University of Stirling

Sonja Schaeffler

Department of Psychology

Are Affective Speakers Effective Speakers?

– Exploring the Link Between the Vocal Expression of Positive Emotions and Communicative Effectiveness

Thesis submitted in fulfilment of the requirements for the degree of PhD.

August 2006

This research has been funded through a Faculty Studentship by the University of Stirling and a Fellowship by the German Academic Exchange Service (DAAD).

Please note that the author's name has changed since the submission of this thesis, and that previous publications of the author (referred to in this thesis) have appeared under her maiden name *Sonja Biersack*.

Abstract

This thesis explores the effect of vocal affect expression on communicative effectiveness. Two studies examined whether positive speaker affect facilitates the encoding and decoding of the message, combining methods from Phonetics and Psychology.

The first study explored the relationship between perceived affective state and communicative effectiveness. To obtain a range of naturally occurring variation in affect, 200 speakers were asked to report their emotional state using the Brief Mood Introspection Scale after providing speech samples elicited in a Map Task paradigm in which speakers described a predefined route. In addition, one group of listeners (N=80) listened to a sentence with identical verbal content across speakers to rate how happy the speaker sounded; another group of listeners (N=600) was asked to draw the route on a map with just the landmarks while listening to the recorded descriptions. A measure of description quality in terms of linguistic content was obtained, as well a measure of listener comprehension in form of a route deviation estimate.

The results showed that perceived happiness, which was, interestingly, not correlated with self-reported happiness, contributed to listener comprehension over and above linguistic content. The happier speakers sounded the better listeners followed their descriptions, independently of verbal content. Perceived happiness was associated with faster speech rate, higher F_0 and F1, wider F_0 range and lower values for jitter. Path analyses showed that vocal correlates of positive affect influence communicative effectiveness in two ways: directly, by increasing perceptual salience (e.g., a raise in F1 can be associated with increased loudness), and indirectly, by signalling positive affect (a raise in F1 can be associated with a spreading of the lips, i.e. smiling), which, in turn, may improve rapport with the listener.

The aim of the second study was to test the impact of positive affect on communication in a highly affectively charged speech register, child-directed speech (CDS). The affective nature of CDS is attributable to the fact that interaction with their own baby serves as a powerful inducer of positive mood in mothers. A prosodic disambiguation task was administered to 24 mothers who where asked to address syntactically ambiguous sentences such as *"Touch the cat with the spoon"* to their 2 to 3.5 year old child and to an adult confederate trying to elicit an action complying either with the instrument or the modifier interpretation of the prepositional phrase. In addition, 24 non-mothers were tested addressing an imaginary toddler and an imaginary adult, thereby eliminating positive affect that might be elicited by real interaction with a small child. All speech samples were rated for perceived happiness by 20 independent listeners.

The findings showed that pitch increase and perceived happiness were more pronounced in CDS, and more so in mothers than in non-mothers. In contrast, only non-mothers, but not mothers, showed increased prosodic disambiguation in CDS, as indicated by changes in the duration of pauses after the first noun. This effect was corroborated by a forced choice test in which 48 listeners judged the intended meaning of each sentence. The results suggest that when expressing positive affect, speakers tend to emphasise affective prosody at the expense of linguistic prosody. In the case of CDS, this communication strategy may be more effective as it can serve to elicit attention in the child.

Taken together, Study 1 and 2 suggest that the effect of positive speaker affect on communicative effectiveness is primarily mediated by perceived affective prosody which may facilitate listener attention and improve rapport between interlocutors. In some situations, the para-linguistic features of affective prosody may enhance linguistic prosody. However, when placed in conflict with each other, as in Study 2, affective prosody appears to override linguistic prosody. In sum, the findings of this thesis lend support to the idea that affect expression impacts human communication independently of informational content and structure of the message.

Acknowledgements

A PhD is always a joint effort. Not just by you and your supervisor, but by you, your supervisor, your friends, family and colleagues. A lot of support is needed to see through a project like this and to keep up the motivation.

