

UNIVERSITY OF TWENTE

MASTER THESIS

How Turn-Taking Influences the
Perception of a Suspect in Police
Interviews

Author:

Rifca Peters
s1266276
post@rifca.nl

Committee:

Dr. ir. Rieks op den Akker
Drs. Merijn Bruijnes
Dr. Mariët Theune

July 25, 2014



Human Media Interaction

Abstract

In human conversations interlocutors do not necessarily adhere to the ‘one-at-a-time’ [17] speaking rule. People often speak at the same time; sometimes this overlapping speech is problematic, sometimes it’s not. Moreover, people do not always take-up the next turn even if directly addressed by a question.

In this thesis we present our study on conversational behaviour in police interviews between a police officer and a suspect. Interviews are a special type of speech exchange system; structured by question-answer adjacency pairs and with pre-defined roles for interlocutors: interviewer or interviewee [18]. Interviews have a different turn-taking system than ordinary conversation; the interviewer can use pre-utterances leading to a question and the interviewee is obliged to wait and respond. Investigative police interviews are a specific type of interview where there might be a lot at stake for the suspect and the suspect can become quite emotional. There is an asymmetric relation in power and control in favour of the police officer; suspects can negotiate power and control by displaying resistance to their role of interviewee [11]. Moreover, suspects have the ‘right to silence’ protecting them from self-incrimination. However, suspects do have a choice between a silent response and verbal non-response. In police interviews violations of turn-taking can be the product of emotion or strategic choice; silences and overlapping speech can have different meanings.

In a previous conversation analysis of police interviews [16] we attempted to find factors that can explain the meaning and function of silences and overlaps in these interviews. In the present study we looked at how the interlocutors in police interviews were perceived by observers. In a controlled experimental setting we attempted to isolate the influence of turn-taking behaviour on the perception of the emotion and interpersonal stance of the interlocutors. This study is part of a project where we develop an embodied conversational agent (ECA) that acts as a suspect in real-time interactions with a human police interviewer. Such an ECA should be capable of showing behaviour appropriate given an internal state of the agent and the context of the conversation. The question then is: how does turn-taking behaviour, realised using synthesized speech, in police interviews influence the perception a human observer gets from a virtual suspect?

To answer this question we set up a controlled perception study with variations of extracts from police interviews of our DPIT-corpus [14]. The variations differed only in the relative timings of the start of the speech, resulting in either

overlapping talk, in a gap between turns or in bridged speech: no overlap and no gap between utterances. The perception of a suspect reported by participants included high individual differences. Participants did report minor differences in perception of the suspect between turn-taking variations: turn-taking with gaps was associated with higher affiliation, face and rapport and lower deception for the suspect; overlaps were associated with higher power for the suspect. There was a difference in influence of turn-taking on the perception of a suspect between police officers and non-police participants: police trainees reported the lowest perception of affiliation, face and rapport if the suspect provided a delayed response resulting in a gap while the other participants reported the lowest perception of affiliation, face and rapport if the suspect started speaking in overlap.

The study received low response and participants ended prematurely. To evaluate the perception study, we conducted a meta-analysis in which people participated while thinking out loud. This revealed that people experienced difficulties in forming an impression of the interlocutors because they missed contextual, non-verbal and prosodic information. Moreover, the disfluency in synthesized speech had a comical effect and influenced the perception of the interlocutors. Last but not least, participants had difficulties noticing the differences between variations and varied their basis for perception of the interlocutors between variations.

To look at how perception was influenced by turn-taking we isolated the turn-taking variations as much as possible. Short audio stimuli were generated with synthesized speech; excluding prosody and non-verbal cues. The resulting stimuli were monotone unnatural conversations which lacked contextual information. The stimuli diverged too much from real-life situations rendering it next to impossible to form an impression of the interlocutors. By isolating turn-taking while keeping context, content and nuisance constant we hoped that perception in turn-taking variations would suffer equally allowing us to uncover the effect of turn-taking on perception without hindrance from other factors. However, confusion by the test situation has influenced the ratings, thereby rendering the effect of turn-taking unclear.

We unceasingly believe that personality, emotional state and interpersonal stance play a role in display of turn-taking behaviour, and turn-taking behaviour to be a factor when forming an impression of an interlocutor. However, turn-taking behaviour is one of many factors of an interlocutor's behaviour in a certain content; it seems impossible to measure the isolated effect of turn-taking on the perception of an interlocutor. Considering the goal of the project —the development of a virtual suspect capable of appropriate (turn-taking) behaviour—we don't need to know the influence of every individual factor instead we might better focus on unravelling the combinations of factors that define a situation and appropriate behaviour to evoke a certain impression or personality.



Voor oma, je zou zo trots zijn.

Preface

September 2013, following a course in conversational agents and a research assignment on turn-taking in police interviews, I dove into this project. I had big dreams and great ambition, but soon the reality check struck me. Graduating is not about grandeur but about learning to scope, concretize, and letting go. This thesis is the result of this long journey and the completion of my master program Human Media Interaction at the University of Twente. The thesis is original work by the author; a work in progress paper of the perception study has been published in the Chi Spar*s proceedings (see Appendix D).

In retrospective, I'm wondering how this project evolved to where it is now. I am always advocating that in Human Media Interaction it is not so much the fundamental truth, but the user experience that counts; in my work I'm happy if there is alignment between the user experience and requirements. At the start of this project I thought to study user experience, but I soon realized we knew too little to prototype and evaluate a model. I started to search for existing knowledge and in my head I tried to solve the puzzle of how people 'decide' their turn-taking behaviour. With every new piece of information the puzzle became more complicated and I got stuck inside my head. And there, disguised as an answer to my puzzle, was the proposal of a structured fundamental study. I shouldn't have been that naive, I know that fundamental research won't solve a phenomenological puzzle, but at least I hoped it would mark the pieces.

Completing this journey would have not been possible without the many people who helped me during the project. I firstly thank my supervisors Rieks op den Akker and Merijn Bruijnes. Without their confidence and support on both process and content I would have never completed this thesis on time. In addition I really appreciate Mariët Theune for her contribution by reading and revising the thesis. I want to thank the Dutch police academy, in particular Imke Rispens and Arend de Vries. They opened their doors for us, taught about police interviewing and provided us with the DPIT-corpus. Moreover, many people are really appreciated for their participation in the study. Last but not least, I want to express my gratitude to Jeroen Berendsen; without his help I wouldn't have been able to program my own survey tool and his support and understanding empowered me to travel on this journey.

I did let go the big dream, scoped, concretized and pulled through. But now this project is finished, the dream is still there and I cannot wait to start my next adventure in hopes of solving a puzzle one day.

Abbreviations

CA	Conversation Analysis
ECA	Embodied Conversational Agent
TRP	Transition Relevance Place
TCU	Turn Constructional Unit
DPIT	Dutch Police Interview Training
FTA	Face Threatening Act
WOz	Wizard of Oz

Contents

1	Introduction	1
I	Literature Review	4
2	Methods	5
2.1	Search Strategy	5
2.2	In- and Exclusion Criteria	5
2.3	Review Procedure	6
2.4	Article Selection	6
2.5	Article Characteristics	7
3	Results	8
3.1	Power	8
3.2	Rapport	11
3.3	Agreement	12
3.4	Deception	12
3.5	Profession	14
4	Discussion	15
4.1	Summary	15
4.2	Limitations and Implications	16
4.3	Hypotheses	17
II	Perception Study	18
5	Related Work	19
6	Methodology	21
6.1	Approach and Procedure	21
6.2	Stimuli	22
6.3	Pilot Study	22
6.3.1	Participants	23
6.3.2	Design	23

6.3.3	Measurement and Data Coding	23
6.3.4	Data Analysis	24
6.3.5	Results	24
6.3.6	Lessons Learned	25
6.4	Design	26
6.5	Measurement	27
6.6	Data Coding	28
6.7	Participants	28
6.8	Data Analysis	28
6.9	Assumptions and Limitations	29
7	Results	30
7.1	Perception per Factor	30
7.1.1	Power	30
7.1.2	Affiliation	32
7.1.3	Face	32
7.1.4	Rapport	33
7.1.5	Deception	33
7.2	Discussion of the Results	34
8	Evaluation	35
8.1	Feedback	35
8.2	Meta-Analysis	36
8.2.1	Stimuli	36
8.2.2	Measurement	38
8.2.3	Approach	39
8.2.4	Tool	40
8.3	Perception Scores	40
8.4	Discussion	42
8.4.1	Summary	42
8.4.2	Limitations and Implications	42
8.4.3	Conclusion	44
9	Conclusion	45
A	Survey	50
B	Stimuli	55
C	Results	60
D	Chi Spar*s	63

Chapter 1

Introduction

In human conversation we try to adhere to a ‘one-at-a-time’ approach. Sacks, Schegloff and Jefferson [17] proposed a systematic for smooth turn-taking, offering a set of rules to provide next-turn allocation to one interlocutor and thereby minimizing gap and overlap. However, moments of overlapping speech or silences occur frequently in human conversation [19].

Interlocutors can perform either one of the actions speaking or being silent. In a two party conversation variations in the activity of both interlocutors results in four possible dialog states: self-speaking, other speaking, none speaking and both speaking [7]. Transitions between dialog states create a conversation interaction pattern. In this report we adhere to the definitions used in [7]:

- Overlap_W A within-speaker overlap; during consecutive speech of one interlocutor the other starts speaking initiating overlapping speech and stops speaking resolving overlapping speech. Resulting interaction pattern: SELF-BOTH-SELF.
- Overlap_B A between-speaker overlap; an occurrence of overlapping speech where a speaker transition takes place. Resulting interaction pattern: SELF-BOTH-OTHER.
- Pause A within-speaker silence; a silence between two utterances of a single interlocutor. Resulting interaction pattern: SELF-NONE-SELF.
- Gap A between-speaker silence; a silence in which a speaker transition occurs. Resulting interaction pattern: SELF-NONE-OTHER.
- Bridged A sequence of latched utterances; neither overlapping speech nor audible silence is present between two utterances of one or more interlocutors.

Turn-taking behaviour is observable by vocal analysis but not every silence or moment of overlapping speech is the same; they are often communicative in their own right [4, 11, 24]. Schegloff [19] defined four classes of overlapping speech: 1) cooperative overlap e.g. by utterance completion to assist a speaker,

2) backchannel to signal e.g. active listenership, 3) non-problematic overlapping speech such as chorus and 4) interrupt where the speaker did not finish the utterance and did not yield the floor. An interlocutor can have different reasons to be silent, especially in police interviews in which the suspect likely wishes to avoid self-incrimination and is legally protected to do so. In previous conversation analysis [16, 14] we looked at the interpretation of turn-taking behaviour in police interviews and found that there is a relation between the interpersonal stance of the suspect and the interpretation of silences. E.g. a silent response from a suspect with a positive stance was interpreted as timidity opposed to a suspect with a hostile stance whose silences were associated with withdrawal.

Contrary to the dynamic turn-taking in human conversation, turn-taking in current natural dialogue systems is often restricted to ‘one-at-a-time’. Conversational agents are limited to listening or speaking and listening is initiated either on a moment predetermined by the system or whenever the user makes a sound; resulting in an unnatural human-system interaction. Exceptions are dialogue systems that allow more free turn-taking behaviour [22].

This thesis is part of a project where we are working towards a computational model for human-like turn-taking behaviour for an embodied conversational agent (ECA). This ECA will act as a suspect in a social skill training serious game supporting police interview training. Rich ECA turn-taking behaviour, including pauses, interrupts, and hesitation, is expected to support a more natural human-system interaction. Moreover, appropriate turn-taking behaviour can support the display of personality and emotional state of the virtual suspect. However, extent and contextual condition of the relation between turn-taking and perception of a suspect agent remains unclear.

This thesis examines the relation between turn-taking and the perception of a suspect in a police interview. Police officers receive training on recognition and strategic use of interactional phenomena such as dominance [2]. Due to this experience, their perception of a suspect may be different from untrained people. By a literature review we collected existing knowledge on factors that influence turn-taking in police interviews. In a perception study we investigated the influence of turn-taking on the impression that observers get from a suspect in simulated police interviews by looking at differences in perception scores for extracts of police interviews in which turn-taking was systematically varied. In an evaluation we assessed the feasibility of the chosen approach.

This report is organized as follows. In part I we present the approach (chapter 2) and the results (chapter 3) of our literature review on factors influencing turn-taking in police interviews. Part I is concluded with a discussion of the results leading to hypotheses about the influence of turn-taking on perception of a suspect in chapter 4. These hypotheses were tested in the perception study presented in Part II. In chapter 5 we present related perception studies. The approach of the perception study is presented in chapter 6 followed by the results in chapter 7 and an evaluation of the study in chapter 8. The report is concluded in chapter 9 with some final remarks.

Part I

Literature Review

Chapter 2

Methods

In [16] we report about previous literature where we investigate the factors that influence the interpretation of meaning of overlapping speech and silence in police interviews. In a systematic literature review we extended this knowledge and investigated the relation between affect and turn-taking in investigative police interviews. In this chapter we present the search strategy, selection process and criteria and the characteristics of the articles meeting these criteria.

2.1 Search Strategy

An active search on Google Scholar and Scopus was conducted between October 29th 2013 and December 23th 2013 for articles published since the year 2000. Our search query (“Conversation Analysis” AND (Police AND (interview OR interrogation))) returned 2330 results on Google Scholar and 21 results on Scopus. Our search query (“Turn Taking” AND (Police AND (interview OR interrogation))) returned 2620 results on Google Scholar and 2 results on Scopus.

2.2 In- and Exclusion Criteria

For inclusion in this review articles were required to be published in English or Dutch. Selected articles needed to incorporate investigative police interviews and consider turn-by-turn interaction between a police officer and a non-professional (suspect, witness or victim). As the study was exploratory the search field was kept broad and all articles that discussed any affective factor possibly contributing to the understanding of turn-taking behaviour were included. Articles that described or evaluated police interviews with a focus on (improvement of) interviewing skills and/or general outcome —without consideration of turn-by-turn interaction— were excluded.

2.3 Review Procedure

Article selection was done in a four-step procedure, this procedure was repeated for each search result. In the first step the titles within the search results were screened, articles with a high chance of failing to meet the inclusion criteria were removed. Also references to textbooks and duplicates for Scopus and Google scholar were removed. In the second step abstracts of the remaining articles were screened, filtering out duplicates between the search queries, articles not available to us and articles failing to meet the inclusion criteria. In the third step the remaining 54 articles were fully read; those articles meeting the inclusion criteria were summarized for documentation in step four.

2.4 Article Selection

The search for (“Conversation Analysis” AND (Police AND (interview OR interrogation))) returned 2351 results of which 242 textbooks were removed and 2023 articles were removed based on title. From the remaining 86 articles based on the abstract were removed: 9 duplicates, 17 articles for which no full text was available to us and 30 articles which did not meet the inclusion criteria. Additionally 21 articles were removed after examination of the full text because they failed to meet the inclusion criteria. The search for (“Turn Taking” AND (Police AND (interview OR interrogation))) returned 2622 results of which 221 textbooks were removed and 2295 articles were removed based on title. From the remaining 107 articles based on the abstract were removed: 46 duplicates and 37 articles which did not meet the inclusion criteria. Additionally 16 articles were removed after examination of the full text because they failed to meet the inclusion criteria. In total four articles that did present a non-affective factor (i.e. typing activity or presence of an interpreter) were excluded. One article presenting a theoretical framework extensively reported in [16] was excluded from this review.

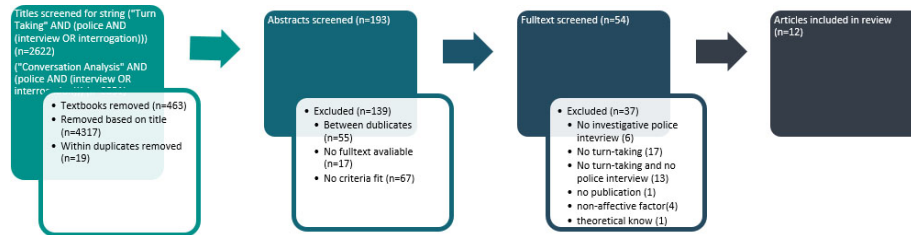


Figure 2.1: article selection procedure and number of articles meeting the in- and exclusion criteria for each step.

2.5 Article Characteristics

Eleven articles presenting turn-by-turn interaction in police interviews have been included in the review. All reported analysis of conversation (of which one a controlled experiment). The conversation analyses included three times Australian, once Dutch, four times British/UK, once Brazilian, once Iranian and once British and/or Swedish conversations. The reported affective factors influencing turn-taking were: 5 times power —of which once related to ethnicity—, once agreement —related to preference—, twice deception —once in multi-suspect interviews—, twice rapport and once profession related to cooperation, power and face. One article did not mention turn-taking but did discuss rapport in police interviews. An overview of the included articles and their characteristics is presented in Table 2.1.

Table 2.1: characteristics of the selected articles. Each row presents the characteristics for one article. Each column presents a characteristic for an article: ‘Article’ name of the author and year of publication, ‘Turn-Taking’ the studied turn-taking behaviour, ‘Factor’ the affective factor related to turn-taking, ‘Method’ research methodology and range (CA=Conversation Analysis) and ‘Country’ origin of the study materials and/or participants.

Article	Turn-Taking	Factor	Method	Country
Benneworth 2009	Hold floor	Power	CA 11 int 1 case	UK
Heydon 2011	Silence	Agreement	CA 13 int.	Australia
Komter 2003	Silence	Deception	CA 20 audio int. 1 case	Netherlands
Vrij et al. 2012	Overlapping speech	Deception (Collective)	Controlled 43 pairs	UK/Sweden
Momeni 2011	Overlapping speech	Power	CA 50 cases	Iran
Yoong 2010	Overlapping speech	Power	CA 1 int.	Australia
Haworth 2006	Silence and overlap	Power	CA 1 case	UK
Ostermann 2003	Silence and overlap	Profession	CA 26 audio int.	Brazil
Jones 2008	Overlapping speech	Power (& Ethnicity)	CA 20 int.	England/Wales
Fogarty et al. 2013	Silence and overlap	Rapport	CA 1 child case	Australia
Abbe & Brandon 2013	–	Rapport	review	–
Walsh & Bull 2011	–	Rapport	CA 142 int.	England/Wales

Chapter 3

Results

In this chapter we present the results of the literature review on the relation between affect and turn-taking in police interviews. The findings from the selected articles are ordered by factor suggested to influence the turn-taking behaviour or perception of this behaviour and presented below.

