# A Multimodal Database for Mimicry Analysis

Xiaofan Sun[1], Jeroen Lichtenauer[2], Michel Valstar[2], Anton Nijholt[1], Maja Pantic[1, 2]

[1] Human Media Interaction, University of Twente, Enschede, NL
{x.f.sun, a.nijholt}@ewi.utwente.nl.
[2] Department of Computing, Imperial College, London, UK
{j.lichtenauer, michel.valstar, m.pantic}@imperial.ac.uk

**Abstract.** In this paper we introduce a multi-modal database for the analysis of human interaction, in particular mimicry, and elaborate on the theoretical hypotheses of the relationship between the occurrence of mimicry and human affect. The recorded experiments are designed to explore this relationship. The corpus is recorded with 18 synchronised audio and video sensors, and is annotated for many different phenomena, including dialogue acts, turn-taking, affect, head gestures, hand gestures, body movement and facial expression. Recordings were made of two experiments: a discussion on a political topic, and a role-playing game. 40 participants were recruited, all of whom self-reported their felt experiences. The corpus will be made available to the scientific community.

**Keywords:** Mimicry database, interaction scenario, synchronized multi-sensor recording, annotation, social signal processing, affective computing.

## 1 Introduction

To study the phenomena in social interactions between humans in more detail and to allow machine analysis of these social signals, researchers are in need of rich sets of labelled data of repeatable experiments, which should represent situation occurring in daily life [1], [2]. This data could then be used to develop and benchmark new methods for automatic detection and recognition of such behavioural cues. Having sufficient labelled/unlabelled data of mimicry episodes and detailed expressions is a prerequisite for automatically detecting and analyzing mimicry occurring in social interactions. Mimicry episodes are difficult to collect and detect mainly because they inherently involve the temporal interaction between two or more persons. They are unpredictable and relatively rare, which makes it difficult to elicit mimicry displays without deliberately designing recording scenarios. Good experiment scenarios are based on existing social psychology literature, increasing the chance of recording clear, salient, and high-quality cues that relate to displays of mimicry.

There is no doubt that mimicry occurs in the most basic interaction, which is a dyad. Mimicry can be expressed in both in auditory and visual channels. However, obtaining multi-modal sensor data that can be used for multi-modal analysis is a challenge in itself. The recording of different modalities requires different equipment,

and different equipment necessitates different expertise to develop, set up and operate [3], [4]. In summary, to create a database that will contribute to the research of mimicry, we need interdisciplinary knowledge, including social psychology and engineering, as well as methodological solutions to combine and fuse the sensory data from a diversity of multimodal equipment. This is probably the main reason that we currently lack such a mimicry database

In addition, manual labelling of spontaneous mimicry is time consuming and requires trained annotators. It is also a subjective process, lacking strict guidelines how to perform the annotation. Thus, even if recordings are rich in expressions of spontaneous mimicry, there is no way of attaining a set of consistent and reliable labels. Due to these difficulties, nearly all of the existing databases are artificial and, to different extents, acted [5]. As a result, though mimicry has attracted increasing attention from researchers in different research fields, automatic mimicry analysis is not seriously addressed in current computer science and machine analysis.

Recently created databases containing emotional expressions in different modalities can be used as a reference for creating a mimicry database. These databases mostly consist of audio, video or audiovisual data [6], [7], [8], [9], [10]. Although many of these databases can be considered to contain naturalistic data, none of them were designed to capture episodes of mimicry. Another issue is that, because they were not designed to capture mimicry, there is no well-synchronized view of all partners in a conversation. This makes the automatic analysis and annotation of mimicry in these databases difficult, if not impossible.

One of the notable databases with spontaneous reactions is the Belfast database (BE) created by Cowie et al. [11]. Even though this database consists of spontaneous reactions in TV talk shows and is rich in body gestures and facial expressions, the context was less effective in evoking mimicry.

Some naturalistic or induced-spontaneous datasets of human-human or human-computer interactions might not contain a large number of explicit mimicry episodes. Nevertheless, they could be useful in training tools for the automatic detection of cues that do not directly indicate mimicry but could be relevant to e.g. human affect, which probably is a factor affecting mimicry. For example, the AMI meeting corpus [10] consists of 100 hours of meeting recordings in which people show a huge variety of spontaneous expressions. The data, mostly centred on the idea of enacting meetings, is related to mimicry of dominant and submissive nonverbal behaviours. Tiedens and Fragale [12] have demonstrated that people may react to others who display dominance with dominant displays of their own, and similarly respond to submissive behaviours with mutual submission. Both are referred to as postural mimicry.

