The Virtual Noise-Sensitive Table
3.3	The Physical Noise-Sensitive Table
3.4	LED board used in the table
3.5	The first version of <i>Reflect</i> 28
3.6	Triangular microphone configuration used in <i>Reflect</i>
3.7	The basic system architecture of <i>Reflect</i>
3.8	The second version of <i>Reflect</i>
3.9	Visualizations on <i>Reflect</i>
3.10	Summary of the different properties of the versions of the table
4.1	Participation of group members in the choice shift task
4.2	Excerpt from the murder mystery task material
4.3	Visualizations used in first study 39
4.4	Answers to questions about visibility and obtrusiveness
4.5	Boxplot showing difference in participation balance
4.6	Change in participation for extreme participators
4.7	Error levels in estimating speakers and topic focus
4.8	Average shift from perfect balance in topic discussion
4.9	Rate of participation in the case study group
4.10	The state of the table at specific points for the case study group
5.1	Some emotions in the arousal-valence space

5.2	Voice analysis tool screenshot	52
5.3	Three and five second windowed samples	52
5.4	Weights for samples when computing global arousal	54
5.5	System architecture after adding prosody analysis	55
5.6	The average arousal level of subjects across conditions	57
5.7	Average arousal split by gender	58
5.8	Correlation between amount of speech and arousal	59
5.9	Amount of speech across condition by gender	60
5.10	Evolution of arousal of all groups over time	61
5.11	Evolution of arousal over time for individual groups	62
5.12	Evolution of arousal for one group	63
5.13	Excerpt of transcript from one group in high arousal	63
5.14	Excerpt of transcript from one group in low arousal	64
6.1	The portable version of <i>Reflect</i>	66
6.2	Sounds generated by tapping the <i>AudioButton</i>	69
6.3	Basic functions of the <i>AudioButtons</i>	69
6.4	System architecture after introducing <i>AudioButtons</i>	71
6.5	Answers to Question 1	72
6.6	Answers to questions 2, 3 and 5	73
6.7	Answers to questions 4, 6 and 7	73

Khaled Bachour

PhD in Computer Science

Chemin de Rionza, 13 1020 Renens, Switzerland \bowtie khaled.bachour@gmail.com $\overset{()}{=} +41 \ 79 \ 812 \ 20 \ 90$

	Education
2005–2010	PhD in Computer Science , <i>École Polytechnique Fédérale de Lausanne</i> , Switzerland. Augmenting Face-to-Face Collaboration with a Low-Resolution Semi-Ambient Feedback
2002–2005	Master in Computer Science, American University of Beirut, Lebanon.
1999–2002	BS in Computer Science , <i>American University of Beirut</i> , Lebanon. With High Distinction
1990-1999	Lebanese Baccalaureate , <i>International College, Beirut</i> , Lebanon. Mathematics Track
	Languages
Arabic	Native speaker
English	FluentExcellent in written and spoken language. IELTS Score 8.5.
French	Advanced Very good in written and spoken language.
	Skills
	Programming
Systems	Linux, Windows Software LATEX, MS Office, SPSS, Adobe Pho- toshop, Adobe Premier
Programming	Java, C/C++, Python, MPI, OpenGL, C#, ASP .NET, SQL, MATLAB
	Work Experience
	Research
2006–2010	Research Assistant , <i>École Polytechnique Fédérale de Lausanne</i> , Switzerland. Interactive furniture for collaborative learning.
2002–2003	Research Assistant , <i>American University of Beirut</i> , Lebanon. Incremental Knowledge Acquisition with object-oriented database systems
	Teaching
2008,2009	Information Technology Project , <i>Teaching Assistant</i> , School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne. Course coordination, lab supervision, grading exams and managing student assistants.
2002, 2005	Database Systems and Java Programming courses , <i>Teaching Assistant</i> , Department Computer Science, American University of Beirut. Lab supervision and grading assignments.

Web Development

2004 Web Developer, *Engicon*, Amman, Jordan. Developed ASP .NET CMS for company website.

Film-makingInterests and ActivitiesFilm-makingWritten, directed and participated in several amateur short films.PhotographyAmateur photographer with some local activities.OthersCooking, music, cinema, badminton

Publications

Journal Articles

2010 K. Bachour, F. Kaplan, P.Dillenbourg. "An interactive Table for Supporting Participation Balance in Face-to-Face Collaborative Learning". In IEEE Transactions on Learning Technologies, pp. 203-213, July-September, 2010.

Other Papers

- 2008 K. Bachour, F. Kaplan, P.Dillenbourg. "Reflect: An Interactive Table for Regulating Face-to-Face Collaborative Learning". In Lecture Notes in Computer Science, Springer, Berlin / Heidelberg.
- 2009 F. Kaplan, S. Do-Leng, K. Bachour, G. Y. Kaw, C. Gault, and P. Dillenbourg. "Interpersonal Computers for Higher Education". Interactive Artifacts and Furniture Supporting Collaborative Work and Learning, pages 129-145, Springer US.
- 2010 K. Bachour, H. S. Alavi, F. Kaplan, P. Dillenbourg. "Low-Resolution Ambient Awareness Tools for Educational Support.". CHI 2010 Workshop: Next Generation of HCI and Education. Atlanta, GA.