3.1 Power

In investigative police interviews turn transitions are pre-specified by asymmetric question-answer adjacency pairs. Only the police interviewer has the right to ask questions, thereby having control over the agenda and turn-management.

The possible extent of an officer's power and control over the conversation is demonstrated in Benneworth's analysis of an interview with a paedophilic suspect in the UK [3]. The officer strategically used closed and polar questions designed to elicit responses that align with the polarity of the interrogative. By strategic use of the statement-agreement adjacency pair the need for audible agreement was eliminated. The officer took up the role of storyteller; produced a prolonged account of the offence and maintained the floor and limiting. The contribution of the suspect was limited to infrequent and minimal denials. Contrary to ordinary conversations the officer maintained the floor during pauses in his speech. The officer maintained the floor while thinking what to say by use of: fillers such as 'umm' as delaying devices or by audible in-breaths to signal incompleteness of the story/utterance in progress. An example is shown in Extract 3.1.

According to Momeni [13] the asymmetrical relationship is not only pre-specified but includes local properties too. Momeni distinguished four types of dominance: 1) quantitative dominance (amount of talk), 2) interactional dominance ('strong' vs. 'weak' interactional behaviour), 3) semantic dominance (topic-change) and 4) strategic dominance (important interventions). The pre-defined roles provide the officer with more chances to talk and the possibility to easily interrupt the interviewee. Momeni analysed 50 Iranian court cases and

found nine different types of interruptions used by police officers to get more turns of talk: 1) interrogative interruption, to get information by a question; 2) confirmative interruption, to get confirmation; 3) non-confirmative/informative interruption, to threaten or insult; 4) information-objection interruption, to get information by objection; 5) declarative-information interruption, to get information by giving information; 6) non-information-objection interruption, objection not aimed at getting information; 7) confess interruption, derived from presupposition or from previous confession; 8) cooperative interruption, to cooperate not necessarily to elicit information and 9) corrective interruption, to correct own words.

Yoong [27] considered interruptions by police interviewers a sign of assertion of power. Yoong adhered to the definition of a social norm as: ‘the accepted or required behaviour for a person in a particular situation’ providing an expectation of appropriate behaviour for a given situation. In police interviews the norm for suspects is to have their power reduced, cooperate and fulfil their role as interviewee by answering questions. Interruptions were seen as violations of turn-taking because they break into a turn in progress and challenge the speaker’s control of turn completion. Yoong analysed an Australian conversation between an officer and a suspect and found that interruptions were used to maintain power and control by showing disinterest in certain suspect responses and direct the suspect back to the topic of interest. An example is shown in Extract 3.2.

Haworth [6] claimed that, despite the pre-specified advantage for the police interviewer, power and control are constantly under negotiation and open to challenge and resistance. The author distinguished four features influencing power and control: 1) topic, 2) question type, 3) question-answer adjacency pairs and 4) institutional status. Analysis of a high profile case in the UK—a doctor suspected of murdering numerous of his patients—revealed that control was achieved on a turn-by-turn basis through the following strategies and techniques:

- Recognition-interrupts by the suspect; challenging topic introduction by breaching the expected question-answer adjacency pair and the interviewer’s right to set the topic agenda.
- The suspect giving the illusion of compliance by providing a seemingly legitimate, syntactically correct but un-satisfactory response.
- The suspect challenging control by introducing a reformulated question and providing an answer to that question instead of to the question asked.
- The suspect challenging the status of the officer after signals of incompetence reflected in pauses, hesitations and incompleteness of the officer’s utterance.
- The suspect providing minimal responses (yes/no), even as response to an open explanation-seeking questions.

- The suspect controlling the floor and topic by taking extended turns after an explanation-seeking question on which only the suspect had knowledge.
- The suspect interrupting any attempt of the officer to ask a question during extended turns.
- The officer granting the suspect discursive freedom to talk ‘out of turn’ by not taking the next turn despite pauses, to support the overall goal of the interview to get the interviewee to talk.
- The officer sanctioning an unsatisfactory answer by repetition or rephrasing of the question.
- Topic continuation or discontinuation by the officer despite the response of the suspect.

After analysis of 20 police interviews conducted with Afro-Caribbean and white British suspects in England and Wales, Jones [9] concluded that high frequency of overlaps was attributed to powerful talk and dominance. Interesting is the fact that Jones analysed the data for whether or not the proposition in the overlap were taken up by the interruptee. Responsivity appeared to be lower for the Afro-Caribbean suspects compared to white British suspects granting higher power and control to the police officer in interviews with Afro-Caribbean suspects. According to Jones the racial inequality observed in the overlapping speech was the result of the suspects’ behaviour.

4	Police	okay you’ve been arrested on suspicion of indecently assaulting Sarah .hhhh
6	Suspect	mmm. (2.6)
8	Police	what Sarah does describe is that umm (.)
9		she would regularly go into your house ummm (3.2)
10		quite unexpectedly sometimes ummm (4.1)
11		she talks about going into your front room and...

Extract 3.1: the officer maintains the floor, despite length of pauses, by using fillers (extract taken from [3]).

- 1 Police so how would you describe Kafeel as a as a personality?
- 2 Suspect I mean I know him as a as a professional
...
he doctor ah the university there ah in his ah
umm back home when he was ()
- 3 Police ah hum
- 4 Suspect and he was quite hard working
...
his his parent's are well respected doctors
- 5 Police yeah
- 6 Suspect and ah they are moderate ah moderate umm Muslims
...
I've seen them and-
- 7 Police do do you know how umm ah Kafeel met Bilai?

Extract 3.2: the officer interrupts the suspect to prevent turn completion and direct the suspect back to the topic of interest (extract taken from [27]).

3.2 Rapport

Rapport in police interviews was studied in a review by Abbe & Brandon [1] and the analysis of 142 police interviews by Walsh & Bull [26]. The authors of both studies agree that rapport is a critical step in the development of a professional relationship between an officer and a non-professional. Abbe and Brandon considered rapport a prerequisite for techniques used in police interviews while Walsh and Bull focused on the importance of rapport to gain trust.

Both articles claimed that initial building of rapport is not sufficient. Rapport should be established and maintained to influence the quality and outcome of police interviews. Rapport was built by display of calmness and signs of equality [26]. Rapport can be degraded by note-taking that negatively influenced attention and by pseudo-rapport, i.e. faked positivity and attention [1]. A better perception of rapport by a witness was suggested to increase trust and improve responsiveness, cooperation, agreement and recall of information [1, 26]. A better perception of rapport by suspects was suggested to increase talkativeness and openness [26].

Based on the analysis of an interview with a child witness Fogarty [5] argued that progressivity displayed rapport. Progressivity was characterized by smooth collaborative completion of actions within a given sequence of interaction. Stretches, fillers and pauses in the speech of the interviewee were signs of discomfort and disrupted the progressivity. Strategic use by the officer of continuers to display hearing and understanding and of silences to leave sequential space to the interviewee, helped to restore progressivity.

The institutional character and the asymmetrical relationship in police inter-

views influenced perception of rapport. Abbe and Brandon claimed that, within the positivity component of rapport, perceived status reflected competence (respect) rather than warmth (liking) [23]. Abbe and Brandon looked beyond rapport at operational accord: a relationship in which interlocutors share at least some goals and experience mutual affinity or respect; including agreement on roles, expectations and desired outcomes. Abbe and Brandon claimed that shared understanding of the situation and rules decreases the importance of mutual positivity; coordination remains critical and might take a complementary form where the interviewer is dominant and interviewee submissive. In interest-based compliance the interviewer must establish authority and credibility to gain control and demonstrate competence. Respect must be elicited from the interviewee but need not be reciprocated. In police interviews interviewers have a high degree of control enabling them to exercise interest-based social influence: the ability to influence by promising that there is something to gain by compliance.

3.3 Agreement

Heydon [8] analysed what happened when suspects attempted to provide a silent response to a question or request in investigative police interviews and concluded that suspect contributions were constrained in allowable turn type. Non-immediate responses to accusations were not interpreted as denial by the police interviewer and the interviews were continued by the officer as though the suspect accepted the accusation. An example is shown in Extract 3.3 where the suspect neither agreed nor contradicted the attribution of having seen the broken glass while the police officer continues assuming acceptance. Alternatively a suspect can access the ‘right to silence’ by providing a non-preferred response i.e. offering verbal non-conformation.

3.4 Deception

Komter [10] showed that the turn-taking pattern and interpretation of silence depended on the stage of the interview. The author investigated persuasion to admission of guilt and found that the suspect was strategically transformed to confessing in four stages. Each stage resulted in a different turn-taking and interpretation of silences. At the start of the interview the suspect was in fact deceptive and displayed resistance by evasive or defensive responses while the officer displayed distrust. The suspect repeatedly provided a polite and complete sentence but failed to provide a satisfactory and legitimate response to the question asked. Requests for self-blame were not taken-up by the suspect. Silent responses were repaired by the officer by addition of injunctions. In the second stage the officer was less accusatory and the suspect was more compliant. The officer attempted to get the suspect to agree with the unreasonableness of her behaviour. The suspect provided syntactically correct responses. Nevertheless

the officer was not satisfied by the answers and invited the suspect to elaborate. This second stage was characterized by longer stretches of talk by the interviewer and long silences. These silences were allowed by the officer to highlight the inadequacy or absence of the suspect’s answers. Silences after invitation for self-blame were interpreted as non-contradicting. Gaps after a suspect’s response were used to highlight the damaging implications of the answer. The officer took responsibility for the silences and initiated repairs, giving the officer control over the length of silences and their effects. Examples of these silences are shown in Extract 3.4. In the third stage the officer attempted to connect to and draw conclusions from earlier statements from the suspect. During the discussion of these earlier statements, both interlocutors have the same knowledge of the topic. In this stage the suspect was cautious and withholding and displayed a lack of understanding. The silences were shorter and no gaps were present after the suspect’s responses. Silent responses by the suspect were interpreted as non-response. In the last stage the suspect was persuaded to tell the truth story and admitted to the crime. Non-verbal responses were highlighted by the officer resulting in a long silence followed by an explicit request by the officer for a verbal response from the suspect.

Vrij et al. [25] reported the results of a controlled study on cues for deception in multi-suspect interviews and found interruptions as a factor. The authors analysed investigative interviews of 21 pairs of truth tellers and 22 pairs of liars with a total of 86 participants (25 male, 61 female) and coded interruptions. Truth tellers made more interruptions ($M = 8.57, SD = 8.45$) than liars ($M = 2.73, SD = 2.96$). Additionally, truth tellers said more in the interview ($M = 1544$ words, $SD = 763$) than liars ($M = 1039$ words, $SD = 531$). The results were consistent with the hypothesis that truth tellers adopt a ‘tell all’ approach. Interruption between speakers to add information and correct each other were present in joint recall. For liars the ‘keep it simple’ approach was common which resulted in fewer interactions between suspects.

433 Police uh you saw the glass shatter to the ground,
(0.4)
434 Suspect I just kept walking.
(0.2)
435 I just got in the car AND ROB
(0.6)
436 me friend said what the hell’s going on,
(0.4)
437 whadcha do,
(1.2)
438 Police so you didn’t bother saying anything to them.=
439 =that that the glass was broken, or.

Extract 3.3: the suspect neither agrees nor contradicts the attribution of having seen the glass while the officer continues assuming acceptance of the attribution (extract taken from [8]).

1 Police ... jij begint haar te bijten=
 2 =alles ... waar je dochters bij staan.
 (77)
 3 wat vind je daar nou van
 (1)
 4 Suspect niet goed natuurlijk
 (5)
 5 Police waarom doe je het dan
 (9)
 6 geef me es één reden waarom dat je dit hebt gedaan
 (2)
 7 Suspect ik weet niet
 (2)
 8 Police voor die ene a- armband die jij wou hebben,=
 9 Suspect =ja
 (1)
 10 Police ja was dat het=
 11 Suspect =ik denk dat dat het was ja
 (6)
 12 Police vind je het niet belachelijk

Extract 3.4: the officer's pauses draw attention to the suspect's position of 'no-defence' after an accusation (line 2) or 'officially absent' response to a question (line 5), gaps after the responses by the suspect (lines 4 and 11) highlight the damaging implications of the answers (extract taken from [10]).

3.5 Profession

Analysis of intragender differences in face work between female police officers and feminist social workers in institutionalized conversations with female victims in Brazil showed different interaction patterns in police interactions compared to social workers [15]. Ostermann considered silent responses less cooperative. Silent responses endanger the smooth flow of the conversation and create opportunities for turn-taking violations that constitute face-threatening acts. Ostermann found that: 1) officers strategically used silences to gain control over the conversation; by being silent the officer forced the victim to speak, 2) silent responses by officers occurred in clusters or as response type while those of social workers were characterized by single occurrences and 3) after a silent response by the victim social workers took up responsibility for the flow of the conversation while police officers did not necessarily do so. Ostermann reported police officers using interruption and topic change as a strategy to gain control and signal an answer to be unsatisfactory. The author concluded that interactional patterns were predicted by task-orientedness and ideological stances. Police officers were more likely to provide non-responses and change topic, social workers used more cooperative strategies.

Chapter 4

Discussion

The results from the literature review provided us with some suggestions of factors related to turn-taking in police interviews. In this chapter we discuss the implications of, and hypotheses derived from these findings.

4.1 Summary

Yoong [27] showed that police officers interrupted suspects to prevent them from turn completion. These deliberate interruptions were considered signs of assertion of power [13, 27]. Due to the asymmetric question–answer adjacency pairing, police interviews are structured to provide the officer with control over the conversation [3]. Haworth [6] showed that power was under constant negotiation and reported recognition interrupts, minimal responses, taking extended turns, and interruptions of questions as techniques used by suspects to gain control in police interviews. Vrij [25] suggested that truth tellers adopt a ‘tell all’ approach resulting in a talkative mood opposed to liars who adopt the ‘keep it simple’ approach resulting in a less talkative mood. A more in-depth analysis of silence during stages of deception and truthfulness by Komter [10] suggested that resistance by evasion or defence was a sign of deception and silences after a statement or question were associated with a non-contradicting position of the suspect. This absence of denial was often highlighted by the officer by allowance of a long silence. To be considered relevant, denial should be provided immediately following or interrupting an accusation [8]. Rapport is considered a critical step in eliciting trust and building a relationship in professional interaction and therefore a prerequisite for techniques used in police interviews, e.g., to get cooperation from the interviewee [1, 26]. Suspects talked more openly in harmonious interactions and cooperation and agreement were increased. Discomfort —considered a lack of rapport— was displayed by stretches, fillers and pauses in the speech of the suspect [5].

4.2 Limitations and Implications

Articles that included discussion of the relation between power and turn-taking mostly agreed on the existence of a pre-specified asymmetrical relationship granting higher power and control to the police interviewer. However, the effect of the role of the interlocutor on turn-taking remained unclear. Interruption was reported a sign of power for both officer [13, 27, 15] and suspect [6]. Silent responses were considered less cooperative for both officer and suspect [15] and minimal (silent) responses by a suspect were considered an attempt to gain power and control [6].

Most research is done from the perspective of the police officer. The reasons reported for turn-taking behaviour by the police officer were mostly strategic. An exception were silences while forming a sentence; these silences were accompanied by strategic fillers to signal incompleteness of the utterance [3]. Strategic reasons for silence were to: highlight absence, inadequacy or damaging implication of a suspect's response [10]; gain control by forcing the other to speak [15]; or restore rapport by providing the other with chances to speak [5]. Types of strategic overlapping speech were: interruptions to gain control by prevention of turn completion [27] or to signal a response to be unsatisfactory [15]; and backchannels to encourage elaboration [3]. Little is known on how this behaviour was interpreted by a suspect, Haworth [6] reported silence and hesitation in officer speech being interpreted as a sign of incompetence. The reasons for a suspect's turn-taking behaviour were less clearly strategic. Silent responses can be included in an evasive strategy to display resistance [10]; evasive verbal responses were common too. However, evasion is not necessarily a strategic choice and instead can be a reaction to emotions such as fear and shame. Strategic turn-taking behaviour was present in: silences used to give the illusion of compliance with the role of interviewee and interruptions to challenge power and control by interrupting topic or question initiation [6]. The 'right to silence' makes interpretation of silence an interesting subject for studies. Interpretation of interruptions have not been included in any of the studies. Silences by a suspect were interpreted as: a signal of low rapport [26], damaging cooperation and agreement [26, 1]; a signal of discomfort, indicating a low level of rapport [5]; acceptance of an accusation [8]; a non-contradicting response [10] or the absence of a response [10].

The interpretation of a silent response by a suspect was shown to be dependent on the type of officer utterance. Consistent with the preferred response in a statement-agreement adjacency pair, non-response of a suspect following a statement by the officer was interpreted as agreement [8, 3, 10]. In an accusation-denial adjacency pair a non-response was interpreted as absence of denial and thereby confirmation [8]. These findings suggest that timing and placement of denial are key to its recognition and that the choice to access 'right to silence' is a high risk strategy for suspects. However, Heydon does not make a distinction between truthful and deceptive suspects. Neither did Heydon consider on whom the burden of proof rests; as suggested to be of influence when interpreting the meaning of silence [4]. The common assumption that a submissive

interlocutor is passive and less talkative seems to be moderated in police interviews. We argue that a submissive suspect complies with the role of interviewee and the preference to take up the next turn in a question-answer adjacency pair, whereas a dominant or hostile suspect displays resistance by withholding a response.

4.3 Hypotheses

Each hypothesis predicts, based on the literature, how a variation in turn-taking influences the perception of the suspect's power, rapport and deception. We formulated hypotheses following the same pattern for each of these factors: a turn-taking feature influences the perception of a factor. For the definition of turn-taking features we adhere to the classification of Heldner and Edlund [7]; *Pause*: a silence between two consecutive utterances of one interlocutor, *Gap*: a silence between utterances of two different interlocutors, *Overlap*: a moment of overlapping speech where two interlocutors speak at the same time and *Bridged*: sequential utterances without an audible silence in-between. Specifications for turn-taking patterns and additional interpersonal or contextual factors are included if predictions were suggested to be true only under these circumstances.