The SEMAINE corpus [6] consists of recorded audio-visual conversations with annotation for five affective dimensions (arousal, valence, power, anticipation and intensity). It uses the Sensitive Artificial Listener (SAL) technique, described in [13] as ”a specific type of induction technique that focuses on conversation between a human and an agent that either is or appears to be a machine and it is designed to capture a broad spectrum of emotional states”. In the SEMAINE corpus, each participant has a conversation with four different emotionally coloured virtual agents, in which mimicry-relevant cues, such as emotional mimicry can probably be found.

Our proposed database intends to become a valuable resource for research of mimicry. This research, in turn, will allow conversational agents to improve their

social interaction capabilities in face-to-face communication by recognising mimicry and responding appropriately by instantiating psychological theories through the use of nonverbal cues. From the automatic understanding of mimicry and other social signals, and prediction of how these signals might affect social situations, applications can be derived that can help people improve their social skills.

## 2 Mimicry Perception Conversation Recording

In this paper we describe a novel dataset called the MAHNOB HMI iBUG Mimicry database, or MHi-Mimicry-db for short, created to allow research in the automatic detection of mimicry. Our goal was to collect recordings of behaviour with as many occurrences as possible in which people are acting either identically and simultaneously or with a significant amount of resemblance and/or synchrony.

Besides collecting the expected (mimicry) behaviour, the data should also enable the analysis of those behaviours on a social interaction level. In order to explore the human perception correctly, the analysis of social interaction should at least include the relation between those behaviours, their function, and the intention behind them.

### 2.1 Interaction Scenario

As a general starting point in our database design, specific hypotheses in the field of social psychology determined what kind of scenarios would be suitable for our recordings. We chose to design our recording scenario such that the collected data allows us to test two hypotheses about mimicry that have been posed in the literature:

**Hypothesis 1**. Agreement-/Disagreement-Mimicry occurs in conversations when the participants agree with each other as well as when they do not agree with each other, with a higher frequency or amount of mirroring during agreement than during disagreement. Moreover, mimicry occurs in conversations in with there is the intention to gain acceptance from an interaction partner through conforming to that person's attitudes, opinions, and behaviours [14], [15], [16], [17], [18].

**Hypothesis 2**. Affiliation-Mimicry has the power to improve social interaction. That is: when individuals communicate, one partner who wants to affiliate with others may intentionally engage in more mirroring of them; in contrast, when they want to disaffiliate they intentionally engage in less mirroring [19], [20], [21], [22], [23], [24].

Based on the theoretical foundations of the above two mimicry hypotheses, we designed two conversational scenarios. The first scenario is a debate, and the second scenario is a role-playing game where one participant plays the role of a homeowner who wants to rent out a room, and the other participant plays the role of a student who is interested in renting the room.

### 2.2 Procedure

The recording includes two experiments. In Experiment A, participants were asked to choose a topic from a list. Participants were then asked to write down whether they agree or disagree with each statement of their chosen topic. The discussion is held between the participant and a confederate. Participants are led to believe that the confederate is a fellow naïve participant. Participants were asked to start the conversation by presenting their own stance on the topic, and then to discuss the topic with the other person, who may have different views about the topic.

Every topic has a list of statements regarding that topic associated with it. In the pre-recording assessment, the participants note their (dis)agreement with these statements. This is used as a reference for annotating, possibly masked, opinion or attitude. During the discussion participants and confederates express agreement and disagreement, and show a desire to convince the other person of their opinion.

In Experiment B, the intent was to simulate a situation where two participants want to get to know each other a bit better and need to disclose personal and possibly sensitive information about them in the process. Participants were given a communication assignment that requires self-disclosure and emotional discovery. Participant 1 played a role as a student in university who was looking for a room to rent urgently Participant 2 played a role as a person who owns an apartment and wants to rent one of the rooms to the other one.

Participants are not sure about their partner's preference at the beginning, so the hypothesis is that they will try to get more information from their partners first, only gradually showing more sensitive personal information to the other. Moreover, their conversation partners may not want to expose many details to them until s/he decides whether the participant is someone they like or not. However, they have the same goal, which is to share an apartment, so they have the tendency of affiliation.