First of all thank you so much to my supervisor Vera Kempe who has not just been a great supervisor, but also an exceptionally gifted researcher, strategist, cook, chauffeur and friend! Many thanks to my second supervisor Professor Leslie Smith for some inspiring discussions; all the staff and postgrads who made me feel so welcome at the Department; particularly Bruce Sutherland, Bob Lavery, Shona MacLean and Stephen Stewart who helped me to get by in a Windows world; Kay Faichney for guiding me safely through the administrative jungle; everyone from the departmental basketball team for providing the much-needed comic relief; our Postgrad Tutor David Donaldson for teaching me the meaning of 'proactive'; Ranald Macdonald for indispensable help on Path Analysis, and most importantly, Lynn Williams and Rachael Fox who have been brilliant friends and a great support.

I also have to thank the helpful staff from the Glasgow Science Centre (GSC), who made it possible to recruit participants and run studies at this great place; all the RAs who helped running the studies, Hannah Blair, Adele Fern, Kathryn Mortimer, and especially Graham Samson and John Thoresen; Aileen Schmidt, Playgroup leader at the Department of Psychology for helping me recruit mothers and toddlers; and most importantly, the participants of my studies, both at Stirling University and at the GSC. Thanks also to colleagues from the speech science community who have given me feedback during those three year, at conferences and in form of reviews for submissions, particularly Sarah Hawkins and Randy Diehl for helpful comments on an earlier draft of Study 1.

Many thanks to my friends and family in Germany, Sweden and in Scotland, particularly my parents who have always supported me in my pursuit to find a career that I would really enjoy, Susen Rabold, who has been such a good friend and such great support; Klaus Härtl, my good friend and Latex specialist; Jan van Dorn for crucial advice on shimmer, jitter, formants, the principle of GIGO and the PhD God; my friends from the early days of studying Phonetics, the 'Trinker', and all the others at the Department of Phonetics, University of Munich, who have continued to give me advice and support me over all the years I have been abroad, particularly Phil Hoole, Florian Schiel, Christoph Draxler, Klaus Jänsch and Professor Tim Tillmann.

And finally, thank you so much to Felix Schaeffler, my partner in life and science. I will always be deeply grateful for your love, help, unfailing patience and continuous support.

Contents

List of Figures

List of Tables

Part I

Background

Chapter 1

Introduction

1.1 Aims and Contents of the Thesis

This thesis explores the impact of affect expression on the process of communication, i.e. the information conveyed in the message and the comprehension of the message by the listener. While previous research has mainly aimed at finding acoustic correlates for the expression and recognition of various emotional states, I focused on positive affect in the speaker and its influence on communicative effectiveness. By communicative effectiveness, I mean the extent to which the listener is able to comprehend the message in the way intended by the speaker.

The work is interdisciplinary in character and combines methods from Phonetics and Experimental Psychology. It seeks to employ phonetic methodology in ecologically valid studies that treat speech as behaviour and to explain the observable variation in it. In light of the different backgrounds this thesis might draw its readership from, readers are referred to Appendix E where a Glossary will give brief definitions of the psychological and phonetic terms used in this thesis.

The issues raised in this thesis should be of interest for researchers from speech sciences as well as psychology. However, speech and its immediate impact on the listener is not just relevant to researchers interested in speech, it is relevant to everyone who relies on efficient communication, intends to impart it, or model it for technical applications.

1.2 Chapter Layout

Chapter 1 outlines the aims and structure of the thesis and provides an overview over the contents and structure of each chapter.

Chapter 2 provides definitions for ‘emotion’, ‘affect’, ‘mood’ and ‘valence’ and outlines the components and functions of emotions. It then sets emotions in the context of psychological research and justifies why in this thesis only one emotional state, ‘happiness’ is examined.