Hypothesis 1 *In interactions where a suspect initiates Overlap, the suspect is perceived to have higher power than when a suspect provides a Bridged response or a delayed response resulting in a Gap.*

Hypothesis 2 *In interactions where a suspect provides a silent response to an open-question, resulting in a Pause between sequential officer utterances, the suspect is perceived to have higher power than when a suspect provides a vocal and syntactically correct response.*

Hypothesis 3 *In interactions with an audible Pause between consecutive suspect utterances, the perceived level of rapport is lower than when sequential turns were Bridged.*

Hypothesis 4 *In interactions with an audible Pause between sequential suspect turns, the suspect is perceived as more deceptive than in Bridged sequential turns.*

Hypothesis 5 *In interactions with a Gap between a question from an officer and the answer by a suspect, the suspect is perceived as more deceptive than in Bridged or Overlap up-take of question-answer adjacency pairs.*

Hypothesis 6 *In interactions where a suspect continues talking in Overlap, maintaining or receiving the floor, the suspect is perceived as having more power than when a suspect yields the floor.*

Part II

Perception Study

Chapter 5

Related Work

The influence of turn-taking on the impression that people get from an agent was studied earlier in a perception study by Ter Maat and Heylen [20]. They showed that variations in turn-taking strategies create different impressions of e.g. friendliness, rudeness and arousal. The authors argued that turn-taking norms put on constraints that prevent people from being rude or impolite and that turn-taking is influenced by emotions. Based on this assumption they suggested that observing turn-taking behaviour can provide information about the personality, identity and feelings of a person.

Ter Maat and Heylen defined turn-taking by two features: *start-up* and *overlap resolution*. Start-up described the moment an agent started speaking. Overlap resolution described the agent's action during overlapping speech. For both features three strategies were defined based on the work of Schegloff [19]. The start-up strategies were: exactly when the other agent finished speaking (*At*), with some delay (*After*) or before the other agent is finished (*Before*). The overlap resolution strategies were: stop speaking (*Stop*), continue normally (*Normally*) or continue with raised voice (*Raised*). In a conversation simulator two agents were scripted to use a combination of strategies of the two features (excluding *Before+Stop*) resulting in different interaction patterns. The resulting interaction patterns corresponded with the silence and overlap classification by Heldner and Edlund [7] introduced in Chapter 1. *At* resulted in bridged utterances, *After* in a gap, *Before+Stop* in overlap_W and *Before+Normally or Raised* in overlap_B . The speech of the agents contained prosodic features but no recognizable content (words). Participants were asked to rate one agent on several 5-point scales measuring personality, emotional state and interpersonal stance.

Based on the ratings, Ter Maat and Heylen concluded that the *Before* strategy was perceived as more unfriendly, rude, cold and active compared to start speaking *At* or *After*. An agent starting *At* the end of the other agent's speech was perceived as the most pleasant. An agent starting *After* was perceived as the most submissive. When including both start-up and overlap resolution strategies the scores for positivity, friendliness, agreeability, respect, pleasant-

ness, attentiveness, warmth and responsibility were highest for the *At+Stop* condition. The scores on negativity, unfriendliness, disagreeability, rudeness, distance, unpredictability, un-attentiveness and cold were highest for the *Before+Raised* condition.

In a follow-up study [21] the authors showed that it is possible to alter the impression that people have of an agent by variation in turn-taking. Participants actively interacted with a virtual interviewer. The interviewer anticipated the participant’s end of turn and asked a next question using one of the three start-up strategies (*At*, *After* or *Before*). Interviewer gender and conversation topic were varied between participants. Afterwards participants were asked to rate the interviewer on 27 semantic differential scales measuring personality, emotion, social skill and interview skill. The scales were grouped by four factors: agreeableness, assertiveness, conversational skills and rapport.

The authors concluded that starting *Before* the end of the participant’s turn evoked a perception of a less agreeable and more assertive interviewer. Starting *After* had opposing associations evoking an impression of a more agreeable and less assertive interviewer and a perception of higher rapport. A significant difference was found between the ratings for the male and female interviewer. However, it was unclear if this difference was caused by gender or the quality of synthesized speech.

In the first study speech with no recognizable content was used, excluding the influence of context and content. The second study differentiated from the first in the specific interview setting and participant’s active involvement in the conversation. The latter study did not look at the influence of the context and possible meaning conveyed in silence. Both studies presented similar results suggesting a relation between turn-taking and the perception of personality and emotion independent from engagement, context and content.

We extended these perception studies by including police interview specific content and roles. In earlier conversation analysis [16] we showed that contextual factors such as the role of the interlocutor, interview phase and topic played an important role in the interpretation of silence and overlap. Based on the literature review, we argue that within the police interview setting there are some differences in the relation between turn-taking and perception of an agent compared to cooperative conversations. Moreover, we focus on the perception of the suspect interviewee whereas in [21] the participants were asked to rate the interviewer.

Chapter 6

Methodology

The goal of the perception study is to discover how variations in turn-taking influence the impression people get from a virtual suspect in police interviews. Based on extracts of police interviews between a police officer and a suspect we created audio stimuli with variance in timing of turn-taking. These stimuli were presented to participants who were asked to fill in a survey on their impression of the suspect's power, affiliation, face, rapport and deception after each stimulus. A pilot study was conducted to evaluate the stimuli and survey. In this chapter we present the methodology of the perception study. First we describe the approach and stimuli, next we describe the details and results of the pilot study, then the details of the main perception study are presented and the chapter is concluded with the assumptions and limitations of the study.

6.1 Approach and Procedure

The chosen online survey approach enabled people to participate independent of time and place and is considered suitable to gather quantitative data from a larger group of people. Measurement of perception by various 7-point Likert scales (see section 6.5) provides a response format suitable for comparison between different participants and/or stimuli.

Participants were instructed to browse to the survey website on a computer with loudspeakers. The participants were ensured confidentiality of their data and provided with limited information about the study. To avoid biasing participants they were not informed about our interest in turn-taking variations. On the survey website the participants played an audio stimulus and provided a rating of their impression of the interlocutors by scoring the Likert scales. The participants were allowed to play the audio file repeatedly. After completion of their ratings the participants could advance to the next stimulus.

Table 6.1: characteristics of the stimuli. Each row presents the characteristics for one extract. Each column presents a characteristic for the extract: ‘Extract’ extract number, ‘Source’ origin of the extract, ‘S gender’ gender of the suspect in the created stimuli, ‘H’ hypothesis tested with the extract, ‘Original’ sub number of the stimulus with the original turn-taking, ‘Variation1’ sub number of the stimulus and the first variation in turn-taking and ‘Variation2’ sub number of the stimulus and the second variation in turn-taking.

Extract	Source	S gender	H	Original	Variation1	Variation2
E1	vanBron1a—10:18.938	male	1, 5	b) bridged	c) overlap	a) gap
E2	vanBron1a—25:57.224	male	1	c) overlap	b) bridged	a) gap
E3	Wassink1—6:44.258	female	1, 5	b) bridged	c) overlap	a) gap
E4	Huls3—29:03.210	female	1, 5	a) gap	b) bridged	c) overlap

6.2 Stimuli

We selected extracts from the DPIT-corpus [14] that demonstrate or contradict one or more of the hypotheses from section 4.3. The selected extracts contained conversations that occur in real police interviews ensuring a certain degree of realism in content and interaction. For each extract altered versions were created in which the turn-taking was systematically adjusted while maintaining the content of the conversation as much as possible. Names were replaced by fictive names of similar length. Minor adjustments were made to keep turn-taking patterns and content consistent.

The stimuli were generated using text to speech to minimize the influence of prosody and because the intended game environment will include synthesized speech. We used Ivona (<http://www.ivona.com>) text to speech because this software provided the most fluent words and sentences of the software available to us. All stimuli were generated using a single male and a single female Dutch voice to minimize voice bias. Interlocutors in a single stimulus were of opposing gender to increase identification; gender of officer and suspect were counterbalanced over all stimuli. Utterances were recorded and edited to control timing using Audacity (audacity.sourceforge.net).

An overview of the stimuli is presented in Table 6.1. Case descriptions and extract transcriptions are available in Appendix B.

6.3 Pilot Study

We conducted a small pilot study testing Hypothesis 1 to evaluate the approach and check the feasibility of stimuli and measurement. Two groups of participants completed a survey containing two stimuli and provided feedback afterwards. In this section we present an excerpt of the methodology and results.

Table 6.2: stimuli for and design of the pilot study.

Extract	Source	original	group A	group B
E1	vanBron1a—10:18.938	b) bridged	b) bridged	b) bridged
E2	vanBron1a—25:57.224	c) overlap	c) overlap	b) bridged

6.3.1 Participants

The participants of the pilot study were 8 (former) HMI or Psychology students (5 male, 3 female, aged 25+ with the majority between 25 and 34 years old (n=5)). All participants were naive to the goal of the study. Participants were equally divided in groups A (n=4) and B (n=4), homogeneity of variance was violated for age.

6.3.2 Design

The design of the pilot study was a 2*2 between-subject design. All participants listened to a baseline stimulus E1_b. The second stimulus was systematically varied between groups A (E2_c original overlap, n=4) and B (E2_b bridged, n=4). After each stimulus the perception of both interlocutors was measured. Feedback on the approach, stimuli and measurement was gathered afterwards in an unstructured interview.

6.3.3 Measurement and Data Coding

To measure the perception of power, affiliation, face, rapport and deception, various statements and bipolar adjective pairs were presented after each stimulus. In total 36 scales were presented; 6 statements, 15 bipolar pairs for the suspect and the same 15 bipolar pairs for the officer. Perception scores were measured on a numeric interval scale [1-7]. Values were corrected to align polarity. A missing value was coded 0. Perception scores were collected by a free-to-use online survey tool (<http://www.thesistools.com/web/?id=395827>).

Feedback on the study was written down arranged by topic (textual, scales, stimuli, procedure or tool). Similar feedback provided by different participants was clustered and tallied in a response matrix.

Bipolar pairs:

- Affiliation: friendly-hostile, together-opposed, cooperative-competitive
- Power: dominant-submissive, strong-weak
- Face: polite-impolite, respectful-disrespectful, autonomous-dependent, approving-disapproving
- Rapport: attentive-inattentive, close-distant, positive-negative
- Deception: deceptive-honest, rejecting-accepting, resistant-cooperative

Statements:

- Rapport: coordination, harmony, attention
- Deception: defensiveness, 2*evsion

6.3.4 Data Analysis

The data consisted of perception scores collected at interval level. The data was not normally distributed, therefore non-parametric tests were done. We checked for spurious relations using two-tailed Spearman correlation on perception scores for the suspect in E1. One-tailed Spearman correlation was done to check for correlations between scales designed to measure a single factor. The hypotheses were tested with Mann-Whitney test on perception scores for the suspect in E2.

6.3.5 Results

The survey response data contained 8 independent measures of perception, measured by 36 scales (6 statements, 15 bipolar pairs for the suspect, 15 bipolar pairs for the officer), for two stimuli. Values for variance, standard deviation and error of mean were high. Missing values were detected, one participant failed to complete the survey, two participants missed one or more scales.

Correlation

The Spearman correlation statistics reported non-significant correlations for E1 between perception scores and group or age. Exceptions were significant correlations between: group and perception scores on “accepting–rejecting” ($r_s = -.77, p(\text{two-tailed}) < 0.05$), age and perception scores on “accepting–rejecting” ($r_s = -.73, p(\text{two-tailed}) < 0.05$) and age and scores on evasiveness ($r_s = .80, p(\text{two-tailed}) < 0.05$). The presence of one or two correlations out of 21 scales could occur based on chance; no influence of unsystematic variation in age or differences between groups was assumed.

The Spearman correlation statistics reported a significant correlation for E1 between all scales designed to measure affiliation. The two scales designed to measure power showed no significant correlation. The four scales designed to measure face showed a significant correlation between four out of six scale pairs. The six scales designed to measure rapport showed low correlation; only three pairs were correlated significantly. Correlation statistics for the six scales designed to measure deception showed a total of six significant correlations: two between the three statements, one between the three bipolar pairs and three between combinations of one statement and one bipolar pair. With an exception for affiliation, correlation between scales intended to measure a single factor was weak; no assumption could be made that the scales indeed measured a single underlying factor. Moreover, significant correlations were reported between scales designed to measure different factors, e.g. between scales for face and affiliation.

Hypothesis Testing

Hypothesis 1 predicted a difference between the two groups for the perception scales measuring power: “submissive–dominant” and “strong–weak”. Higher values for group A were expected. No significant difference was reported between perception scores for “submissive–dominant” for group A ($Mdn = 6$) and B ($Mdn = 7$), $U = 3.00$, ns , $r = -.45$. The perception scores on strong–weak” for group A ($Mdn = 6$) did not differ from group B ($Mdn = 6.5$), $U = 4.00$, ns , $r = -.29$. Hypothesis 1 was not supported.

Participant Feedback

After completing the survey participants were asked to report any errors or noteworthy observations. Participants reported problems disambiguating some of the bipolar scales and negative statements. The audio quality (and overlapping speech or grammar errors) had a negative effect on hearing and understanding of speech. Four participants reported voice bias; perceiving the male voice as more pleasant. Some participants experienced difficulties understanding the task at hand and needed the baseline stimulus to get acquainted. The survey tool proposed some limitations: audio files were only accessible in Chrome web browser, questions couldn’t be marked as required, question numbers were not consecutive and the tool offered confusing marketing statements at completion.

6.3.6 Lessons Learned

The high variance, standard deviation and error of mean in perception scores increased the probability that differences were based on chance. No assumption can be made that differences between groups were caused by systematic variations. Two participants reported voice bias in favour of the male interlocutor, this was reflected in the data by a more positive perception of the male suspect compared to other participants. The possible influence of this bias was high due to the low sample number. A larger dataset is needed to find and remove or correct outliers. Moreover, perception is subjective and the extent to which people respond to differences in turn-taking is likely to increase the individual variation in perception scores. The sample of the pilot study was too small to ground any conclusions. We expect differences to be subtle and possibly be overshadowed by individual variation. A within-subject design allows us to look at individual differences between perceptions scores for variations in turn-taking.

Significant correlation was present between scales designed to measure a single factor for affiliation and face, indicating that the scales did measure a single underlying factor. The statements designed to measure rapport were reported to be ambiguous and showed no correlation with the other scales for rapport. Added value of these statements could not be found and might best be removed. During re-evaluation of the literature on rapport we discovered that perception of competence and respect were suggested to reflect the positivity component of rapport within the police interview setting [1]. Additional scales

for these aspects are expected to improve the measurement of perception of rapport. The scale “close–distant” was reported to be ambiguous and should be reformulated. While designing scales to measure face and rapport careful consideration of the different components of these factors is necessary. Correlations between the scales to measure deception were scarce. A possible explanation is that detection of deception is difficult; especially without non-verbal cues. The absence of correlation between the scales intended to measure power could be explained by the ambiguity of the “strong–weak” scale as reported by participants. Alternatively, we suggested a statement measuring the aspect of control as associated with power. Some significant correlations between scales designed to measure different factors were present, this can be explained by the relation between these factors. E.g. a positive influence of rapport on cooperation and agreement as suggested by Abbe and Brandon [1].

Based on the results of the pilot study some changes were made to the design and measurement of the main perception study. Additionally, a new online survey tool was created. Due to the individual differences in perception scores comparison between groups is next to impossible; for the main perception study we changed to a within-subject design. Some scales were removed or introduced as suggested above. The resulting scales are presented in Section 6.5. Additionally, some minor changes in formulation of scales and grammar of statements were made based on the feedback of the participants. Moreover, the order of scales was slightly changed to cluster all scales concerning the same interlocutor. We introduced an exercise audio file and scales to allow participants time to get acquainted with the synthesized voices and survey format. A custom online survey is created to fit the study requirements (e.g. marking all scales required) and allow a clear presentation.

6.4 Design

The main perception study was designed with a mixed design; comparing scores for four different extracts (E1, E2, E3, E4), each with three variations: a) gap, b) bridged and c) overlap (see Section 6.2). A limitation of four stimuli for each participant was proposed to limit the time needed to complete the survey to under 30 minutes. The first three extracts had a 3*3 within-subject design to avoid unsystematic individual differences between participants. The three extracts were divided among three groups; within one group all participants rated all three variations of a single extract. The order in which variations were presented was counterbalanced within each group. The last extract was compared between subjects; each group scored a different variation of E4. An overview of the study design is presented in Table 6.3.

Participants were assigned to a group and order based on the current total number of survey website visitors; the first visitor was assigned group A order 1, the second visitor group B order 1, the third visitor group C order 1, the fourth visitor group A order 2 etc.

Table 6.3: perception study design.

Group	Order	1 th Stimulus	2 nd Stimulus	3 th Stimulus	4 th Stimulus
A	1	E1 _a	E1 _b	E1 _c	E4 _a
A	2	E1 _b	E1 _c	E1 _a	E4 _a
A	3	E1 _c	E1 _a	E1 _b	E4 _a
B	1	E2 _a	E2 _b	E2 _c	E4 _b
B	2	E2 _b	E2 _c	E2 _a	E4 _b
B	3	E2 _c	E2 _a	E2 _b	E4 _b
C	1	E3 _a	E3 _b	E3 _c	E4 _c
C	2	E3 _b	E3 _c	E3 _a	E4 _c
C	3	E3 _c	E3 _a	E3 _b	E4 _c

6.5 Measurement

To measure how the interlocutors were perceived, various statements and bipolar adjective pairs were presented after each stimulus. All statements and pairs were measured on a 7-point semantic differential scale.

The scales included among others the characteristics of interpersonal stance [2] and the aspects of rapport [23]. The scales were based on and validated by the work of Ter Maat [20, 21], reviewed by police academy personnel and were evaluated in a pilot study (see section 6.3.6). Scales were counterbalanced for polarity. The scales were the same for each stimulus —except for references to gender which were altered to comply with the gender of the interlocutors in the stimulus under assessment. The impression of the suspect was measured by all scales; the impression of the police officer was measured by a selection of 6 scales to minimize the length of the survey. An overview of the scales is presented below; scales selected for the officer are depicted in *italics*. The full survey (in Dutch) is available in Appendix A.

Bipolar pairs:

- Affiliation: *friendly-hostile*, together-opposed, *cooperative-competitive*
- Power: *dominant-submissive*
- Face: *polite-impolite*, autonomous-dependent, *approving-disapproving*
- Rapport: positive-negative, attentive-inattentive, close-distant, *competent-incompetent*, respectful-disrespectful, careless-responsible
- Deception: deceptive-honest, rejecting-accepting, resistant-cooperative

Statements:

- Power: in control
- Deception: defensiveness, evasion

6.6 Data Coding

All perception scores were coded by counting numbers 1-7, where a higher value was associated with high perception of that aspect of a factor e.g. 7=friendly and 1=hostile. Ratings for counterbalanced scales were transformed to align with other scales measuring the same factor. A missing value was coded 0. Missing values were excluded from the analysis.