To rule out mixed gender effects, experiments included either all male participants and confederates, or all female. After recording both sessions, participants finished a personality questionnaire and two separate experiment questionnaires, which were designed to measure the experienced affect and attitude during the two sessions.

### 2.3 Self-Report of Participants

Nonconscious behavioural mimicry has been explained by the existence of a perception-behaviour link [26]; watching a person engage in certain behaviour activates that behavioural representation, which then makes the perceiver more likely to engage in that behaviour herself. Chartrand & Bargh [27] experimentally manipulated behavioural mimicry to explore the consequences for liking a person. They argued that perception of another person's behaviour automatically causes nonconscious mimicry, which in turn creates shared feelings of empathy and rapport. Perspective taking, or the ability to adopt and understand the perspective of others, is one component of empathy [28]. The ability to take the perspectives of others increases behavioural mimicry, suggesting that individuals who are able to affiliate with group members because of their ability to understand others also routinely use mimicry behaviour [29]. Few researchers use actual social interaction corpora to

detect human postures to recognize mental states. In our experiments we considered that behaviour presentation in interaction is inherently linked to personality traits (confidence, nervousness, etc.) so personality questionnaires have been included.

28 male and 12 female students from Imperial College London (aged 18 to 40years) are participants. Each was paid 10 pounds for participating in the study, which took about 1.5 hours. Two male confederates and one female confederate were from the iBUG group at Imperial College London. All participants were assigned to each other randomly. Four personality questionnaires were finished before attending the experiment. These were: 1) Big-Five Mini-Markers FNRS 0.2, 2) The Aggression Questionnaire including four subscales: physical aggression =.85), verbal aggression (= .72), anger (= .83), and hostility (= .77). 3), Interpersonal reactivity index consisting of four 7-item subscales, including Fantasy (FS), Perspective Taking (PT), Empathetic Concern (EC), and Personal Distress (PD), and 4) Self-Construal scale composed of 15 items made up the Independent self-construal subscale, and the remaining 15 items corresponded to the Interdependent self-construal subscale.

## 3 Synchronized Multi-Sensor Recording Setup

The recordings were made under controlled laboratory conditions using 15 cameras and 3 microphones, to obtain the most favourable conditions possible for analysis of the observed behaviour. All sensory data was synchronized with extreme accuracy.

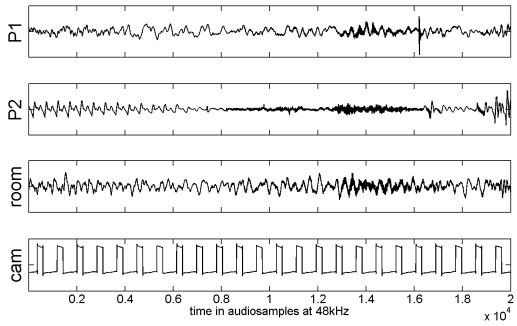


**Fig. 1.** 4 tracks recorded in parallel by the audio interface. From top to bottom: head microphones of participants 1 and 2, room microphone, and camera trigger

### 3.1 Audio Channels

Three channels of sound were recorded using a MOTU 8pre[1] eight-channel interface (Fig. 1). Channel 1 and 2 contain the signal from head-worn microphones, type AKG HC 577 L. Channel 3 contains the signal from an AKG C 1000 S MkIII room microphone. This channel can be used to obtain a noise estimate for noise reduction in the first two channels. Most of the noise originates from the building ventilation system,

which was controlled remotely. The background noise cannot be assumed constant as the ventilation system was sometimes switched on or off during a recording.

### 3.2 Camera Views

Three types of cameras have been used: An Allied Vision Stingray F046B, monochrome camera, with a spatial resolution of 780x580 pixels; two Prosilica GE1050C colour cameras, with spatial resolutions of 1024x1024 pixels; and 12 Prosilica GE1050 monochrome cameras, with spatial resolutions of 1024x1024 pixels. Different sets of cameras have been set up to record the face regions at two distance ranges: 'Far' corresponds to a distance range for upright poses and 'Near' corresponds to forward- leaning poses. The focal length and focus of the cameras have been optimized for the respective distance range. The best camera view to use for a facial analysis depends on a person's body pose in each moment. The cameras were intrinsically and extrinsically calibrated. See figure 2 for the camera views.