Chapter 3 argues that emotions and communication are necessarily intertwined. This chapter is divided into three parts. Part 1 discusses the impact of positive affect on information processing as well as the link between positive affect and social cognition. Part 2 introduces the Brunswikian Lens Model adapted by Scherer for research on vocal affect expression and gives an overview of features that have been identified to be crucial for the production and perception of positive affect expression, respectively. Part 3 outlines how predictions are derived for this thesis. To this end, three possible models for the impact of vocal affect on communication are introduced.

Chapter 4 presents the first study. This study examines whether the vocal expression of positive affect enhances communicative effectiveness. Communicative effectiveness is determined by how well listeners perform in a Map Task. The degree of vocal affect is determined by listeners' affect ratings.

Chapter 5 presents the second study. Communicative effectiveness is examined for a specific verbal communication problem: the disambiguation of ambiguous sentences by prosodic means. Positive affect is elicited as part of a speech register that in particular accommodates the listener: child-directed speech of mothers addressing their children. An overview over research on child-directed speech is given, and the Prosodic Bootstrapping Theory is introduced as an important conceptualisation. The study is divided into two parts, contrasting speech of mothers with speech of non-mothers. The first part presents data of mothers talking to their own child and to an adult, the second part presents data of childless women talking to an imaginary child and an imaginary adult.

Chapter 6 summarises the results of the studies presented in this thesis and comprises a general discussion of the impact of vocal affect on speech communication. Theoretical implications are considered, and a number of issues for future research are identified.

Chapter 2

The Psychology of Emotions

2.1 Terminology

The terminology in emotion research is far from consistent (for an overview see Alpert & Rosen, 1990). This inconsistency can be accredited to, amongst other things, the fact that emotion-related concepts play a role in different disciplines where the focus of research and with that the definitions vary strongly (cf. Alpert & Rosen, 1990). At the same time, terms such as emotion, affect and mood seem to be so readily part of our daily language use that many authors only vaguely, if at all, define how they use those terms (see Russell, 2003, for a comprehensive discussion of this issue). A common way of distinguishing, for example, 'affect' and 'mood' is in terms of duration, with mood supposed to last longer than affect (cf. for example American Psychiatric Association, 1980). Alpert & Rosen (1990) argue against such a simplification and emphasise that *"affects do not become moods, no matter how long they persist"* (p. 239). In light of the inconsistencies found in the literature with regards to emotion-related terminology, for which a definition of categories by duration alone is only one example, it seems mandatory to, at this point, define how those terms will be used in the current thesis.

Emotion: An emotion can in a first step be described as a mental state that occurs spontaneously rather than through conscious effort, and moves an organism to action (cf. The American Heritage Dictionary of the English Language, 2004) This simplified conceptualisation forms the basis for all definitions relevant for the scope of the current thesis. Scherer (1986) defines emotion in more detail as *"a series of interrelated adaptive changes in several organismic subsystems following antecedent events evaluated to be of major relevance to an organism's goals"* (p. 146). This definition is in turn in accordance with Ekman who gives similar characteristics for what he refers to as 'basic emotions', a limited number of transient, discrete, emotional states such as happiness, sadness, anger, fear and disgust (see for example Ekman, 1992). In the present work the term emotion will be used in Ekman's sense of 'basic emotions' (further discussed in Section 2.2).

Affect: The term affect is often used synonymously to 'emotion' (cf. for example Scherer, 1986) or it is used to name a broader category of emotion-related experience (Forgas, 1995). I will use the term in this latter sense to refer to emotion-related behaviour that can occur for a longer time span, as a prolongation of a certain emotion. Due to its extension in time, affect can be expected to be less intense than an emotion in terms of both experience and display. Affect in this definition still presupposes an elicitation through appraisal of significant events, and has with that still a causal determinant.

Mood: The term mood refers to an observable state with a longer time course than an emotion or an affective state. It is often described as a diffuse affect state (see, for example, Scherer, 2003). Importantly, the elicitation through appraisal of a significant event is not a prerequisite. States like apprehension, euphoria and irritation can be described as moods (Ekman, 1992).