6.7 Participants

90 police academy trainees were contacted by police academy personnel and asked to participate in the study. 4 people responded to this call of which 2 completed the survey. Additionally 3 police academy personnel participated of which 2 completed the survey. Other participants were contacted through a mailing list of the University of Twente or social media; 16 people responded to this call for participation.

In total 13 participants completed the survey (male $n=6$, female $n=7$). 4 participants belong to the target group; police officers or trainees. All but one of the 9 non-police participants were students (MBO $n=2$, HBO $n=3$, WO $n=3$)¹. Although the survey requested a name of the study only three participants provided this information (Computer Science $n=1$, Mathematics $n=1$, administration $n=1$). Three external participants reported prior experience with police interviewing, once as a witness and twice as both witness and victim. None of the participants reported experience as suspect or professional. One external participant reported to have participated in a similar study before. All participants were naive to the goal of the study and native or proficient Dutch language users. Participants were divided in three groups; each group rating different stimuli (see section 6.4). Due to the algorithm for group assignment and non-participating website visitors the participants were unequally divided in groups A (police $n=2$, non-police $n=2$), B (non-police $n=2$) and C (police $n=2$, non-police $n=5$).

6.8 Data Analysis

The number of participants was too low for the intended statistical analysis. Instead we did a qualitative analysis of the data looking for differences in individual perception scores between turn-taking variations and for agreement in differences between participants in a single group.

¹For an explanation of the Dutch education system we refer to: http://en.wikipedia.org/wiki/Education_in_the_Netherlands

6.9 Assumptions and Limitations

A potential limitation of our study is the use of auditory-only stimuli. This removed the interference of non-verbal behaviour. However, non-verbal behaviour is undoubtedly important for the perception of an agent and will be available in the envisioned game environment. Studies on the perception of ECAs that incorporate verbal and non-verbal agent behaviour are required in the police domain.

The stimuli were generated using synthesized speech as will be the case in the envisioned game. The quality of the synthesized speech had a negative effect on hearing and understanding of the speech. Sound fragments such as “uhmm”, backchannels and partial words appeared mostly problematic and therefore were excluded during creation of the stimuli. Although we are interested in perception based on turn-taking, from the pilot feedback we learned that people did focus on understanding the content. We cannot prevent people from incorporating content and context into their judgement. Contextual information was documented for all stimuli allowing us look at the possible influence of content and context on differences in perception scores.

The role of the participants was different from the role users have in the intended game environment. In the present study the participant was a third person observer whereas in the game they will be actively involved in interaction with the virtual suspect. The work of Ter Maat [20, 21] reported more or less equal results in perception studies with observing and interacting participants but reported higher influence of gaps on perception for interacting participants compared to observers. This suggested that any effects found in observers are expected to be present when actively involved too.

All stimuli were short extracts (between 25 and 40 sec). To the best of our knowledge, no research has been done to investigate the relation between vocal stimuli length and perception agreement. In [14] we showed that inter-annotator agreement was low for single utterances, but global patterns became evident over longer periods.

We assumed that all participants were familiar with and have a somewhat similar understanding of the adjectives used in the scales. This assumption is based on the usage of terminology that is widely accepted in study of the factors, but is threatened by the necessity to translate all adjectives to Dutch. The Dutch translations were designed to align with police academy study materials and was reviewed by police academy personnel. However, not all participants were affiliated to the police and might have a different understanding of the used terms. Moreover, the training police officers receive on recognition and strategic use of interactional phenomena such as dominance [2] might influence their perception of an interlocutor. It is feasible and was planned for to have police officers participate in the perception study.

Chapter 7

Results

In this chapter we present the results of the qualitative analysis of the data (see Appendix C) obtained from the perception study. We analysed the perception scores for each participant and checked if the variations in turn-taking result in a different perception of the suspect on each of the factors.

7.1 Perception per Factor

To analyse the perception of each factor we computed the median score for all scales designed to measure a single factor (see Table 7.1). Factors with a high variance (> 2) showed low agreement between individual scales (depicted in grey in Table 7.1). Some details were not reflected in the median perception scores. Therefore, we counted the combined perception scores of all individual scales designed to measure a single factor and analysed the differences in combined perception scores between turn-taking variations. A visual representation of these differences is depicted in Figure 7.1. Below we present, for each factor, the differences in perception scores for the suspect between stimuli with varying turn-taking.

7.1.1 Power

The median scores from participants in group A and B reported only minor differences in perception of power between turn-taking variations ($\leq 1, 5$). EA1 and EA2 reported a slightly higher perception of power if the suspect started speaking with some delay resulting in a gap; this contradicts hypothesis 1. Visual analysis of the combined perception scores from participants in group A and B showed weak and inconsistent differences in perception of power; suggesting no influence of turn-taking on the perception of power.

The median perception scores from five participants in group C reported the highest perception of power when the suspect started speaking in overlap and the lowest perception of power in turn-taking with gaps. PC2 reported

Table 7.1: median scores for perception of the suspect for each participant, factor and turn-taking variation. Rows depict perception scores from one participant. Columns depict the data for each participant: ‘Participant’ participant number (P=Police academy trainee or personnel, E=non-police participant), ‘Stimuli’ rated stimuli ordered by presentation order, ‘Median perception score’ median of perception scores for scales designed to measure a single factor (grey depicts a variance > 2), ‘Power’ median scores for power, ‘Affiliation’ median scores for affiliation, ‘Face’ median scores for face, ‘Rapport’ median scores for rapport, ‘Deception’ median scores for deception, ‘a’ median scores for gap, ‘b’ median scores for bridged and ‘c’ median scores for overlap.

Participant	Stimuli	Median perception score														
		Power			Affiliation			Face			Rapport			Deception		
		a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
PA1	E1 _a , E1 _b , E1 _c	1,5	2	2	6	6	6	4	4,5	5	5	5	6	1	1	1
PA2	E1 _b , E1 _c , E1 _a	6	5,5	6	2	2	2	3,5	3,5	3,5	4	4	3	6	5	6
EA1	E1 _a , E1 _b , E1 _c	4	4,5	3,5	5	5	3	5	4,5	3,5	5	5	5	2	3	4
EA2	E1 _c , E1 _a , E1 _b	4	3,5	3,5	7	7	5	4	5	3,5	6	5	6	2	2	1
EB1	E2 _a , E2 _b , E2 _c	5,5	5	5,5	1	1	1	4,5	4,5	4,5	2	3	5	5	3	3
EB2	E2 _c , E2 _a , E2 _b	4	4	4	3	3	3	3	3	3,5	3	3	3	5	5	5
PC1	E3 _b , E3 _c , E3 _a	4	5	6	2	5	4	3	4	3,5	3	4	4	3	5	3
PC2	E3 _b , E3 _c , E3 _a	2	5	5	1	5	5	3	5,5	4,5	2	5,5	4,5	1	5	5
EC1	E3 _a , E3 _c , E3 _c	2,5	3,5	4	4	6	4	5,5	4,5	3	6	4	2	2	5	7
EC2	E3 _a , E3 _c , E3 _c	3	4	5,5	7	5	3	4	4	4	5	5	3	1	2	6
EC3	E3 _b , E3 _c , E3 _a	2	5,5	6,5	7	7	2	5	6,5	4	7	7	2	2	1	6
EC4	E3 _c , E3 _a , E3 _b	3,5	4	4,5	4	4	4	4	4	3	4	4	4	3	4	4
EC5	E3 _c , E3 _a , E3 _b	3	4	5	6	3	3	4,5	4	4	5	4	3	2	3	6

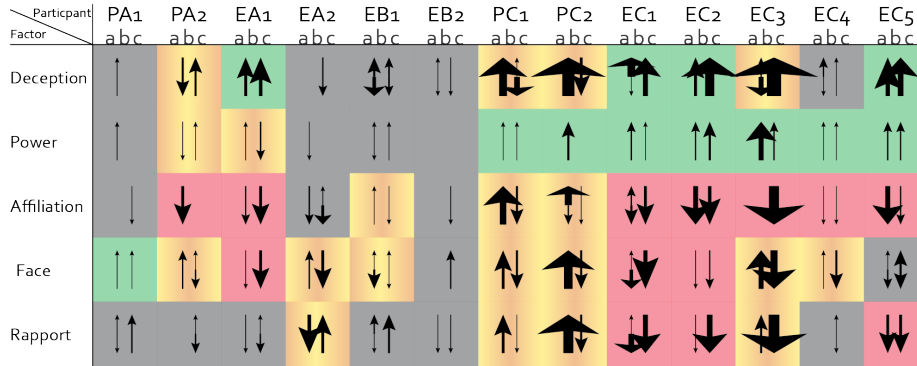


Figure 7.1: visual representation of the combined perception score differences for each participant and each factor between turn-taking variations: a) gap, b) bridged and c) overlap. Arrows depict the direction and amount of difference between two variations; a wider arrow indicates a stronger difference. Colours support the direction: grey weak or inconsistent, green increasing, red decreasing, yellow to orange gradient increasing and orange to yellow gradients decreasing.

low perception of power in turn-taking with gaps but did not report a difference in perception between bridged and overlapping turn-taking. The combined perception scores from participants in group C showed a weak but consistent increase in perception of power when the suspect started speaking sooner; the lowest perception of power was reported for turn-taking with gaps and the highest perception of power was reported for overlapping turn-taking. The results from group C support hypothesis 1; suggesting a higher perception of power if a suspect starts speaking in overlap.

7.1.2 Affiliation

Median scores from non-police participants in group A reported the lowest perception of affiliation when the suspects started speaking in overlap. The combined perception scores from participants in group A showed a decrease in perception of affiliation if the suspect started speaking sooner; perception of affiliation was decreased for bridged turn-taking compared to gap and decreased for overlapping turn-taking compared to bridged. The results from group A suggest a negative influence of speed of turn-taking on the perception of affiliation.

Median scores from participants in group B showed no differences in perception of affiliation. Visual analysis of combined scores showed weak and inconsistent differences.

Median perception scores from EC2, EC3 and EC5 reported, similar to group A, the lowest perception of affiliation if the suspect started speaking in overlap. However, EC2 and EC5 reported lower perception of affiliation in bridged turn-taking compared to gap too. PC1 and PC2 reported the lowest perception of affiliation if the suspect started speaking with a delay resulting in a gap. The combined perception scores showed lower perception of affiliation if the suspect started speaking in overlap compared to bridged turn-taking for all participants in group C. A decrease in perception of affiliation in bridged turn-taking compared to gap was present for all non-police participants. Contradicting, the police trainees (PC1 and PC2) showed a strong increase in perception of affiliation in bridged turn-taking compared to gap. The resulting turn-taking variations with lowest and highest perception of affiliation for each participant are depicted in Table 7.2. Police and non-police participants showed a noteworthy difference of perception; police trainees perceived the least affiliation in turn-taking with gaps whereas other participants perceived the least affiliation in overlapping turn-taking.

7.1.3 Face

The median scores from participants in groups A and B showed only minor differences in the perception of face between turn-taking variations ($\leq 1,5$); no consistent direction was found. The combined perception scores of face showed weak and inconsistent differences between turn-taking variations for participants in groups A and B.

The median scores from PC1 and PC2 reported a slightly higher perception of face in bridged turn-taking and the lowest perception of face in turn-taking with gaps. For non-police participants in group C the differences between median scores were minor ($\leq 1,5$) and showed no consistent direction. The combined perception scores showed a lower perception of face if the suspect started speaking in overlap compared to bridged turn-taking. However, PC1, PC2 and EC3 reported a strongly increased perception of face for bridged turn-taking compared to gap; they differed in strength resulting in a different hierarchy of perception of face for turn-taking variations. The resulting turn-taking variations with lowest and highest perception of face for each participant are depicted in Table 7.2. There is some agreement was perception of face within police participants and within non-police participants, but the influence of turn-taking on the perception of face remains unclear.

7.1.4 Rapport

The median perception scores reported no agreement on perception of rapport between participants in group A. The combined perception scores of rapport showed weak and/or inconsistent differences between turn-taking variations for participants in groups A and B. If any conclusions can be made based on these results, it is that the extracts or scales were not suitable to measure perception of rapport.

The median scores from participants in group C showed a weak tendency of lower perception of rapport if the suspect started speaking in overlap. However, the police trainees reported opposing scores perceiving lowest rapport in turn-taking with gaps. The combined perception scores showed lower perception of rapport for overlapping turn-taking compared to bridged for 6 of the 7 participants in group C. PC1, PC2 and EC3 reported an increase in rapport for bridged turn-taking compared to gap while other participants reported a decrease in rapport for bridged turn-taking compared to gap.

7.1.5 Deception

The median scores from participants in groups A and B showed no consistent difference in perception of deception between turn-taking variations. The combined perception scores for deception showed weak and/or inconsistent differences between turn-taking variations for group A. No differences were visible in group B.

The median scores reported higher perception of deception for overlapping turn-taking compared to gap for 6 out of 7 participants in group C. The combined perception scores showed an increase in perception of deception if the suspect started speaking sooner for participants EC1, EC2 and EC5. EC3 reported a decrease in perception of deception in bridged turn-taking compared to gap but a major increase in perception of deception in overlapping turn-taking compared to bridged. PC1 and PC2 reported a major increase in perception

Table 7.2: turn-taking variations with the lowest and highest perception score per factor for each participant. “—” depicts absences of a single highest or lowest perception score for that factor.

Participant	Power		Affiliation		Face		Rapport		Deception	
	High	Low	High	Low	High	Low	High	Low	High	Low
PA1	—	a) gap	—	c) overlap	c) overlap	a) gap	c) overlap	—	—	a) gap
PA2	—	b) bridged	a) gap	—	c) overlap	a) gap	—	c) overlap	—	b) bridged
EA1	b) bridged	c) overlap	a) gap	c) overlap	a) gap	c) overlap	a) gap	c) overlap	c) overlap	a) gap
EA2	a) gap	—	a) gap	c) overlap	b) bridged	c) overlap	a) gap	b) bridged	—	c) overlap
EB1	c) overlap	—	b) bridged	—	—	b) bridged	c) overlap	a) gap	a) gap	—
EB2	—	—	—	c) overlap	c) overlap	—	a) gap	c) overlap	—	c) overlap
PC1	c) overlap	a) gap	b) bridged	a) gap	b) bridged	a) gap	b) bridged	a) gap	b) bridged	a) gap
PC2	—	a) gap	b) bridged	a) gap	b) bridged	a) gap	b) bridged	a) gap	b) bridged	a) gap
EC1	c) overlap	a) gap	a) gap	c) overlap	a) gap	c) overlap	a) gap	c) overlap	c) overlap	a) gap
EC2	c) overlap	a) gap	a) gap	c) overlap	a) gap	c) overlap	a) gap	c) overlap	c) overlap	a) gap
EC3	c) overlap	a) gap	b) bridged	c) overlap	b) bridged	c) overlap	b) bridged	c) overlap	c) overlap	b) bridged
EC4	c) overlap	a) gap	b) bridged	c) overlap	b) bridged	c) overlap	—	—	c) overlap	a) gap
EC5	c) overlap	a) gap	a) gap	—	—	a) gap	a) gap	c) overlap	c) overlap	a) gap

of deception for bridged turn-taking compared to gap and a decrease in perception of deception for overlap compared to gap. The resulting turn-taking variations with lowest and highest perception of deception for each participant are depicted in Table 7.2. Eight participants agreed on the lowest perception of deception in turn-taking with gaps. These results do not support hypothesis 5; predicting a higher perception of deception if a suspect provided a delayed response resulting in a gap between the utterance of a police officer and the suspect.

7.2 Discussion of the Results

Some minor differences in perception scores between variations of turn-taking were present. The differences reported by participants in groups A and B were minor and inconsistent, indicating no influence of turn-taking on the perception of the suspect. Strong and consistent differences were reported in group C; indicating turn-taking variations for E3 have more influence on perception of the suspect than variations for E1 and E2. Possible explanations are a more suitable original extract or a better implementation of turn-taking variations in the stimuli.

Hypothesis 1 predicted a higher perception of power if the suspect started speaking in overlap with the police officer. The hypothesis was supported by the results from group C. Hypothesis 5 predicted a higher perception of deception if the suspect provided a delayed response resulting in a gap. The hypothesis was not supported by the results from group C.

Noteworthy are the contradicting perception scores between police academy personnel or trainees and other participants. E.g. in group C the police trainees reported the lowest perception of affiliation, face and rapport if the suspect provided a delayed response resulting in a gap while the other participants reported the lowest perception of affiliation, face and rapport if the suspect started speaking in overlap. This supports our suggestion (see section 6.9) that the training police officers receive influences their perception of a suspect.

Chapter 8

Evaluation

In this chapter we evaluate the perception study. The survey response was low and people tended to stop (10 out of 23) before completion. A meta-analysis was conducted to investigate what caused people to stop prematurely.

Seven people who had not participated in the perception study before completed the survey while thinking out loud, after completion an unstructured interview was conducted to discuss the survey and participant's thoughts. People who participated in the study before were asked to report errors, problems and other noteworthy findings; six participants answered to this call for feedback.

8.1 Feedback

Feedback was provided by six participants who participated in the study before (3 male, 3 female, 1 police officer, 5 non-police participants).

One participant who is affiliated to the police provided critical notes on the feasibility of the approach. Stating that perception is largely generated based on non-verbal behaviour and prosody, he argued that absence of this information created a large gap between reality and the test situation rendering it next to impossible to provide a rating for perception of the interlocutors.

Two participants provided feedback on the scales. One participant, with low education level, reported problems understanding the adjectives. One participant reported puzzlement while rating the scales, he had difficulties to decide if overlapping speech was dominant because the interlocutor was taking the lead or helpful because a quick response was provided.

The five participants not affiliated to the police were asked to report if they finished the survey and why they had chosen to stop or continue. One participant reported to have completed the survey without considering to quit. One participant reported to have quit prematurely, the reason given for this was:

- Similarity of stimuli and scales caused a lack of concentration and were interpreted as technical failure (n=1).

Three participants reported to have completed the survey but admitted to have considered quitting, reasons given for this were:

- Similarity of stimuli and scales caused a lack of concentration and were interpreted as technical failure (n=2).
- Difficulties in understanding the conversations due to the synthesized monotone speech, word repetition and/or overlapping speech (n=2).
- Too many scales (n=1).
- Too many answer possibilities (7) for each scale (n=1).