**Fig. 2.** Simultaneous views from all the cameras

### 3.3 Audio/Video Synchronization

The cameras are synchronized by hardware triggering [30], and configured to have exposure intervals around the same centre at 58 frames per second. To synchronize between audio and video, we recorded the camera trigger signal as a fourth signal, in parallel with the audio channels. Since the analogue inputs of the 8Pre are sampled using the same clock signal, an event in one of the channels can be directly related to a temporal location in all other channels. The camera trigger pulses can be easily detected and matched with all the captured video frames, using their respective frame number and/or time stamp. The final error of the synchronization is well below 20μs.

# 4 Annotation

The database has been segmented into speech acts, and annotated for a number of social signalling cues, as well as conscious and nonconscious higher-level behaviours.

## 4.1 Segmentation into Episodes of Interest

In our data, Experiment A includes two parts: presentation and discussion. In the presentation part, it is obvious that interviewees play a role as speakers while the interviewers listen all response from listeners is on the involvement or understanding level. For example, understanding can be expressed by nods. So it is natural that the range of nonverbal behaviour expressed by a listener is small, often limited to cues such as nodding, smiling, and certain mannerisms. On the contrary, in the discussion part, interviewers and interviewees both need to express an actual response, i.e. to give feedback on a communicative level. Even more interesting is that people often only mimic another's behaviour when they are playing the same role in interactions. In other words: people may not immediately mimic the speakers' behaviours while listening, and they may, instead, express a consensus response (since they are functioning on the involvement or understanding level). But when the former listener subsequently takes on the role of speaker, s/he often mimics their counterparts' behaviour that was expressed during the previous turn. This complies with one of the most important factors that can affect mimicry - similarity: The similarity of roles played in interactions. In Experiment B, the participants have complete similarity of conversational goal, which is to find a roommate successfully.

In summary, the analysis of relevance among mimicry and social interactions can be extended not only for recognizing human affect, but also for judging relationships (roles) and interaction management (turn-taking).

Annotation Steps:
*Segmentation into episodes according to utterance tokens acquired from participants*
*Annotation of speakers and listeners*
*Annotation of behavioural cues for both participants separately*
*Annotation of mimicry*

In our annotation tool, options for behavioural cues are predefined. After the annotation of episodes and behavioural cues, the tool can automatically compare whether the selected options are the same for both participants, from which the mimicry label (PRESENT/NOT PRESENT) is derived.

## 4.2 Annotation within Segments

For the episodes of interest, more detailed annotations are included, consisting of behavioural expression labels, mimicry/non mimicry labels, and social signal labels. In the interface of the annotation software, the first item that is provided concerns the behavioural expression labels: smile, head nod, headshake, body leaning away, and body leaning forward. When the video data is played, the annotator has to enter the

time when a particular cue was observed, and choose a suitable label from the list. Cases where none of the available labels are appropriate for a certain expression are also taken into other account. Secondly, in order to learn more about the intent behind those behavioural expressions, for each behavioural expression the label of ”conscious” and ”unconscious” is also recorded. For unconscious behaviours, a SOCIAL SIGNAL EXPRESSION has to be chosen. This can be e.g. understanding, agreement, liking, confused, or uncertain. For conscious behaviour, a DESIRED GOAL has to be chosen. For example: to flatter others, to emphasize understanding, to express agreement, to share rapport/empathy, to increase acceptance. Since it is sometimes difficult, or even impossible, to specify a unique reason for mimicry, space is provided to include a comment.

Current annotation considers visual behaviour and participants' roles in each conversation. Further annotation will include the participants' affect and implied social signals relative to mimicry. It will be mainly based on the questionnaires taken during the experiments.

## 5 Overview and Availability

The MHi-Mimicry database is made freely available to the research community through a web-accessible interface (http://www.mahnob-db.eu/mimicry). The dataset consists of 54 recordings. Of these, 34 are of the discussions (Experiment A) and 20 recordings are of the role-playing game (Experiment B). The data contain imagery of 43 subjects (40 participants and 4 confederates). The durations of Experiment A are between 8 and 18 minutes, with an average of 15 minutes. The duration of Experiment B is between 4 and 18 minutes, with an average of 11 minutes. At the time of recording, all the participants ranged in age from 18 to 40 years. Of the participants 26% are female and 95% of the participants come from southern Europe.