Valence: The term valence belongs to the dimensional representation of emotions, an approach that has been suggested to represent grades of emotions more adequately (Schröder, 2004; Mehrabian & Russell, 1974; Ortony, Clore & Collins, 1988). Valence refers to the dimensional scale of 'pleasantness'. i.e. positive/negative or liking/disliking. Other common dimensions in this model are 'activation' (or 'arousal') i.e. the readiness to act in some way, and 'power' (or 'control', 'dominance'), the associated social status (Schröder, 2004).

For many aspects of emotion research, valence models, which place emotions on continuous scales, and categorical models, which consider emotions as discrete entities, are difficult to reconcile. However, a recent theoretical conceptualisation proposed by Russell (2003) has been trying to show that it is possible to integrate the dimensional perspective with the categorical one. Russell puts the emphasis on what he calls 'core affect', *"that neurophysiological state consciously accessible as the simplest raw (nonreflective) feelings evident in moods and emotion"* (p. 148). This core affect is describable as a point in a dimensional space comparable to that of the valence models (see above). Core affect constitutes the basis of an emotional episode, which comprises, amongst others, a physiological response, a display of the experienced 'emotion' and an appraisal of the situation on the basis of the emotion, similar to the components of emotion Ekman (1992) proposes (discussed in more detail in Section 2.2). What makes an emotion categorical, as opposed to the continuously scaled core affect, is the fact that the experience of it becomes an emotional meta-experience, a categorisation of one's state based on the evaluation of all aforementioned components. This categorisation is closely connected to labels for emotions such as happiness, anger, fear, etc., which are acquired through social learning.

For the purpose of this thesis it is emotions, in the categorical sense, that are of main interest. Emotions are what speakers display, and what speakers are able to give self-reports about, and they are what listeners observe.

The next section will return to the components of emotion, describe them in more detail and thereby clarify the purpose and nature of emotions, in particular as a response system to significant events in the environment.

2.2 The Purpose and Nature of Emotions

According to Ekman (1992) emotions mediate quickly and efficiently reactions to significant events, thus contributing crucially to decision-making, including long-term planning and, more generally, to the "dealing with fundamental life-tasks". In line with that, emotions are rarely neutral, but mostly either negative or positive.

Darwin (1872) argued that emotions are universal and that they evolved via natural selection, an argument that has since been corroborated by evidence that emotions are controlled by one of the, in evolutionary terms, oldest parts of the brain, the limbic system. Basic emotions are displayed very early on, even though a fully differentiated set of emotions will only develop within the first years of life (Camras, 1992).

To distinguish emotions, and specifically 'basic emotions' from other affective phenomena such as moods and emotional attitudes, Ekman (1992) has proposed nine characteristics inherent within basic emotions. These characteristics reflect the notion of emotion being a *"'syndrome' with various components (e.g., physiological arousal and expression and feeling) in response to an evaluation of significant events in the environment."* (Scherer, 1986, p. 146).

1. *Distinctive universal signals:*

   Each basic emotion is associated with a distinctive, universal display. Particularly in the area of facial expressions, robust evidence for this has been provided (see Ekman, 1989, for an overview, but see also Russell & Fernández-Dols, 1997, for possible counter evidence). There is less consistent evidence in the area of vocal expressions (see Section 3.2).

2. *Presence in other primates:*

   This is an important prerequisite for the notion of basic emotions being innate and shaped by natural selection. One theory as to why emotions might be present in both humans and other primates assumes that all primates' brains are designed around reward- and punishment-evaluation systems with pre-specified goals, and that emotions play a crucial role as regulators within these evaluation systems

(Rolls, 2000). Accordingly, Johnson-Laird & Oatley (1992) hypothesise that each emotion prompts the organism in a direction which in the course of evolution has proven more successful than other solutions in similar goal-relevant circumstances. Thus, happiness can be, for example, seen as being associated with the maintenance or attainment of goals, while fear can be seen as being associated with the expectation of failure to achieve a goal. There is also evidence, though not based on systematic studies, that primates *display* emotions somewhat similarly to humans (Chevallier-Skolnikoff, 1973; Redican, 1982). Overall, however, research on the exact makeup of emotions in other primates is still largely lacking.