One participant reported to have completed the survey because the differences in time length of the stimuli provided reassurance that the survey worked correctly and similarities were intended.

8.2 Meta-Analysis

A total of 7 people who had not participated in the study before participated in the meta-analysis (male n=5, female n=2, all non-police).

Participants were asked to complete the survey while thinking out loud and being observed by the researcher. Participants were instructed to complete the survey as they normally would have done and to ignore the presence of the researcher. After completion of the survey participants participated in an unstructured interview discussing their thoughts and actions. During this interview participants were informed about the goal of the meta-analysis.

8.2.1 Stimuli

All participants made remarks about the synthesized speech present in the stimuli. All but one participant started laughing while listening to one or more stimuli. Participants experienced hindrance in hearing and understanding of speech which made it difficult to provide a rating for perception of the interlocutors. Participants were confused by the similarities in the first three stimuli which influenced their rating for perception of the interlocutors.

Understanding of Speech

One participant reported that the speech sounded peculiar, but was able to understand the conversation and did not experience hindrance from it. Six participants reported to experience hindrance in understanding due to the synthesized speech by:

- variation in speech pace (n=1);
- lack of prosody (n=3);
- lack of fluency (n=3) and

- sounding like foreign speech (n=1).

Three participants reported that the overlapping speech influenced the hearing and understanding of the conversation. One participant reported word or syllabi repetition to hinder hearing and understanding.

Five participants replayed one or more stimuli. One participant did not repeat the sound file because that would take too much time.

Influence of Understanding on Perception

Three participants reported that the difficulties in understanding impeded their ability to provide a rating for perception of the interlocutors. Five participants reported that their impression of the interlocutors was influenced by the synthesized speech in one or more of the following manners:

- The monotone synthesized voice sounded unfriendly (n=2).
- The lack of prosody hindered forming of an impression of the interlocutors (n=3).
- The pauses in synthesized speech sounded unnatural (not needed) (n=1).
- The perception of an interlocutor might be based on prejudice due to a lack of understanding (n=1).

Stimuli Similarity

Two participants noticed the differences between stimuli without being confused. Three participants noticed minor differences between the stimuli but were confused by the similarities in the stimuli. Two participants noticed no apparent difference between the three variations of one extract. One participant noticed the difference between the second and the third stimulus, resp. gap and bridged turn-taking. The following differences were observed and mentioned:

- M1 The second stimulus (E2_c) was more chaotic and the third stimulus (E2_a) had non-overlapping turn-taking which was perceived as more cooperative.
- M2 The second stimulus (E3_c) included faster paced responses resulting in overlapping speech and the third stimulus (E3_a) included slower turn-taking.
- M3 The second stimulus (E1_a) contained no overlapping speech perceiving the suspect as more cooperative.
- M5 The second stimulus (E1_c) included interrupting speech by the suspect perceiving the police officer as more submissive and the third stimulus (E1_a) contained non-overlapping turn-taking which was perceived friendlier, trustworthy and equivalent.

- M6 In the first stimulus (E1_c) the police officers was leading which was perceived dominant, the second stimulus (E1_a) included no overlapping speech but unnatural long silences were reported and the third stimulus (E1_b) was similar but with slightly shorter silences.

Influence of Similarity on Perception

All but one participant reported that the similarity of the stimuli had influenced their decision on ratings for perception. Participants expressed the following thoughts:

- M1 The ratings for the first stimulus (E2_b) were based on content, ratings for the second stimulus (E2_c) were based more on behaviour and the ratings for the third stimulus (E2_a) elicited more hesitation due to explicit considerations of different factors.
- M2 Was inclined to provide the same rating for the second stimulus (E3_c) based on the impression from the first stimulus (E3_b) because of the similarities. However, minor changes were expected; the ratings for the first stimulus (E3_b) were based on first impression and the ratings for the second stimulus (E3_c) included a more conscious consideration of the influence of faster turn-taking.
- M3 The ratings for the third stimulus (E1_b) were based on the impression of the second stimulus (E1_a).
- M4 No differences were noticed; rating the same stimulus twice was considered useless resulting in low motivation.
- M5 Was inclined to rate each stimulus as if all information was new and provided a rating relative to the previous stimulus.
- M6 The ratings for the third stimulus (E1_b) were loosely based on the perception from the second stimulus (E1_a) while taking into account that the silences were a bit shorter.
- M7 No differences were noticed between stimuli; rating for the second stimulus (E2_a) and the third stimulus (E2_b) were based on memory of perception from the first stimulus (E1_c).

8.2.2 Measurement

Three participants reported that the scales were hard to rate based on a short sound fragment. One participant selected the middle score if a decision felt impossible. Two participants felt the competence scale did not match with the information provided in the stimuli. One participant felt the politeness scale did not match with the information provided in the stimuli. Two participants made a remark about the large amount of scales per stimulus.

Six participants explicitly mentioned that they based their responses on content by:

- Saying “ze zegt...” (translation: she is saying...) while providing a rating for a specific scale (n=1).
- Mentioning being unable to provide a response because the conversation did not include substantive content (n=2).
- Explicit reference to the content being perceived as polite (n=1).
- Mentioning one interlocutor was leading because the other was taking-up the topic (n=1).
- Mentioning rating based on competence on the topic of the conversation (n=1).

Six participants explicitly mentioned that they included behaviour to base their rating:

- Making a remark on shifting focus from content to behaviour between stimuli (n=1).
- Making a remark on shifting focus from content to behaviour between scales (n=2).
- Conscious consideration of the influence of changes in speed of turn-taking (n=2).
- Mentioning to provide a rating based on interaction (n=2).
- Mentioning to consider behaviour as factor to select a rating (n=1).

8.2.3 Approach

Two participants (both noticed no difference between stimuli) reported doubts regarding the goal of the study; one participant suggested that influence of frustration was under investigation and one participant suggested that the researchers were interested in the mean of individual ratings for a single stimulus.

Three participants reported that the similarity of stimuli would have caused them quit prematurely if no researcher had been present, reasons given for this were:

- The feeling that the study is useless (n=1).
- Doubting the relevance of the study (n=2).
- Assuming a technical error (n=1).

Four participants explicitly mentioned that they were pleased that the fourth stimulus was a different extract.

Two participants suggested a between-subject design, presenting four different extracts.

8.2.4 Tool

A technical error was observed once. The second sound file did not play, the participant was able to complete the survey by downloading the file and playing it locally. During this session an error occurred at the end: the demographic questions were not loaded.

8.3 Perception Scores

All participants of the meta-analysis completed the survey, their perception scores are depicted in Table 8.1. In this section, for each participant, we compare the remarks made during survey completion and/or interview with the perception scores for the first three stimuli; three variations of one extract.

M1 reported to perceive the third stimulus ($E2_a$) as more cooperative compared to the second stimulus ($E2_c$). This was consistent with the perception scores; indicating higher perception of affiliation, positive face and rapport and lower perception of deception and power for the suspect for $E2_a$ compared to $E2_c$.

Participant M2 mentioned to have been inclined to repeat the rating provided for the first stimulus. This was consistent with the presence of only minor differences in perception scores between stimuli.

Participant M3 noticed no difference between the second stimulus ($E1_a$) and the third stimulus ($E1_b$). This was consistent with the data showing almost identical perception scores. The difference between overlap and no overlap in the first ($E1_c$) and second ($E1_a$) stimulus was noticed and resulted in higher perception scores on affiliation, face and rapport, lower perception scores on deception and slightly lower perception scores for power for the suspect for $E1_a$ compared to $E1_c$.

Participant M4 noticed no difference between the first three stimuli. The data showed major differences in perception scores between stimuli on various scales but no clear pattern was apparent.

Participant M5 was very conscious of the differences between the three stimuli and reported to perceive the police officer as more submissive in $E1_c$ and the suspect as more friendly and trustworthy in $E1_a$. This was consistent with the perception scores for the suspect in $E1_c$ which were lower for affiliation and higher for power and deception compared to $E1_b$. The officer was perceived higher on affiliation and lower on power in the $E1_c$. In turn-taking with gaps ($E1_a$), perception of affiliation and face were higher and the perception of power and deception was lower for the suspect compared to overlap ($E1_c$).

Participant M6 reported a difference between the first ($E1_c$) and the second ($E1_a$) stimuli, considering the interrupted officer in the first stimulus leading and dominant. This was consistent with high perception scores for the police officer for the first stimulus. Participant M6 noticed only a slight difference between the second ($E1_a$) and the third ($E1_b$) stimuli. However, perception scores for the third stimulus ($E1_b$) showed more resemblance to the first stimulus ($E1_c$)

than to the second stimulus (E1_a).

Participant M7 noticed no difference between the three stimuli and reported to have provided a rating based on memory of perception from the first stimulus. However, a pattern similar to that of M6 was present in the data; showing differences between E2_c and E2_a and similarities between E2_c and E2_b. Participants M6 and M7 rated different extracts but were presented with the same order of variations.

Table 8.1: perception scores from the participants of the meta-analysis. Each row depicts the scores for one stimulus by one participant. Each column depicts the perception scores on one scale.

Participant	Order	Stimulus	Statements			Suspect Adjectives																Police Adjectives					
			Evasion	Defensiveness	In Control	Friendly-Hostile	Together-Opposed	Cooperative-Competitive	Dominant-Submissive	Polite-Impolite	Autonomous-Dependent	Approving-Disapproving	Positive-Negative	Attentive-Unattentive	Close-Distant	Competent-Incompetent	Respectful-Disrespectful	Deceptive-Honest	Rejecting-Accepting	Resistant-Cooperative	Careless-Responsible	Together-Opposed	Cooperative-Competitive	Dominant-Submissive	Competent-Incompetent	Polite-Impolite	Approving-Disapproving
M1	1	E2 _b	7	7	2	1	1	2	7	1	5	1	1	4	2	3	1	5	7	7	3	2	3	6	5	4	3
	2	E2 _c	7	7	5	1	1	1	7	1	5	1	1	4	2	3	1	6	7	7	2	3	3	3	3	3	3
	3	E2 _a	5	5	4	3	2	2	4	2	4	3	2	4	3	4	3	5	5	5	4	6	5	4	6	6	3
	4	E4 _b	5	5	4	5	5	5	2	5	1	4	5	5	3	5	5	3	3	3	5	5	5	6	0	0	0
M2	1	E3 _b	4	5	2	2	3	4	5	2	4	3	3	4	3	4	3	5	5	5	4	5	4	5	5	5	3
	2	E3 _c	5	5	3	3	3	4	5	3	4	3	3	4	3	4	3	5	5	5	3	5	5	4	5	5	4
	3	E3 _a	5	5	3	3	3	4	4	4	4	3	3	3	3	4	4	5	5	5	4	5	4	5	5	5	4
	4	E4 _c	3	5	3	3	4	3	5	4	5	4	4	5	4	4	3	5	5	6	4	5	5	5	5	5	3
M3	1	E1 _c	3	5	4	3	3	3	5	4	5	3	4	4	3	4	4	4	4	5	4	5	5	4	4	6	5
	2	E1 _a	3	3	4	5	5	5	4	5	4	4	5	5	4	5	4	5	4	3	3	5	3	4	4	3	3
	3	E1 _b	3	3	4	5	5	5	4	5	4	5	5	5	5	4	5	4	3	3	5	3	3	4	4	3	3
	4	E4 _a	5	5	3	3	3	3	3	3	5	4	3	4	3	4	3	4	4	5	3	4	5	4	4	5	4
M4	1	E1 _a	2	2	5	3	3	3	3	3	4	4	3	3	3	3	4	5	4	5	5	5	3	5	6	5	4
	2	E1 _b	5	5	3	3	3	3	3	3	5	5	3	5	3	3	3	3	5	5	4	3	3	6	6	5	3
	3	E1 _c	5	5	3	3	3	3	5	3	5	3	3	3	3	3	3	5	5	5	3	3	3	5	3	3	3
	4	E4 _a	6	5	2	3	3	5	3	3	5	3	3	3	3	3	4	5	3	5	3	5	5	6	5	5	5
M5	1	E1 _b	3	3	2	3	6	4	3	6	4	3	4	3	3	4	5	3	3	3	4	3	5	6	6	5	4
	2	E1 _c	3	3	3	3	3	3	5	5	4	3	3	3	3	5	4	5	5	5	5	5	5	3	6	5	4
	3	E1 _a	2	3	4	6	5	6	4	5	4	5	4	5	5	5	5	3	3	2	5	6	6	4	6	6	5
	4	E4 _a	5	5	3	3	3	4	5	3	5	3	3	2	3	3	3	6	5	5	3	6	6	6	6	5	4
M6	1	E1 _c	3	2	5	6	6	6	3	5	2	6	6	7	4	6	5	2	2	2	4	6	6	6	5	4	5
	2	E1 _a	2	1	4	6	7	7	3	6	2	5	4	5	3	2	5	3	3	3	3	7	6	5	3	2	6
	3	E1 _b	2	2	2	6	6	6	3	5	3	5	5	6	5	2	5	2	2	2	5	5	6	5	6	4	6
	4	E4 _a	7	7	5	2	1	2	6	2	5	2	2	6	1	4	2	5	7	7	2	5	6	5	2	6	3
M7	1	E2 _c	7	6	3	2	2	3	6	1	3	2	2	2	3	2	2	5	6	6	3	2	2	7	5	3	3
	2	E2 _a	6	6	5	2	2	2	5	2	5	2	3	2	2	3	2	5	5	6	3	3	3	6	5	3	2
	3	E2 _b	6	6	5	2	2	3	6	2	5	2	1	2	2	3	2	5	5	6	3	3	3	5	5	3	3
	4	E4 _b	6	7	2	5	5	3	2	6	2	3	3	5	5	5	5	4	3	3	5	5	5	6	3	5	3

8.4 Discussion

The results of the meta-analysis provided us with some information to back our suggestions as to why participants quitted prematurely. In this section we discuss the implications of the meta-analysis.

8.4.1 Summary

From the meta-analysis we learned that:

- People not always noticed the differences between stimuli as a result they thought that a (technical) failure occurred or they felt disinclined to answer the same scales over again.
- The synthesized speech lacked prosody, hindering forming of an impression of the interlocutors.
- Overlapping speech hindered hearing and understanding, this effect might be strengthened by the synthesized speech.
- Disfluency in speech by overlapping speech, silence and repetition was considered unnecessary in synthesized speech, causing a comical effect.
- People switched between content and behaviour to determine their rating, the weight of each factor in the forming an impression of the interlocutors remains unclear.
- People who did noticed the variation in turn-taking between stimuli sometimes provided a general description of influence on impression; however, this was not always clearly visible in the data.
- Differences between overlap and no overlap were noticeable for more participants than differences between bridged turn-taking and turn-taking with gaps.

8.4.2 Limitations and Implications

In the present study we are interested in the effect of turn-taking behaviour on perception of a suspect. To measure this effect we isolated turn-taking behaviour as much as possible; non-verbal behaviour and prosodic features were excluded from the stimuli. The isolation of turn-taking and synthesized speech is clearly different from real-life experiences. By keeping content and context similar while varying turn-taking we kept noise similar for all ratings allowing us to compare ratings and uncover the influence of the varied turn-taking. However, the similarity of the stimuli confused participants and influenced their perception and the factors they based their rating on. If people would have based their rating solely on behaviour we would have expected to see differences between ratings for different turn-taking variations. If people would have based

their rating solely on content we would have seen no difference between stimuli. No clear patterns in differences between stimuli were visible in the data, indicating that people indeed changed the basis of their rating per scale and/or stimulus. When people based their rating for one stimulus on content while for another stimulus on behaviour comparing the ratings is not useful.

The speech errors common in human speech seem to be less forgivable when included in synthesised speech. Synthesized speech is able, and thus was expected, to be fluent causing a comical effect when including speech errors. This comical effect was possibly distracting people, hindering the forming of an impression. Moreover, the dissimilarity in reaction to and perception of speech errors in synthesized speech versus human speech means that any results from the test situation cannot be assumed to be true in real-life situations and vice versa.

We noticed that students trained in research and relatives who participated to do the researcher a favour were inclined to neglect the awkwardness of the test situation; they did as asked and completed the survey. Police officers participated from a different perspective, questioning the relevance of an experimental setting derived from actual police interviews. They are trained to pay attention to all verbal and non-verbal signals and not used to interpreting an isolated factor.

The methodology of the present study was based on previous successful perception studies by Ter Maat [20, 21]. These studies had a similar goal: to investigate how turn-taking influences the perception people get from an agent to support the development of a model for an ECA executing turn-taking strategies to evoke a certain impression or personality. Underlying is the assumption that turn-taking violations occur in everyday conversations and that these violations are influenced by emotions and relate to a certain internal state. A minor difference was present in the scope of turn-taking strategies. In [20] both timing of start-up and resolution of overlapping speech were included opposed to [21] and the present study that focused on timing of start-up only.

The studies of Ter Maat and the present study share a similarity in approach: using simulated audio-only conversations, systematically varying turn-taking and measuring perception for each conversation by semantic differential scales. However, there were some differences between the present study and each of the studies by Ter Maat. The first study [20] included unrecognizable speech created from human speech; recordings of human speech were edited in a manner that content was lost while prosody was maintained. The present study used synthesized speech, included content and excluded prosody. In the second study [21] content was included and synthesized speech was used. This study differed from the present study by the role of the agent and the participant. The participant was actively involved in the conversation with an agent, which was controlled by a WOz timing the start of speech based on human-predicted turn ending and using a specific turn-taking strategy. The agent had the role of interviewer asking open-questions in a casual topic. In the present study participants were outside observers. In [20, 21] participants rated an agent with either an undefined role or the role of interviewer. In contrast, in the present study

we were interested in the impression people get from a suspect interviewee.

The first study [20] found some differences in perception between turn-taking variations, indicating that people were able to provide a rating of impression based on timing of speech and prosody. In [21] the authors acknowledged that similar turn-taking behaviour can relate to different internal states and that turn-taking functions as a cue for what is happening in the conversation. This highlights a limitation of [20]; providing suggestions on the influence of turn-taking in conversations with no content. However, in real-life and applications including ECAs there is content. We believe that content influences the perception people get from an agent and argue that it is impossible to develop a model for an ECA that acts appropriately in different situations based on results excluding contextual factors such as content. Moreover, we argue that the institutional police interview setting moderates the perception of turn-taking behaviour, making it undesirable to exclude content.