All 18 sensor tracks are available in the database, as well as an audio-visual overview track that combines all views and the two audio tracks from the head-mounted microphones (Fig. 2). This overview track is intended for human inspection and labelling of the data. A large amount of metadata is stored in the database, and a search interface makes it possible for researchers to collect the data they require.

## 6 Conclusion and Future Work

This is the first accurately synchronized multimodal database of natural human-to-human interaction aimed at the study and automatic detection of mimicry. Although it is not the first database to address natural human-human interaction, the range of sensors, the multi-resolution synchronized views of both participants, and the high accuracy of the multi-sensor, multi-modal synchronization provides many new opportunities to study (automatic) human behaviour understanding in general. In the future, our work will mainly contribute to affective computing and human-machine interaction. In particular, we aim to contribute to (1) the understanding of how human mimicry works and subsequently, the development of automatic mimicry analyzers,

(2) the improvement of the recognition of social and affective attitudes such as (dis)agreeing and (dis)liking through mimicry information, and (3) knowledge about the timing and the extent to which mimicry should occur in human-machine interaction by generating mimicry behaviour in agents. This technology would also strongly influence science and technology by, for example, providing a powerful new class of research tools for social science and anthropology. While the primary goal of such an effort would be to facilitate direct mediated communication between people, advances here will also facilitate interactions between humans and machines.

## Acknowledgment

This work has been funded in part by the European Community's 7th Framework Programme [FP7/2007–2013] under the grant agreement no 231287 (SSPNet). The work of Jeroen Lichtenauer and Maja Pantic is further funded in part by the European Research Council under the ERC Starting Grant agreement no. ERC-2007-StG 203143 (MAHNOB). The work of Michel Valstar is also funded in part by EPSRC grant EP/H016988/1: Pain rehabilitation: E/Motion-based automated coaching.

## References

1. Pantic, M., Gunes, H.: Automatic, dimensional and continuous emotion recognition. Int'l Journal of Synthetic Emotion, vol. 1, no. 1, pp. 68–99 (2010)
2. Vinciarelli, A., Dielmann, A., Favre, S., Salamin, H.: A database of political debates for analysis of social interactions. IEEE Int'l Conf. Affective Computing and Intelligent Interfaces, vol. 2, pp. 96–99 (2009)
3. Savran, A., Ciftci, K., Chanel, G., Mota, J., Viet, L.H., Sankur, B., Akarun, L., Caplier, A., Rombaut, M.: Emotion detection in the loop from brain signals and facial images. eNTERFACE report (2006)
4. Lichtenauer, J., Valstar, M., Jie, S., Pantic, M.: Cost-effective solution to synchronized audio-visual capture using multiple sensors. IEEE Int'l Conf. Advanced Video and Signal Based Surveillance, pp. 324–329 (2009)
5. Gatica-Perez, D.: Automatic nonverbal analysis of social interaction in small groups: A review. Image and Vision Computing, vol. 27, no. 12, pp. 1775–1787 (2009)
6. McKeown, G., Valstar, M.F., Cowie, R., Pantic, M.: The SEMAINE corpus of emotionally coloured character interactions. IEEE Conf. Multimedia and Expo, pp. 1079–1084 (2010)
7. Littlewort, G., Bartlett, M.S., Fasel, I., Susskind, J., Movellan, J.: Dynamics of facial expression extracted automatically from video. IEEE Int'l Computer Vision and Pattern Recognition Workshop, pp. 80–80 (2004)
8. Coan, J., Allen, J.: Handbook of Emotion Elicitation and Assessment. Oxford University Press (2007)
9. Devillers, L., Vidrascu, L., Lamel, L.: Challenges in real life emotion annotation and machine learning based detection. Neural Networks, vol. 18, no. 4, pp. 407–422 (2005)