3. *Distinctive physiology:*

   Distinctive patterns of autonomic nervous system activity have been shown for anger, fear, disgust and sadness, suggesting that emotions facilitate motor behaviour adaptation appropriate for the events that have triggered the emotion. For example, fleeing from a predator is an adaptive action for fear, which is in line with evidence that blood goes to large skeletal muscles at times of fear (cf. Levenson, Ekman & Friesen, 1990). According to Ekman (1992), this is a direct contradiction to theories that consider emotion a social construction lacking biological correlates (Averill, 1980; Cornelius, 1996).

4. *Distinctive universals in antecedent events:*

   If emotions are innate and universal it can be assumed that there are common elements to the events that precede and elicit an emotion. These antecedent events can occur internally or externally. Ekman (1992) gives the example of an repellent taste or smell as the antecedent event for disgust. However, the specific details of antecedent events might be variable between cultures and individuals.

5. *Coherence among emotional response:*

   There is not yet conclusive evidence for whether or not there is coherence in emotion expression and the changes in the autonomic nervous system activity subsumed under 3., i.e. whether there is a systematic relationship between display and physiological response. Incoherence might occur due to differences in normative behaviour or individual differences in temperament, personality, (self-)control, etc.

6. *Quick onset:*

   The quick onset of emotions allows the system to respond to important events with very little expenditure of time. This is particularly advantageous where very little time is left for considerations or preparations, such as in situations of immediate danger.

7. *Brief duration:*

   An emotion is only prolonged in its time course if the emotion is evoked again. The brief duration of emotions allows the system to maintain its flexibility and to respond quickly to the next important event. Motor behaviour or changes in the autonomic nervous system activity associated with a particular emotion may last longer than the emotion itself. Ekman (1992) suggests that multiple expressions of a certain emotion over a longer time span might actually be the expression of a sequence of *"repeated but discrete emotion episodes"* (p.186). Notably, brief duration distinguishes emotions from moods and affect.

8. *Automatic appraisal:*

   The occurrence of an emotion is preceded by an evaluation of a situation determining whether a stimulus will evoke an emotion and, if so, which. This type of appraisal operates automatically, allowing a quick response. When appraisal is conscious and deliberate, it might arouse or alert the system but not necessarily evoke an emotion.

9. *Unbidden occurrence:*

   It is the automaticity and quickness of emotion elicitation that makes it difficult for the individual to control emotions. It is, however, also dependent on individual differences how robustly and quickly emotion responses occur.

Taking these characteristics together it can be said that emotions occur rapidly, through automatic appraisal of the environment, with little awareness and with involuntary response changes in expression and physiology – making it a highly efficient and, from a cognitive point of view, uncostly system.

While this is by no means an exhaustive account of the psychological aspects of emotion, it is a base for further considerations regarding the interplay of affect and communication (see Chapter 3).

2.3 Choosing One Emotion: Happiness

Emotions can be seen as an uncostly response system for the human organism. As an object of research, emotions demand, however, complex analyses on different levels. Ekman's extensive breakdown of the components of emotions makes this very clear (cf. Section 2.2). If the aim is a thorough and holistic analysis of the whole communicative path from speaker to listener, as in this thesis, it is sensible to investigate one emotion at a time rather than a whole catalogue of emotions at once. Furthermore, when examining emotions under laboratory conditions, constraints of feasibility have to be taken into account as well as ethical considerations. I chose 'happiness' because:

1. Unlike fear or anger, happiness is easier to assess and elicit in a large number of speakers, particularly if strong mood-induction techniques or disturbing stimuli are to be avoided.