8.4.3 Conclusion

The stimuli lacked prosody and non-verbal behaviour, diverging too much from real-life situations rendering it next to impossible to provide a rating for impression of the interlocutors. Confusion by the test situation influenced the ratings for perception, thereby rendering the effect of turn-taking unclear.

Better text to speech software might increase hearing and understanding. However, people do have different expectations from synthesized speech which influences their perception of an agent. The present study is part of a project where we develop an ECA for a serious game. If the envisioned game includes synthesized speech, a study with the same speech is recommended because the results obtained from studies with human voices would be non-representative for synthesized speech. Including prosody should be considered.

The present approach of isolation of turn-taking seems not appropriate to measure the influence of turn-taking on perception of interlocutors. More information such as prosody and non-verbal information is needed in order to be able to form an impression of an interlocutor. Human behaviour, including turn-taking, is highly complex. Behaviour is influenced by many factors such as emotion, personality and strategic choices and moderated by context. These factors seem too much interrelated to isolate any of them and present a clear answer to *the* influence of each single factor on specific behaviour. Considering the goal of the project—to support the development of an ECA capable of appropriate turn-taking behaviour—it is not required to know the influence of every individual factor. Instead we should focus on unravelling the combinations of factors that define a situation and appropriate behaviour.

Chapter 9

Conclusion

To support the development of a computational model for turn-taking behaviour of a virtual suspect agent we investigated whether variations in turn-taking behaviour lead to differences in the perception of the suspect. The research question was: *What influence do variations in turn-taking behaviour have on the perception of power, affiliation, rapport, face and deception of a virtual suspect?*

Previous studies looked mainly at interaction from the perspective of the police officers. Much research has been done on the strategic choices of officers to improve the general outcome of interviews and ‘right to silence’ has been subject of many studies. A literature study provided us with some suggestions on the relation between turn-taking and affect in police interviews. Interruptions were associated with power, minimal responses were considered a signal of low rapport and rapport improved cooperation and agreeability. The suggestions were general and the effect of the institutional setting and interlocutors’ roles remained unclear. The existence of an asymmetrical relationship between a police officer and a suspect interviewee was widely acknowledged; by their role of interviewer, the police officers have pre-specified advantage in power and control.

In a perception study we aimed to unravel the effect of turn-taking behaviour on the perception of a suspect. To measure this effect we isolated turn-taking behaviour as much as possible: non-verbal behaviour and prosodic features were excluded from the stimuli. We measured the perception scores for stimuli with: a) gaps between turn-taking, b) bridged turn-taking and c) overlapping speech. There were some individual differences between perception scores for the three variations. Agreement in differences were hard to discover due to individual unsystematic variation. If any, we can conclude that slower turn-taking was associated with a perception of higher affiliation, face and rapport and lower deception for the suspect; faster turn-taking was associated with higher power for the suspect.

In an evaluation of the approach we learned that people find it difficult to provide a perception score based on a short sound fragment of synthesized speech. The information is too scarce to get an impression of the interlocu-

tors and disfluency in synthesized speech had a comical effect. In some cases behavioural changes were noticed and influenced the perception scores. However, when people change their basis of rating between content and behaviour comparison of perception scores is crooked.

In the intended game environment turn-taking, non-verbal behaviour, prosody, content and contextual factors will be present and undoubtedly all contribute to the impression a person gets from the virtual suspect. The extent to which turn-taking contributes to this impression is unknown. It might even be the case that the influence of turn-taking behaviour is overshadowed by other factors. It seems impossible to isolate turn-taking behaviour and obtain a valid measurement of perception. The test situation not only caused difficulties in evoking an impression; moreover, it influenced the ratings for perception rendering results non-representative.

Considering the goal of the study —to support the development of a model for an ECA capable of appropriate turn-taking behaviour— we argue that instead of searching for *the* fundamental truth on why people execute certain turn-taking behaviour it might be better to create a model and evaluate this model by validating if certain desired personality traits can be communicated through the ECA’s behaviour in a certain context. To discover how a person perceives an agent in a specific context, it is important to have the test situation closely resembling the intended context. We still believe turn-taking influences the perception people get from an agent and reckon observable human behaviour a good starting point to create a model, but it is next to impossible and unnecessary to unravel the influence of one behavioural factor on the perception of an interlocutor.

Bibliography

- [1] A. Abbe and S. E. Brandon. The Role of Rapport in Investigative Interviewing: A Review. *J. Investig. Psych. Offender Profil.*, 10:237–249, 2013.
- [2] A. Amelsvoort, I. Rispens, and H. Grolman. *Handleiding verhoor*. Stapel & De Koning, 2010.
- [3] K. Benneworth. Police interviews with suspected paedophiles: a discourse analysis. *Discourse & Society*, 20(5):555–569, Sept. 2009.
- [4] M. Ephratt. The functions of silence. *Journal of Pragmatics*, 40(11):1909–1938, Nov. 2008.
- [5] K. Fogarty, M. Augoustinos, and L. Kettler. Re-thinking rapport through the lens of progressivity in investigative interviews into child sexual abuse. *Discourse Studies*, 15(4):395–420, Aug. 2013.
- [6] K. Haworth. The dynamics of power and resistance in police interview discourse. *Discourse & Society*, 17(6):739–759, Nov. 2006.
- [7] M. Heldner and J. Edlund. Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38(4):555–568, Oct. 2010.
- [8] G. Heydon. Silence: Civil right or social privilege? A discourse analytic response to a legal problem. *Journal of Pragmatics*, 43(9):2308–2316, July 2011.
- [9] C. Jones. Uk police interviews: a linguistic analysis of afro-caribbean and white british suspect interviews. *International Journal of Speech Language and the Law*, 15(2):271–274, 2008.
- [10] M. Komter. The Interactional Dynamics of Eliciting a Confession in a Dutch Police Interrogation. *Research on Language and Social Interaction*, 36(4):433–470, 2003.
- [11] D. Kurzon. The right of silence: A socio-pragmatic model of interpretation. *Journal of Pragmatics*, 23:55–69, 1995.
- [12] H. Mazeland. *Inleiding in de conversatieanalyse*. Coutinho, 2003.

- [13] N. Momeni. Police Genre: Interruption and its Classification as a Sign of Asymmetry in Police Interview/Interrogation. *International Journal of Criminology and Sociological Theory*, 4(1):615–625, 2011.
- [14] R. op den Akker, M. Bruijnes, R. Peters, and T. Krikke. Interpersonal stance in police interviews: content analysis. *Computational Linguistics in the Netherlands*, 3:193–216, 2013.
- [15] A. C. Ostermann. Communities of practice at work: gender, facework and the power of habitus at an all-female police station and a feminist crisis intervention center in brazil. *Discourse & Society*, 14(4):473–505, 2003.
- [16] R. Peters. Turn-taking in police interviews. Unpublished Report, available at <http://www.rifca.nl/downloads/ut/rttt-rmp.pdf>, 2013.
- [17] H. Sacks, E. A. Schegloff, and G. Jefferson. A simplest systematics for the organization of turn-taking for conversation. *Language*, pages 696–735, 1974.
- [18] E. A. Schegloff. From interview to confrontation: Observations of the bush/rather encounter. *Research on Language and Social Interaction*, 22:215–240, 1988.
- [19] E. A. Schegloff. Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, 29(01):1–63, Jan. 2000.
- [20] M. ter Maat and D. Heylen. Turn management or impression management? In *Intelligent Virtual Agents*, pages 467–473. Springer, 2009.
- [21] M. Ter Maat, K. P. Truong, and D. Heylen. How turn-taking strategies influence users’ impressions of an agent. In *Intelligent Virtual Agents*, pages 441–453. Springer, 2010.
- [22] K. R. Thórisson, O. Gislason, G. R. Jonsdottir, and H. T. Thorisson. A multiparty multimodal architecture for realtime turntaking. In *Intelligent Virtual Agents*, pages 350–356. Springer, 2010.
- [23] L. Tickle-Degnen and R. Rosenthal. The Nature of Rapport and Its Non-verbal Correlates. *Psychological Inquiry: An International Journal for the Advancement of Psychological Theory*, 1(4):285–293, 1990.
- [24] J. Verschueren. *What people say they do with words: Prolegomena to an empirical-conceptual approach to linguistic action*, volume 14. Ablex Publishing Corporation Norwood, New Jersey, 1985.
- [25] A. Vrij, S. Jundi, L. Hope, J. Hillman, E. Gahr, S. Leal, L. Warmelink, S. Mann, Z. Vernham, and P. A. Granhag. Collective interviewing of suspects. *Journal of Applied Research in Memory and Cognition*, 1:41–44, Mar. 2012.

- [26] D. Walsh and R. Bull. Examining Rapport in Investigative Interviews with Suspects: Does its Building and Maintenance Work? *Journal of Police and Criminal Psychology*, 27(1):73–84, Feb. 2011.
- [27] D. Yoong. Interactional norms in the Australian police interrogation room. *Discourse & Society*, 21(6):692–713, Oct. 2010.

Appendix A

Survey

After a short introduction and example sound file and scales the perception study started. For each stimulus the impression a participant had of the interlocutors was measured by a total of 25 7-point Likert scales; 3 statements for the suspect, 16 bipolar adjective pairs for the suspect and 6 bipolar adjective pairs for the police officer. The appropriate gender was included based on the gender of the suspect and police officer in the stimulus. A more positive association with a factor was measured by a higher rating, counterbalanced scales are depicted in italics. After rating four stimuli some demographic data was collected. All survey questions are in Dutch and depicted below.

Geef aan in welke mate u het eens bent (7-punts schaal volledig mee eens - volledig mee oneens) met de volgende stellingen:

1. De vrouwelijke/mannelijke spreker heeft controle over het gesprek.
volledig mee eens - - - - - volledig mee oneens
2. De vrouwelijke/mannelijke spreker heeft een verdedigende houding.
volledig mee eens - - - - - volledig mee oneens
3. De vrouwelijke/mannelijke spreker probeert het gespreksonderwerp te vermijden.
volledig mee eens - - - - - volledig mee oneens

Geef aan in welke mate (tegenstellingen 7-punts schaal) u vindt dat onderstaande eigenschappen de vrouwelijke/mannelijke spreker beschrijven:

4. vriendelijk - - - - - vijandig
5. *tegenwerkend* - - - - - *samenwerkend*
6. coöperatief - - - - - competitief
7. *onderdanig* - - - - - *dominant*

8. beleefd - - - - - ongemanierd
9. *afhankelijk* - - - - - *onafhankelijk*
10. *afkeurend* - - - - - *goedkeurend*
11. positief - - - - - negatief
12. aandachtig - - - - - onaandachtig
13. *afstandelijk* - - - - - *betrokken*
14. competent - - - - - incompetent
15. respectvol - - - - - onbeschoft
16. *eerlijk* - - - - - *bedrieglijk*
17. afwijzend - - - - - acceptierend
18. weerstandig - - - - - meewerkend
19. *onachtzaam* - - - - - *verantwoordelijk*

Geef aan in welke mate (tegenstellingen 7-punts schaal) u vindt dat onderstaande eigenschappen de mannelijke/vrouwelijke spreker beschrijven:

20. vriendelijk - - - - - vijandig
21. coöperatief - - - - - competitief
22. *onderdanig* - - - - - *dominant*
23. beleefd - - - - - ongemanierd
24. competent - - - - - incompetent
25. *afkeurend* - - - - - *goedkeurend*

Survey impressie van interpersoonl...

student.survey.rifca.nl

Google

Impressie van interpersoonlijke relaties in politieverhoren

Luistert u goed naar het eerste audiofragment en beantwoord daarna de vragen.

Het gaat om uw indruk van de sprekers, er is geen goed of fout antwoord.

In het volgende fragment van een sociaal politieverhoor hoort u een vrouw die verdacht wordt van mishandeling van haar buurvrouw naar aanleiding van hondenpoep. Voorafgaand aan het fragment hebben de verdachte en de agent een informeel gesprek gehad over huisdieren (hun gezamenlijkheid in leven met katten). In het fragment spreken de verdachte en de agent over de bezigheden van de verdachte op de dag waarop de vermeende mishandeling plaats heeft gevonden.

0:53

Geef aan in welke mate u het eens bent (schaal 1-7 volledig mee eens - volledig mee oneens) met de volgende stellingen:

	Eens	Oneens
De VROUWELIJKE spreker probeert het gespreksonderwerp te vermijden.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
De VROUWELIJKE spreker heeft een verdedigende houding.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>
De VROUWELIJKE spreker heeft controle over het gesprek.	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>

Geef aan in welke mate (tegenstellingen 7-punts schaal) u vindt dat onderstaande eigenschappen de **VROUWELIJKE** spreker beschrijven.

Vriendelijk	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Vijandig
Tegenwerkend	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Samenwerkend
Coöperatief	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Competitief
Onderdanig	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Dominant
Beleefd	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Ongemanierd
Afhankelijk	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Onafhankelijk
Afkeurend	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Goedkeurend
Positief	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Negatief
Aandachtig	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Onaandachtig
Afstandelijk	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Betrokken
Competent	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Incompetent
Respectvol	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Onbeschoft
Eerlijk	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Bedriegelijk
Afwijzend	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Accepterend
Weerstandig	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Meerwerkend
Onachtzaam	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Verantwoordelijk

Geef aan in welke mate (tegenstellingen 7-punts schaal) u vindt dat onderstaande eigenschappen de **MANNELIJKE** spreker beschrijven.

Vriendelijk	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Vijandig
Coöperatief	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Competitief
Onderdanig	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Dominant
Competent	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Incompetent
Beleefd	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Ongemanierd
Afkeurend	<input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/> <input type="radio"/>	Goedkeurend

Volgende →

Figure A.1: screenshot of the online survey tool depicting the audio player and the scales measuring perception of a (female) suspect and (male) police officer.

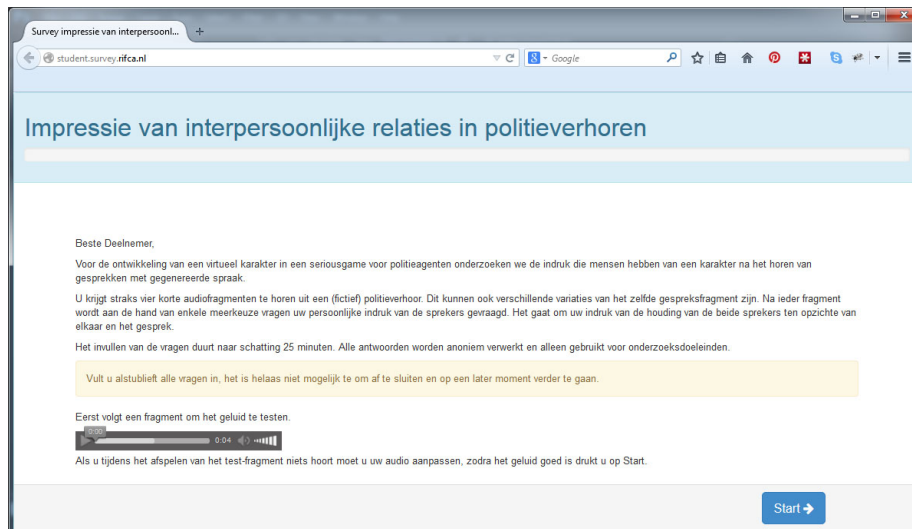


Figure A.2: screenshot of the introduction to the survey.

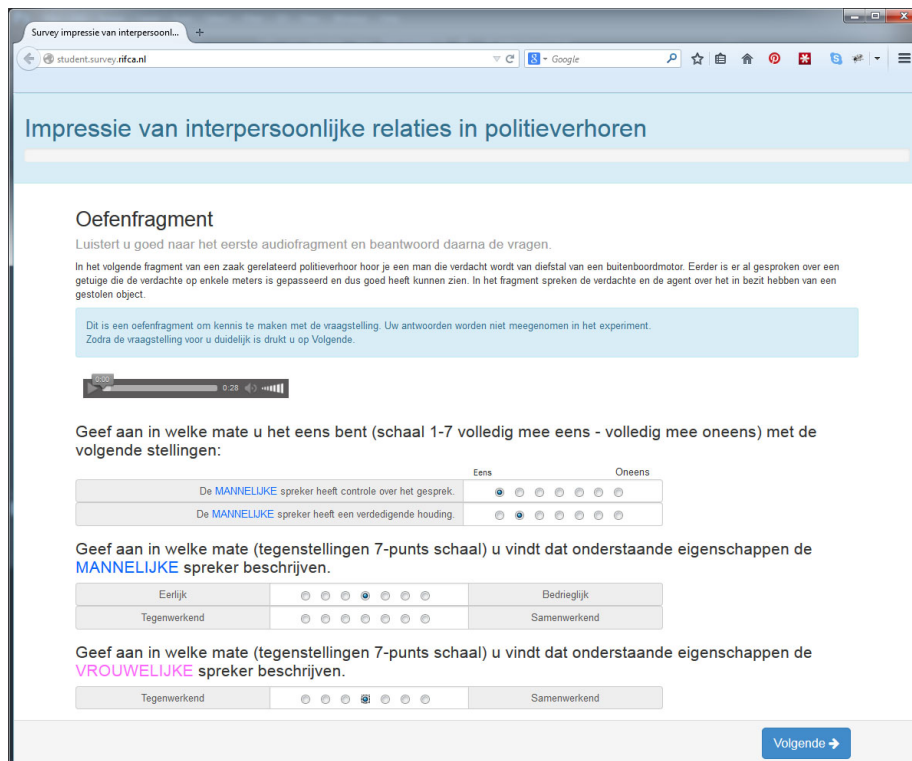


Figure A.3: screenshot of the example sound file and scales.

Survey impressie van interpersoonl...

student.survey.rifca.nl

Google

Impressie van interpersoonlijke relaties in politieverhoren

Tot slot volgen er enkele vragen over uzelf.

Wat is uw leeftijd?

- ☐ 24 jaar of jonger.
- ☐ Tussen de 25 en 34 jaar.
- ☐ Tussen de 35 en 44 jaar.
- ☐ Tussen de 45 en 54 jaar.
- ☐ 55 jaar of ouder.

Wat is uw geslacht?

- ☐ Mannelijk
- ☐ Vrouwelijk

Hebt u ooit te maken gehad met politieverhoren?

- ☐ Nee
- ☒ Ja, als...
 - ☒ slachtoffer ☐ getuige ☐ verdachte ☐ in studie of beroep
 - ☐ overig, namelijk:

Vul uw ervaring met politieverhoor in

Bent u op dit moment student?