10. Carletta, J.: Unleashing the killer corpus: experiences in creating the multi-everything AMI meeting corpus. Language Resources and Evaluation, vol. 41, no. 2, pp. 181–190 (2007)
11. Athanaselis, T., Bakamidis, S., Dologlou, I., Cowie, R., Douglas-Cowie, E., Cox, C.: ASR for emotional speech: Clarifying the issues and enhancing performance. Neural Networks, vol. 18, no. 4, pp. 437–444 (2005)
12. Tiedens, L.Z., Fragale, A.R.: Power moves: Complementarity in dominant and submissive nonverbal behaviour. J. of Personality & Social Psychology, 84, no. 3, pp. 558–568 (2003)
13. Douglas-Cowie, E., Cowie, R., Sneddon, I., Cox, C., Lowry, O., McRorie, M., Martin, J.C., Devillers, L., Abrilian, S., Batliner, A., Amir, N., Karpouzis, K.: The Humaine database: Addressing the collection and annotation of naturalistic and induced emotional data. In Proceedings of the 2nd international conference on Affective Computing and Intelligent Interaction, Springer-Verlag, pp. 488–500 (2007)
14. Brass, M., Bekkering, H., Prinz, W.: Movement observation affects movement execution in a simple response task. Acta Psychologica, vol. 106, no. 1-2, pp. 3–22 (2001)
15. Catmur, C., Walsh, V., Heyes, C.: Sensorimotor learning configures the human mirror system. Current Biology, vol. 17, no. 17, pp. 1527–1531 (2007)
16. Estow, S., Jamieson, J.P., Yates, J.R.: Self-monitoring and mimicry of positive and negative social behaviors. J. of Research in Personality, vol. 41, no. 2, pp. 425–433 (2007)
17. Heider, J.D., Skowronski, J.J.: Ethnicity-based similarity and the chameleon effect. Austin State University, manuscript (2008)
18. Van Swol, L.M.: The effects of nonverbal mirroring on perceived persuasiveness, agreement with an imitator, and reciprocity in a group discussion. Communication Research, vol. 30, no. 4, pp. 461–480 (2003)
19. Baaren van, R.B., Fockenberg, D.A., Holland, R.W., Janssen, L., van Knippenberg, A.: The moody chameleon: the effect of mood on non-conscious mimicry. Social Cognition, vol. 24, no. 4, pp. 426–437 (2006)
20. van Baaren, R.B., Holland, R. W., Steenaert, B., Van Knippenberg, A.: Mimicry for money: Behavioral consequences of imitation. J. of Experimental Social Psychology, vol. 39, no. 4, pp. 393–398 (2003)
21. van Baaren, R.B., Horgan, T.G., Chartrand, T.L., Dijkmans, M.: The forest, the trees and the chameleon: Context dependence and mimicry. J. of Personality and Social Psychology, vol. 86, no. 3, pp. 453–459 (2004)
22. Lakin, J.L.: Exclusion and nonconscious behavioral mimicry: The role of belongingness threat. Ph.D. dissertation (2003)
23. Lakin, J.L., Chartrand, T.L.: Exclusion and Nonconscious Behavioral Mimicry. The social outcast: Ostracism, social exclusion, rejection, and bullying. New York: Psychology Press, pp. 279–295 (2005)
24. Maurer, R.E., Tindall, J.H.: Effect of postural congruence on client's perception of counselor empathy. Journal of Counseling Psychology, vol. 30, pp. 158–163 (1983)
25. Bavelas, J.B., Black, A., Lemery, C.R., Mullett, J.: I show how you feel: Motor mimicry as a communicative act. J. of Personality and Social Psychology, vol. 50, pp. 322–329 (1986)
26. Chartrand, T.L., Jefferis, V.E.: Consequences of automatic goal pursuit and the case of nonconscious mimicry. Philadelphia: Psychology Press, pp. 290–305 (2003)
27. Chartrand, T.L., Maddux, W.W., & Lakin, J.L.: Beyond the perception-behavior link: The ubiquitous utility and motivational moderators of nonconscious mimicry. Unintended thought II: The new unconscious. New York: Oxford University Press, pp. 334-361 (2005)
28. MacDonald, G., Leary, M.R.: Why does social exclusion hurt? The relationship between social and physical pain. Psychological Bulletin, 131, 202 – 223 (2005)
29. Lakin, J.L., Chartrand, T.L.: Using nonconscious behavioural mimicry to create affiliation and rapport. Psychology Science, 14(4), pp. 334 – 339 (2003)
30. Lichtenauer, J., Shen, J., Valstar, M.F., Pantic, M.: Cost-effective solution to synchronized audio-visual data capture using multiple sensors, Report, Imperial College London (2010)