2. Happiness is ubiquitous in everyday life, and it could be expected that there would be no social constraints on displaying happiness in the experiments designed for this thesis.

3. Happiness, unlike many other emotions, is relatively well described and understood, not just in view of facial cues, particularly smiling (e.g. Ekman, 2001; Frank, Ekman & Friesen, 1993), but also in the context of speech communication. Research on child-directed speech (Kitamura & Burnham, 1998; Singh, Morgan & Best, 2002), as well as publications in speech technology (Cahn, 1990; Schröder, 2004) have significantly contributed to this.

4. The display of happiness plays a crucial role in social interaction, especially in the interaction between parents and children, where it is vital for the child's development (Singh et al., 2002; Nicely, Tamis-LeMonda & Bornstein, 1999; Kaplan, Bachorowski & Zarlengo-Strouse, 1999).

However, it has to be understood that it is virtually impossible to examine basic emotions in a laboratory setting. Emotions, as defined by Ekman (1992), are very brief events, possibly too brief to determine, for example, their vocal correlates over the course of a speech task. When looking at these correlates it is probably more adequate to refer to correlates of a less categorical emotion-related state. This is why for the experimental aspects of this thesis I will mainly refer to affect, the in time extended emotion episode (see Section 2.1).

Chapter 3

Positive Affect and Communication

Chapter 2 described the role of emotions for the individual, in particular as a response system to significant events in the environment. This chapter describes the role affect, the in time extended emotion episode, plays for communication between individuals, with the emphasis on speech communication.

The first part of this chapter will demonstrate that the impact of affect is based on the cognitive and physiological changes affective states cause in the speaker, as well as on the signalling function of affect. The second part of this chapter will examine a basic requirement for affect playing a role in verbal communication: the detectability in the speech signal. The last part of this chapter will derive predictions for the two studies of this thesis to test the impact of positive affect expression on verbal communication.

3.1 What Could Make Affective Speakers More Effective Speakers?

The aim of this thesis is to explore whether positive affect causes changes that facilitate the communication process. There are a number of reasons to believe that the cognitive changes that a positive emotional state evokes in the speaker might be beneficial to the communication process. At the same time, it can be assumed that the signalling of positive affect itself will have an impact on the communication process.

3.1.1 Positive Affect and Verbal Working Memory

There is evidence that affect can transiently enhance some cognitive abilities and impair others, for example with regard to working memory. Approach states, evoked by a positive appraisal of an event and linked to activation patterns in the left hemisphere, have been shown to increase verbal working memory capacity, while withdrawal states, evoked by a negative appraisal of an event and linked to activation patterns in the right hemisphere, have been shown to increase spatial working memory capacity (Davidson, 1992; Gray, 2001). Verbal working memory capacity is contributing to speech production as well as speech perception (Caplan & Waters, 1999). Baddeley (1986), who suggested the multi-component model of working memory, proposed the existence of an 'articulatory loop' in the verbal working memory. The articulatory loop is thought to play a critical role in storing or 'buffering' verbal information during the computation of speech input or output, by repeating (and thereby refreshing) information within the decay time of the loop. This process is often compared to a silent, repetitive articulation, similar to what one might do to memorise a phone number, and is thought to facilitate the managing of cognitive load. In experiments, participants' verbal working memory capacity is usually assessed by tasks like the Reading Span Task (Daneman & Carpenter, 1980).

Extensive research on speech perception has demonstrated the benefits of increased verbal working memory capacity (for a overview over this line of research see Caplan & Waters, 1999). Increased verbal working memory capacity has, for example, been shown to help assign the syntactic structure of a sentence to determine sentence meaning (e.g. Waters & Caplan, 1996). Low verbal working memory capacity, on the other hand, has been shown to be particularly detrimental in the processing of syntactically complex sentences (e.g. King & Just, 1991).