- ☐ Ja
- ☐ Nee

Wat is uw huidige of hoogst genoten opleiding?

Vul de naam van uw huidige of hoogst genoten opleiding in

Vul de naam in van de opleiding die u nu volgt of heeft gevolgd.

Hebt u eerder deelgenomen aan een soortgelijk onderzoek?

☐ Ja, ik heb eerder deelgenomen aan een onderzoek naar perceptie van een verdachte in politieverhoor.

Afsluiten ✕

Figure A.4: screenshot of the demographic questions.

Appendix B

Stimuli

The stimuli were generated conversations based on extracts of the DPIT-corpus [14]. Transcriptions of the original extracts (in Dutch) are depicted below. The extracts are transcribed according to the Gail Jefferson convention [12]. Sentences and words present in the original extract that were excluded in the stimuli are depicted in grey, additions are depicted in italics.

Extract E1 is part of a social interviewing phase; the focus of the officer is on getting to know the suspect. The male interlocutor is suspected of involvement in theft and incendiarism. Preceding the extract the suspect was frustrated and the police officer was focussed on making the suspect feel at ease to improve cooperation of the suspect. In the extract the interlocutors discuss the topic of the suspect's occupation. The suspect provides syntactical correct but not necessarily satisfactory responses.

Extract E2 was selected from the same social interview. The male interlocutor is suspected of involvement in theft and incendiarism. Preceding the extract the suspect notices the presence of cameras. The suspect is upset because he has not been informed about the audio-visual registration of the interview. In the extract the interlocutors discuss the impact of the audio-visual registration. The suspect attempts to gain control over the conversation by interrupting the officer.

Extract E3 is part of a different social interview in which the female is suspected of battery against her neighbour. Preceding the extract the interlocutors have had an informal conversation on the topic of their joint experience with cats. In the extract the interlocutors discuss the topic of the suspect's pursuits on the day of the alleged battery. The suspect is providing syntactically correct responses to probing questions.

Extract E4 is selected from a case related interview; the focus of the police officer is on gathering information about the suspect's involvement in the case. The female interlocutor is suspected of money theft from a gas station kiosk. Preceding the extract it became evident that despite earlier statements by the suspect the suspect did visit the kiosk and had the opportunity to take the money. In the extract the interlocutors discuss the suspect's pursuits in the

kiosk. The suspect is waiting to be addressed by a completed question before providing a syntactical correct response.

696 PB2 wat ~~zo~~-*zou* wat voor vak zou jij ambiëren,
(1.68)
698 SB1 ambities,
(0.44)
700 PB2 °nou°
(0.27)
702 SB1 n:-*nou* iik *ik* ben gewoon stoffeerder
(0.5)
704 PB2 oke
(1.45)
706 PB2 hoe lang doe je dat al
(0.66)
708 SB1 ja hoe lang doe ik het al
(1.43)
710 jaar of
(0.22)
712 zes zeven of zo
(0.28)
714 PB2 oke
(1.28)
716 SB1 nah maar d- *dan* maar dan wel ook echt echt gewoon
stoffeerder van van van mooie meubeltjes he
[omitted 6 lines to keep turn-taking consistent]
728 PB2 maar hoe ben je daar zo ingerold want dat is wel lastig
specifiek vak
(0.52)
730 SB1 ja klopt
(1.58)
732 PB2 (maar hoe) ben je daar zo ingerold
(.)
734 SB1 via een oud kennis van me

Extract E1: vanBron1_a — 10:18.938

1609 SB1 JA EN MIJ MIJN ZIELELEVEN HIER BLOOT ZITTEN LEG- GE
 LEGGEN GEVEN ZITTEN ER PSYCHIATERS BIJ=
 1610 PB1 =je hebt je hebt nog helemaal niks[verteld]
 1611 SB1 [ZITTEN] ER PSYCHIATERS
 BIJ[OF WAT]
 1612 PB1 [er zit]
 (.)
 1614 er zitten=
 1616 SB1 =GEEF ANTWOORD OP MIJN VRAAG GOZER *NOU*
 (.)
 1618 PB1 er zit-ten
 1619 SB1 [GEEF] ANTWOORD OP MIJN VRAAG=
 1620 PB1 =blijf
 (.)
 1622 SB1 GEEF ANTWOORD OP MIJN VR^{AA}
 1623 PB1 [(ga zitten)]^G=
 1625 =ga nou zitten-nou
 1626 SB1 [ZIT]TEN=
 1628 =zitten hier psychiaters en psychologen bij
 (.)
 1630 PB1 psycholoog zit er bij ja=
 1631 SB1 =ja=
 1632 PB1 =ja
 (0.64)
 1634 SB1 oke
 (0.75)
 [omitted 4 lines to prevent repetition]
 1644 PB1 .hh=
 1645 =je zegt je zegt ik heb heel mijn zielenleven
 1646 SB1 [(heel goed)]
 1645 PB1 bloot gelegd[maar je] hebt nog helemaal niks verteld
 1647 SB1 [jah
 (.)
 1649 SB1 ik
 (.)
 1651 ik heb niks verteld,=
 1652 PB1 =nee::=
 1653 SB1 =dan moet je die bana-nen uit je oren halen gozer
 1654 PB1 [(je hebt niks)
 (.)
 1656 banaan uit mijn oren?=
 1657 SB1 =ja dan moet je de bananen uit je oren halen want dan
 hoor je dat ik heel veel zeg=
 1658 PB1 =nou wat heb je verteld dan=
 1659 =ik ik kan me het niet herinneren hoor
 1660 SB1 [wat heb je allemaal verteld dan]

411 PW1 jou werk, werk jij trouwens?=
 412 SW1 =nee ik zit nog op school
 [omitted 3 lines]
 (.)
 418 PW1 oké
 (1.18)
 420 wat voor opleiding doe je?
 (.)
 422 SW1 kappersopleiding
 [omitted 3 lines]
 (1.33)
 430 PW1 en waar doe je dat?
 (0.77)
 432 SW1 in eh Zwolle
 (.)
 434 PW1 °in Zwolle°
 (0.47)
 436 °oke°
 (2.29)
 438 en gister waar was je gister,
 (0.34)
 440 SW1 nou heb ik eerst les gehad
 (.)
 442 maar dat was alleen sochtends
 (0.61)
 444 PW1 tot hoe laat?
 (.)
 446 SW1 tot e::h twaalf uur had ik les
 (0.88)
 448 PW1 hmmm en wat heb je toen gedaan?
 (0.48)
 450 SW1 en toen ben ik nog heel even daar de stad in geweest met
 een klasgenootje van mij
 (0.21)
 452 PW1 oke
 [omitted 4 lines]
 (.)
 462 SW1 en toen ben ik naar huis gegaan
 (0.72)
 464 PW1 hoe heet die eh kla- dat klasgenootje van jou,
 (.)
 466 SW1 Tamara
 (0.34)
 470 PW1 jij zegt daarna ben ik naar huis gegaan
 (0.31)
 472 samen met Tamara,
 (0.31)
 474 PW1 of alleen,
 475 SW1 nee alleen

1562 PH3 en eh waar zit die telefoon die zit daar dan
(0.99)

1564 zi- zit is dat echt achter de balie of
(0.46)

1566 SH1 neu nee meer naar de zijkant zeg maar
(.)

1568 PH3 oke
(0.62)

1570 een beetje naast ehh
(0.74)

1572 de balie
(1.05)

1574 SH1 jah
[omitted 1 line containing officer falter]
(3.03)

1578 PH3 ja hoe hoe hoe zag dat daar binnen uit wat heb je
daar gezien in die eh
(1.22)

1580 in die kiosk
(1.89)

1582 SH1 u::hm
(0.29)

1586 er stond een stelling met snoep
(1.13)
[omitted 7 lines]

1602 daar voor stonden eh stonden eh de (de bloemen)=
[omitted 16 lines]

1634 PH3 =ben je nog ergens anders geweest in die kiosk
(0.8)

1636 SH1 nee
(2.11)
[omitted 9 lines]

1656 PH3 heb je daar nog wat meegenomen
(2.15)

1658 SH1 nee
(2.54)

1660 snoep ofzo eh
(1.93)

1662 PH3 °dat vraag ik° aan jou
(0.27)

1664 SH1 nee

Extract E4: Huls3 — 29:03.210

Appendix C

Results

The raw data obtained from the perception study is depicted below. Each table presents the data for one extract. For extract E1, E2 and E3 all three variations (gap, bridged and overlap) were rated by the same participants. The variations of E4 were divided among the three groups, participants in group A rated E4_a (gap), group B E4_b (bridged) and group C E4_c (overlap).

Table C.1: perception scores for E1 by group A, scores are depicted for each participant ordered by turn-taking pattern: a) gap, b) bridged and c) overlap.

Participant	Stimulus	Statements			Suspect Adjectives																Police Adjectives						
		Evasion	Defensiveness	In Control	Friendly-Hostile	Together-Opposed	Cooperative-Competitive	Dominant-Submissive	Polite-Impolite	Autonomous-Dependent	Approving-Disapproving	Positive-Negative	Attentive-Unattentive	Close-Distant	Competent-Incompetent	Respectful-Disrespectful	Deceptive-Honest	Rejecting-Accepting	Resistant-Cooperative	Careless-Responsible	Together-Opposed	Cooperative-Competitive	Dominant-Submissive	Competent-Incompetent	Polite-Impolite	Approving-Disapproving	
PA1	E1 _a	1	1	1	6	6	6	2	6	3	4	5	5	6	5	5	1	2	1	4	7	7	5	5	6	6	
	E1 _b	1	1	1	6	6	6	3	6	3	5	6	5	5	5	5	1	2	2	4	6	7	6	5	6	6	
	E1 _c	1	1	1	6	7	6	3	6	3	6	6	5	5	6	5	6	1	2	2	4	6	7	6	5	6	6
PA2	E1 _a	6	6	6	2	6	2	6	3	4	2	2	5	2	4	4	4	6	6	4	3	3	5	5	3	4	3
	E1 _b	6	6	5	2	2	2	6	3	6	2	2	5	2	4	4	4	5	4	4	2	3	5	3	4	3	
	E1 _c	6	6	6	2	2	2	6	4	5	2	3	4	2	4	3	4	6	6	3	3	3	5	5	3	4	3
EA1	E1 _a	2	2	4	5	6	5	4	5	4	5	6	5	4	5	5	3	2	3	5	5	5	4	5	5	6	
	E1 _b	3	3	5	5	4	5	4	5	4	5	5	4	4	5	5	3	3	4	4	5	5	4	5	5	5	
	E1 _c	4	4	3	3	3	4	4	3	4	4	5	5	3	5	4	4	4	5	3	5	5	3	5	4	5	
EA2	E1 _a	1	1	4	7	7	6	4	6	4	4	6	6	5	6	6	2	2	2	4	5	3	6	6	5	6	
	E1 _b	1	1	4	7	7	4	3	6	4	5	5	5	5	4	6	2	2	2	5	3	4	6	4	6	6	
	E1 _c	1	1	4	5	5	5	3	6	3	3	5	6	6	5	7	1	1	2	4	6	6	7	6	6	6	

Table C.2: perception scores for E2 (group B), scores are depicted for each participant ordered by turn-taking pattern: a) gap, b) bridged and c) overlap.

Participant	Stimulus	Statements			Suspect Adjectives															Police Adjectives						
		Evasion	Defensiveness	In Control	Friendly-Hostile	Together-Opposed	Cooperative-Competitive	Dominant-Submissive	Polite-Impolite	Autonomous-Dependent	Approving-Disapproving	Positive-Negative	Attentive-Unattentive	Close-Distant	Competent-Incompetent	Respectful-Disrespectful	Deceptive-Honest	Rejecting-Accepting	Resistant-Cooperative	Careless-Responsible	Together-Opposed	Cooperative-Competitive	Dominant-Submissive	Competent-Incompetent	Polite-Impolite	Approving-Disapproving
EB1	E2 _a	5	5	5	1	4	1	6	3	5	4	1	4	7	2	1	7	1	2	5	1	7	7	7	1	4
	E2 _b	3	2	3	1	5	1	7	1	4	5	1	3	7	4	1	7	1	5	5	1	1	7	4	1	3
	E2 _c	3	4	4	1	4	1	7	1	5	5	1	5	7	5	1	7	1	3	4	1	1	7	3	5	3
EB2	E2 _a	5	5	3	3	2	3	5	3	4	3	3	3	3	4	3	5	6	6	3	3	3	4	3	3	3
	E2 _b	5	6	3	3	2	3	5	3	4	3	3	3	3	3	3	5	6	5	3	3	3	5	4	3	3
	E2 _c	5	5	3	3	3	3	5	3	5	3	2	3	3	3	3	5	5	5	4	2	3	4	3	3	3

Table C.3: perception scores for E3 (group C), scores are depicted for each participant ordered by turn-taking pattern: a) gap, b) bridged and c) overlap.

Participant	Stimulus	Statements			Suspect Adjectives															Police Adjectives						
		Evasion	Defensiveness	In Control	Friendly-Hostile	Together-Opposed	Cooperative-Competitive	Dominant-Submissive	Polite-Impolite	Autonomous-Dependent	Approving-Disapproving	Positive-Negative	Attentive-Unattentive	Close-Distant	Competent-Incompetent	Respectful-Disrespectful	Deceptive-Honest	Rejecting-Accepting	Resistant-Cooperative	Careless-Responsible	Together-Opposed	Cooperative-Competitive	Dominant-Submissive	Competent-Incompetent	Polite-Impolite	Approving-Disapproving
PC1	E3 _a	2	2	0	2	2	2	4	4	3	3	3	3	3	0	4	3	3	3	3	3	2	4	0	0	0
	E3 _b	5	5	0	5	5	5	5	4	4	5	4	3	6	0	4	4	5	5	4	5	5	5	0	0	0
	E3 _c	2	6	0	4	3	5	6	3	3	4	4	2	6	0	4	3	3	3	4	3	3	4	0	0	0
PC2	E3 _a	1	1	0	1	1	1	2	1	5	4	2	2	2	0	1	3	2	1	2	1	1	4	0	0	0
	E3 _b	4	6	0	5	5	5	5	6	3	6	5	4	6	0	6	5	5	5	5	5	5	5	0	0	0
	E3 _c	3	5	0	5	4	5	5	5	3	5	5	5	4	0	5	4	5	6	4	2	5	5	0	0	0
EC1	E3 _a	2	1	1	4	4	7	4	7	4	4	6	6	4	4	6	2	4	1	7	7	7	4	4	7	7
	E3 _b	2	5	1	2	6	6	6	3	5	4	4	7	2	4	2	1	6	6	6	7	7	4	4	7	4
	E3 _c	2	7	1	2	4	4	7	2	4	2	2	2	2	4	2	4	7	7	4	7	7	2	2	7	6
EC2	E3 _a	1	1	2	7	7	7	4	7	4	4	4	6	5	4	6	1	2	2	4	6	6	4	5	6	4
	E3 _b	1	1	4	3	5	6	4	6	4	4	4	5	5	4	6	3	4	2	4	6	6	4	4	6	4
	E3 _c	2	6	5	2	3	4	6	4	4	4	3	4	3	4	2	6	4	6	4	6	6	4	4	6	4
EC3	E3 _a	1	7	3	7	7	7	1	7	2	4	4	7	4	7	7	1	2	2	6	7	7	4	7	7	6
	E3 _b	1	1	7	7	7	7	4	7	6	6	7	7	4	7	7	1	4	1	7	7	7	4	7	7	4
	E3 _c	7	7	7	1	4	2	6	2	6	4	2	4	2	7	2	6	4	4	4	7	7	2	3	7	7
EC4	E3 _a	3	3	3	5	4	4	4	4	5	3	4	4	4	4	4	3	5	4	4	5	4	5	3	4	4
	E3 _b	3	4	3	4	4	4	5	4	5	4	4	4	4	4	4	3	4	5	4	5	4	5	3	5	4
	E3 _c	3	4	4	3	4	4	5	3	5	3	5	4	4	4	3	3	5	5	3	5	4	5	3	4	3
EC5	E3 _a	2	2	3	5	6	6	3	5	3	4	4	6	5	5	5	2	2	2	6	6	6	4	5	6	4
	E3 _b	2	4	3	3	4	3	5	5	2	3	4	6	2	5	4	2	5	3	5	5	5	3	5	5	4
	E3 _c	2	6	4	3	3	3	2	6	3	5	2	2	5	2	5	3	4	6	6	5	6	6	3	5	6

Table C.4: perception scores for E4 (group A, B and C), scores are depicted ordered by turn-taking pattern/group: a) gap (group A), b) bridged (group B) and c) overlap (group C).

Participant	Stimulus	Statements			Suspect Adjectives															Police Adjectives						
		Evasion	Defensiveness	In Control	Friendly-Hostile	Together-Opposed	Cooperative-Competitive	Dominant-Submissive	Polite-Impolite	Autonomous-Dependent	Approving-Disapproving	Positive-Negative	Attentive-Unattentive	Close-Distant	Competent-Incompetent	Respectful-Disrespectful	Deceptive-Honest	Rejecting-Accepting	Resistant-Cooperative	Careless-Responsible	Together-Opposed	Cooperative-Competitive	Dominant-Submissive	Competent-Incompetent	Polite-Impolite	Approving-Disapproving
PA1	E4 _a	2	2	2	3	5	5	2	5	4	4	3	5	3	3	4	5	5	3	3	4	3	5	5	5	3
PA2	E4 _a	1	5	5	3	4	3	4	5	3	4	4	5	3	5	4	4	4	5	3	5	3	3	5	3	4
EA1	E4 _a	3	6	2	3	6	3	4	3	4	3	3	3	3	3	3	5	5	5	3	5	5	4	5	5	4
EA2	E4 _a	2	6	3	6	6	4	3	6	3	4	4	3	3	4	6	3	2	2	2	4	3	2	6	6	5
EB1	E4 _b	3	5	5	1	4	1	7	1	3	4	4	3	5	5	1	3	1	5	3	1	1	7	4	1	4
EB2	E4 _b	5	4	3	4	3	4	3	5	3	4	4	3	4	3	4	4	3	3	4	2	3	6	4	2	3
PC1	E4 _c	5	5	0	5	5	5	5	4	4	5	4	3	6	0	4	4	5	5	4	5	5	5	0	0	0
PC2	E4 _c	4	6	0	5	5	5	5	6	3	6	5	4	6	0	6	5	5	5	5	5	5	5	0	0	0
EC1	E4 _c	4	7	1	1	1	2	4	1	4	1	1	2	4	4	1	7	7	7	4	7	7	3	2	6	4
EC2	E4 _c	6	6	2	2	2	4	4	2	4	4	2	3	2	4	2	6	6	6	4	6	6	4	4	6	4
EC3	E4 _c	6	6	2	1	2	2	4	2	6	2	2	6	2	4	3	5	6	7	4	7	7	4	7	7	4
EC4	E4 _c	5	5	3	3	3	3	4	3	4	4	4	4	3	4	3	4	5	6	3	4	5	5	4	4	3
EC5	E4 _c	3	5	4	3	3	2	6	3	6	3	3	6	2	4	3	4	6	5	5	5	4	5	5	5	4

Appendix D

Chi Spar*s

On April 3th the perception study proposal and preliminary results from the pilot study were presented during the Chi Spar*s conference at The Hague University of Applied Sciences. The publication is available on the Chi Spar*s website (<http://chi-sparks.nl/2014/proceedings/>) and included below.