Less evidence is available for aspects of speech production. Increased verbal working memory capacity has, however, been reported to facilitate articulatory control (Baddeley, 1986) and has also been shown to be directly linked to increased speech rate (cf. Raine, Hulme, Chadderton & Bailey, 1991). The rationale is that an increase in articulation speed reduces the chunk size of units, thus enabling more information to be rehearsed in the articulatory loop, as the capacity of the verbal working memory is quicker exceeded when chunks are longer. One could also speculate that increased verbal working memory capacity is likely to facilitate particularly taxing aspects of speech production such as producing syntactically complex sentences (similar to what has been observed on the perception side) or taking into account listener perspective by being able to simultaneously process perceptual considerations, for example, in view of different possibilities for sentence interpretation.

Based on the notion that positive affect increases verbal working memory capacity (Gray, 2001), one can hypothesise that via this increased verbal working memory capacity, positive affect facilitates the communication process on both sides of the speech chain, in the speaker as well as in the listener.

3.1.2 Positive Affect and Social Cognition

Central to Ekman's definition of basic emotions is the appraisal of environmental stimuli (see Section 2.2). In the context of speech communication, this appraisal will, amongst other things, necessarily be an emotional appraisal of the interlocutor, and the social situation the speech interaction takes place in, which in turn can have further impact on the communication process itself. This cross-influence of emotional states in social interaction is conceptualised in the Affect Infusion Model (Forgas, 1995), according to which the emotional state elicited by appraisal of one event can directly influence the appraisal of another event.[1] When the appraisal is positive, resulting in a state of positive affect in the speaker, a number of further benefits for the communication process can be anticipated. Positive affect might facilitate greater accessibility of positive memories of social situations (cf. Bower, 1981) which, in turn, might make a speaker more inclined to engage in communication. It might also help to weigh the cost and benefit of various communication strategies (cf. Forgas, 1998b), promoting the achievement of communicative goals.

A number of benefits of positive affect also have been also proposed under Fredrick-

[1] For a definition of 'appraisal' see Appendix E.

son's Broaden-and-Build Theory (e.g. Fredrickson, 2003). Based on her own research and that of Isen (1987), Fredrickson suggests that, *"[...] when people feel good, their thinking becomes more creative, integrative, flexible and open to information"* (p. 333). Global-visual processing tasks employed in Fredrickson (2003) revealed, for example, that participants who experienced positive affect exhibited a broadened pattern of thinking and processing compared to neutral or sad control groups. Similarly, Isen (1987) could show that in word-finding tasks where creativity was a decisive factor, 'happy' participants would score higher. Fredrickson (2003) even believes that the impact of positive affect is not just short-term: *"By momentarily broadening attention and thinking, positive emotions can lead to the discovery of novel ideas, actions and social bonds."* (p. 333). Fredrickson's and Isen's studies are quite different to Forgas' studies in focus, but all of this research points in the same direction: 'Happy' people have seemingly more resources at their disposal.

While this could also mean that positive affect makes a speaker more cooperative (cp. Forgas, 1998b), it has interestingly been shown that 'happy' speakers are not necessarily more polite speakers. Forgas (1998a) found that 'happy' speakers formulate requests in more direct, less polite ways than 'sad' speakers who tend to be more indirect and polite (see also Forgas, 1998a,b, 1999a,b). This behaviour could be attributed to an overestimation of the potential success of their social interaction by the 'happy' speakers, but it might also just reflect the fact that less effort is needed in some aspects of communication when positive affect is signalled.

The signalling of an emotional state itself is important for the communication process.