Figure D.1: presentation at the Chi Spar*s conference — photo by Patrick Deters (<https://www.flickr.com/photos/chi-nederland/13777355413/>).

How Turn-Taking Influences the Perception of a Suspect in Police Interviews

Rifca Peters

University of Twente
Enschede, The Netherlands
post@rifca.nl

Merijn Bruijnes

University of Twente
Enschede, The Netherlands
m.bruijnes@utwente.nl

Rieks op den Akker

University of Twente
Enschede, The Netherlands
h.j.a.opdenakker@utwente.nl

ABSTRACT

We study turn-taking behaviour in non-cooperative dialogue for the development of believable characters in a serious game for conversational skill learning in the police interview context. We describe a perception study to see how participants perceive a suspect's interpersonal stance, rapport, face, and deception when the turn-taking of the subject varies. We influence the perception of the suspect's stance by altering the timing of the start of speech with respect to the ending of the interlocutor's speech. The results of the study contribute to the development of an embodied conversational agent capable of natural human-system conversation with appropriate turn-taking behaviour.

Author Keywords

Embodied Conversational Agent; Turn-Taking; Serious game; Social skill training; Police interview; Believable virtual humans; Experimental perception study

ACM Classification Keywords

I.2.7 Natural Language Processing

INTRODUCTION

In human conversation we try to adhere to a “one-at-a-time” approach. Sacks, Schegloff and Jefferson [15] proposed a systematic, offering a set of rules to provide next-turn allocation to one interlocutor and thereby minimizing gap and overlap. However, moments of overlapping speech or silences occur frequently in human conversation [16]. These silences and moments of overlapping speech are often communicative in their own right [5,11,14]. Emotions and the stance people take towards each other influence turn-taking behaviour. “A clash of opinions also means a clash of turn-taking” [12]. Contrary to the dynamic turn-taking behaviour in human conversation, turn-taking behaviour in current natural dialogue systems is often restricted by a “one-at-a-time” rule. Conversational agents (CAs) are limited to listening or speaking and listening is initiated either on a place predetermined by the system or whenever the user makes a sound, resulting in an unnatural human-

system interaction. Exceptions are the dialogue systems that allow more free turn-taking behaviour [18].

In the context of the COMMIT P2 project we are working towards a computational model for human-like suspect turn-taking behaviour. This model supports the creation of a believable embodied conversational agent (ECA). This ECA will be used in a social skill training serious game for police officers that is currently under development. Rich CA turn-taking behaviour, including pauses, interrupts, and hesitation, is expected to support a more natural human-system interaction. A previous conversation analysis [2] showed that a factor such as the topic of conversation influences the interpersonal stance and the turn-taking behaviour of the suspect. Moreover, the stance of the suspect appeared to be related to the interpretation of suspect silences, e.g., a silent response from a suspect with a positive stance is interpreted as timidity while during a hostile stance it is related to withdrawal. Turn-taking strategies seem to have an effect on the perception of the agent [12,13].

The purpose of this study is to investigate how turn-taking behaviour influences the impression that observers get from a suspect in simulated police interviews. We look at the relation between turn-taking behaviour and perception of power, affiliation, rapport, face, and deception. We use extracts of police interviews in which we systematically vary turn-taking behaviour to study the influence of turn-taking on perception. This study focuses on the police interview setting. Police officers receive training on recognition and strategic use of interactional phenomena such as dominance [3]. Due to this experience, their perception of affective stance may be different from untrained people. The results of this work will inform the creation of a serious game that *police officers* will use to train their interview skills.

This paper is organized as follows. In the next section, we give a brief overview of relevant literature. Next, we present the Research Question and describe Methodology of the perception study. We conclude with a Discussion of the expectations of the results and the relevance of the results for the development of a conversational agent.

RELATED WORK

Literature on theoretical frameworks of and results from conversation analysis on turn-taking in police interviews provided us with some suggestions on which factors influence turn-taking behaviour in police interviews.

Yoong [24] showed that police officers interrupt suspects to prevent them from turn completion. These deliberate interruptions are considered signs of assertion of power

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

Copyright is held by the author(s).

Published in: van Leeuwen, JP, Stappers, PJ, Lamers, MH, Thissen, MJMR (Eds.) *Creating the Difference: Proceedings of the Chi Sparks 2014 Conference*, April 3, 2014, The Hague, The Netherlands.

[14,24]. Due to the asymmetric question/answer adjacency pairing, a police interview is structured to provide the officer with control over the conversation [4]. Haworth [7] claimed that power is under constant negotiation and reported recognition interrupts, minimal responses, taking extended turns, and interruptions of question as techniques used by suspects to access control in police interviews. Vrij [22] suggest that truth tellers adopt a “tell all” approach resulting in a talkative mood opposed to liars who adopt the “keep it simple” approach resulting in a less talkative mood. A more in-depth analysis of silence during stages of deception and truthfulness is given by Komter [10] who suggests that resistance by evasion or defence is a sign of deception and silences after a statement or question are associated with a non-contradicting position of the suspect. This absence of denial is often highlighted by an officer by allowing a long silence. To be considered relevant, denial should be provided immediately following or interrupt an accusation [9]. Rapport is considered a critical step in eliciting trust and building a relationship in professional interaction and therefore a prerequisite for techniques used in police interviews, e.g., to get cooperation from the interviewee [1,23]. Suspects tend to talk more openly in harmonious interactions and cooperation and agreement are increased. Discomfort –considered a lack of rapport– is displayed by stretches, fillers and pauses in the speech of the suspect [6]. In turn-taking we adhere to the terminology proposed by Heldner and Edlund [8], distinguishing two silences: gap and pause, two overlaps: between and within speaker, and bridged turn transitions: a smooth transition with no discernable silence (less than 0.18s) (see Figure 1). The type of question can influence the perception of an utterance. For example, a question directly addressing the suspect requires a response while this is not necessary for a statement. Also, an open-ended question is expected to be followed by an extensive response while yes or no are satisfactory responses for a closed question [17]. The type of question asked is related to the function of a question, e.g., information seeking for open-ended questions and conformation seeking for closed questions [20]. Moreover, case-related question may be more sensitive than small talk.

Ter Maat et al. [12,13] show that the manipulation of turn-taking strategies can lead to different perceptions of an agent on personality scales, interpersonal scales, and emotional scales. They conclude that these strategies can be used in the repertoire of expressive behaviours of agents reflecting these dimensions. We extend on this perception study. Based on the literature review, we hypothesize there is a relation between turn-taking behaviour and perception of power, affiliation, rapport, face, and deception.

RESEARCH QUESTION

To support the development of a computational model for turn-taking behaviour of a virtual suspect agent we evaluate the suggestions presented in the literature review: we assess if turn-taking behaviour is indeed related to the perception of interpersonal stance and investigate possible interaction between factors of interpersonal stance. The main research question is: *‘What influence do variations in turn-taking behaviour have on the perception of power, affiliation,*

rapport, face and deception of a virtual suspect?’. We formulated hypotheses following the same pattern for each of these factors: a turn-taking feature influences the factor. For deception (the other factors are omitted to conserve space):

In interactions with audible pause between sequential suspect turns, the suspect is perceived as more deceptive than in latched sequential turns.

In interactions with a gap between a question from an officer and the answer by a suspect, the suspect is perceived as more deceptive than in latched or overlapping question/answer adjacency pairs.

In interaction with a gap between a statement by an officer and a denial by a suspect, the suspect is perceived as more deceptive than in latched or overlapping denial.

METHODOLOGY

We selected extracts from our police interview corpus [2] and generated them with variance in timing of the start of speech with respect to the ending of the speech of the other interlocutor. These extracts are presented to participants who are asked to fill in a short survey on their perception on the personality, emotional state and interpersonal stance of the suspect after each extract. A pilot study is conducted to evaluate the stimuli and survey.

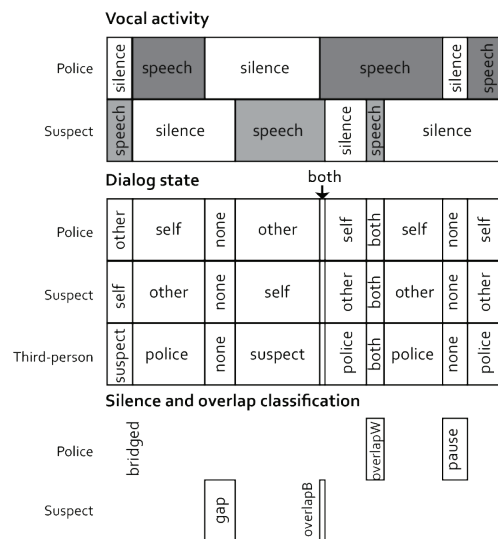


Figure 1: Top: Vocal activity of two speakers. Middle: The dialogue state shows who is speaking (depending on the perspective). Bottom: Classifications of the dialogue state: gap, pause, between-speaker overlap, within-speaker overlap, and bridged speaker transitions.

Participants

Police officers or police trainees are our participants (n=30) as their perception of affective stance may be different from untrained people due to their experience. Participants need to be native or proficient Dutch language users as all stimuli are in Dutch.

Stimuli

The stimuli, extracts from the corpus of Dutch police interview training videos [2], are generated using Ivona (ivona.com) text to speech. To maximize recognition of both speakers they are of opposing gender and the gender of the officer and the suspect are counterbalanced over all stimuli. All stimuli are generated using a single male and a single female voice. The extracts selected from the corpus demonstrate –or contradict– one interpersonal factor, see Figure 2. For each extract an altered version is created in which the turn-taking behaviour is adjusted while maintaining the content of the conversation as much as possible. Names are replaced by fictive names of similar length. Utterances are recorded and edited to vary the turn-taking using Audacity (audacity.sourceforge.net).

Design and Procedure

Participants are seated in front of a computer with loudspeakers. On the computer an online survey is presented. The participant is provided with information about the study and ensured confidentiality of their data. On each page the participant plays an audio file. Each file consists of an extract of a simulated conversation between an officer and a suspect. To distinguish between the officer and the suspect both interlocutors are of opposing gender.

To gather how a suspect is perceived, a survey is presented after each stimulus. The survey is the same for each stimulus –except for gender that is altered to comply with the gender of the speakers in the extract under assessment– and consists of opposing statements pairs to be rated on a 7 point semantic difference scale for: *dominance, friendliness, togetherness, cooperativeness, positivity, agreeability, attentiveness, politeness, respectfulness, autonomy, closeness, resistance, compliance* and *deceptiveness*. The chosen scales include the characteristics of interpersonal stance [3] and the factors of rapport [19]. Questions are counterbalanced for polarity where possible.

DISCUSSION

Previous research investigating police interviews included some aspects of silence or interruption and provided us with

```
410 P till what time?=
412 S =till eeoh twelve o'clock I had lessons
      (0.7)
414 P hmm hmm
      (0.4)
416 S ·Hhhh, then, eh, I went into the city for
      a bit with a classmate
      (.)
418 P ok
      (.)
420 S because eh
      (0.52)
422 yeah we also eh kind of eh needed things
      for in hair, I also do hairstyling, so we
      also needed things to put in hair and then
      I went home
```

Figure 2: Example of the transcription (translated from Dutch) of one stimulus for deception demonstrating the “tell all” approach. The officer asks a question (line 410) a response is immediately provided by the suspect (line 412), the suspect volunteers extended information and repeatedly self-selects (lines 416, 420-422).

suggestions on how personality, emotional state, and interpersonal stance influence turn-taking behaviour in a police interview setting [1,4,6,7,9,10,14,22,23,24]. However, these studies included turn-taking as one aspect within overall suspect behaviour and were not directed at the development of a model for turn-taking behaviour of a suspect. In this study we investigate if the factors influencing turn-taking according to the literature hold for a suspect in Dutch police interviews. We investigate whether variations in turn-taking behaviour lead to differences in the perception of the suspect. The first results will be presented at the Chi Sparks conference.

We expect the results of the study to contribute to the understanding of underlying factors influencing the (unconscious) choices a suspect makes if and when to speak. This understanding of underlying factors is needed to create an embodied conversational agent capable of mimicking human-like turn-taking behaviour which will support a more natural conversation between a human and an ECA. It can show its internal state by showing the appropriate turn-taking behaviour. For example, a virtual suspect in a dominant stance will display behaviour such as interrupts or when the agent has a deceptive stance it will take shorter turns and longer pauses in storytelling. See for an example of this type of agent [21]. The current study will try to determine what appropriate turn-taking behaviour is given the internal state of the agent that it tries to convey.

A potential limitation of our study is the usage of auditory-only stimuli. This removes the interference of non-verbal behaviour. However, non-verbal behaviour is undoubtedly important for the perception of an agent and will be available in the intended game environment. Studies on the perception of ECAs that incorporate verbal and non-verbal agent behaviour are required in the police domain. Also, all stimuli are short extracts (between 25 and 40 sec). However, longer extracts might be necessary for observers to form a consistent perception of the speakers. To the best of our knowledge, no research is done to investigate the relation between vocal stimuli length and perception agreement. In [2] we saw that inter-annotator agreement was low for short fragments, but we showed that global patterns become evident over longer periods.

By investigating the influence of turn-taking behaviour on the perception of a virtual suspect in police interview we aim to support the development of a virtual suspect for use in a social skill training serious game for police officers. By assessing the influence of turn-taking behaviour on the perception we gather knowledge about the extent of importance to model turn-taking behaviour and the appropriate behaviour given a desired interpersonal stance.

ACKNOWLEDGMENTS

This research is supported by the Dutch national program COMMIT.

REFERENCES

1. Abbe, A. and Brandon, S.E. The Role of Rapport in Investigative Interviewing: A Review. *J. Investig. Psych. Offender Profil.* 10, (2013), 237–249.
2. Akker, R. op den, Bruijnes, M., Peters, R., and Krikke, T. Interpersonal stance in police interviews: content analysis. *Computational Linguistics in the Netherlands* 3, (2013), 193–216.
3. Amelsvoort, A. van, Rispen, I., and Grolman, H. *Handleiding Verhoor*. Stapel & De Koning, 2012.
4. Benneworth, K. Police interviews with suspected paedophiles: a discourse analysis. *Discourse & Society* 20, 5 (2009), 555–569.
5. Ephratt, M. The functions of silence. *Journal of Pragmatics* 40, 11 (2008), 1909–1938.
6. Fogarty, K., Augoustinos, M., and Kettler, L. Re-thinking rapport through the lens of progressivity in investigative interviews into child sexual abuse. *Discourse Studies* 15, 4 (2013), 395–420.
7. Haworth, K. The dynamics of power and resistance in police interview discourse. *Discourse & Society* 17, 6 (2006), 739–759.
8. Heldner, M. and Edlund, J. Pauses, gaps and overlaps in conversations. *Journal of Phonetics* 38, 4 (2010), 555–568.
9. Heydon, G. Silence: Civil right or social privilege? A discourse analytic response to a legal problem. *Journal of Pragmatics* 43, 9 (2011), 2308–2316.
10. Komter, M. The Interactional Dynamics of Eliciting a Confession in a Dutch Police Interrogation. *Research on Language and Social Interaction* 36, 4 (2003), 433–470.
11. Kurzon, D. The right of silence: A socio-pragmatic model of interpretation. *Journal of Pragmatics* 23, (1995), 55–69.
12. Maat, M. ter and Heylen, D. Turn management or impression management? *Intelligent Virtual Agents*, (2009), 467–473.
13. Maat, M. Ter, Truong, K., and Heylen, D. How turn-taking strategies influence users' impressions of an agent. *Intelligent Virtual Agents*, (2010), 441–453.
14. Momeni, N. Police Genre: Interruption and its Classification as a Sign of Asymmetry in Police Interview/Interrogation. *International Journal of Criminology and Sociological Theory* 4, 1 (2011), 615–625.
15. Sacks, H., Schegloff, E. a., and Jefferson, G. A simplest systematics for the organization of turn-taking for conversation. *Language*, (1974), 696–734.
16. Schegloff, E.A. Overlapping talk and the organization of turn-taking for conversation. *Language in Society* 29, 01 (2000), 1–63.
17. Snook, B., Luther, K., Quinlan, H., and Milne, R. Let 'em Talk!: A Field Study of Police Questioning Practices of Suspects and Accused Persons. *Criminal Justice and Behavior* 39, 10 (2012), 1328–1339.
18. Thórisson, K.R., Gíslason, O., Jónsdóttir, G.R., and Thórisson, H.T. A Multiparty Multimodal Architecture for Realtime Turntaking. *Intelligent Virtual Agents*, (2010), 350–356.
19. Tickle-Degnen, L. and Rosenthal, R. The Nature of Rapport and Its Nonverbal Correlates. *Psychological Inquiry: An International Journal for the Advancement of Psychological Theory* 1, 4 (1990), 285–293.
20. Tracy, K. and Robles, J. Questions, questioning, and institutional practices: an introduction. *Discourse Studies* 11, 2 (2009), 131–152.
21. Traum, D. Non-Cooperative and Deceptive Virtual Agents. *IEEE Intelligent Systems*, (2012), 66–68.
22. Vrij, A., Jundi, S., Hope, L., et al. Collective interviewing of suspects. *Journal of Applied Research in Memory and Cognition* 1, (2012), 41–44.
23. Walsh, D. and Bull, R. Examining Rapport in Investigative Interviews with Suspects: Does its Building and Maintenance Work? *Journal of Police and Criminal Psychology* 27, 1 (2011), 73–84.
24. Yoong, D. Interactional norms in the Australian police interrogation room. *Discourse & Society* 21, 6 (2010), 692–713.