It will allow the listener to gain insight into the speakers' appraisal of events. This may provide contextual and meta-communicative information, and can allow the listener to compare the speaker's appraisal with his own to estimate agreement. It might also up to a certain degree regulate listener behaviour. The signalling of positive affect might, for example, reinforce the listener's communicative strategy or reactions. The signaling of negative affect, on the other hand, might inhibit certain listener behaviour. Finally, perceiving positive affect might trigger direct physiological or psychological response in the listener. Research on facial expressions of affect suggests that receivers subconsciously 'mimic' affect expressions (Hatfield, Cacioppo & Rapson, 1992; Dimberg & Öhman, 1996; Dimberg, Thunberg & Elmehed, 2000) and it can be assumed that the same might be true for receivers of vocal affect expression. In line with this assumption Bachorowski & Owren (2001) report that perceived vocal affect expression alters listeners' self-reported affect.

However, these functions of signalling do not presuppose that the signalled affective state matches the *genuine* affective state of the speaker. It has been conclusively shown that an affect display does not automatically express genuine affect (Scherer, 1986; Ekman, 1992; Scott, Young, Calder, Hellawell, Aggleton & Johnson, 1997) and that speakers have up to a certain degree voluntary control over this display (Gross, 2002). From an evolutionary perspective, displays of emotional states are thought to have evolved to indicate vital information about important events such as immediate danger, etc. (Ekman, 1992). Some accounts assume that affective expressions have then developed further to manipulate listeners in ways beneficial to speakers (Russell, Bachorowski & Fernández-Dols, 2003). Accordingly, speakers may use affective expressions to capture the listener's attention. For example, exaggerated pitch contours observed in child-directed speech might be employed to capture the attention of the child (Fernald, Taeschner, Dunn, Papousek, Boysson-Bardies & Fukui, 1989). Affective expressions can also be utilised to signal appeasement, i.e. to calm and reassure the listener. Owren & Bachorowski (2003) observed, for example, increased frequency of laughter in interactions with interlocutors of the opposite sex (see also Grammer, 1990). Such manipulation does not have to be disadvantageous for the listener. The attention eliciting features of child-directed speech are thought to contribute greatly to language acquisition and learning (Fernald, 1994; Morgan & Demuth, 1996; Thiessen, Hill & Saffran, 2005), the signalling of appeasement might have great advantages in terms of creating listener rapport. At the same time, there can also be downsides to the signalling of emotional states. If a speaker cannot suppress his genuine emotional state this might be of disadvantage as much as a listener falsely believing he has information about the genuine emotional state of the speaker. Affective expression is, thus, determined by genuine 'leakage' as well as social display rules with signalling function. Overall, however, there is reason to suggest that positive affect facilitates communication at both ends of the speech chain.

The last sections have outlined how positive affect can potentially facilitate communication. The majority of the above mentioned effects presuppose that speakers express their experienced positive affect and that listeners recognise it as such. The following sections will thus give an overview over the vocal correlates of affect encoding and decoding.

3.2 Vocal Affect Expression

Emotions are associated with certain vocal correlates due to the physical response that emotions evoke in the body (Ekman, 1992; Scherer, 1986). From an evolutionary perspective, these physical responses are mammalian elaborations of vertebrae arousal patterns, in which neurochemicals (e.g. dopamine, noradrenaline, and serotonin) increase or decrease the brain's activity level (Oatley, Keltner & Jenkins, 2006; Panksepp, 1998). Changes in the brain's activity level manifest themselves in changes in muscle tension, breathing and posture. Ekman (1992) captures these correlates in his third characteristic of emotions, the distinctive physiology (see Section 2.2). Vocal settings, i.e. the adjustments in the larynx and the oral cavity for speech production, seem to be particularly prone to subtle and at the same time constantly occurring changes of that type. Scherer (1986, p.148) describes this sensitivity of the vocal settings:

> *Take the following simple example: After appraisal of a stimulus as dangerous and requiring action [...], fundamental frequency of the voice (F_0) will increase because of the functional response of the SNS (increased muscle tension); at the same time, salivation will decrease because of the changes in the ANS (sympathetic dominance). Both effects will tend to make the voice sound more high-pitched (changes in vocal fold vibration and vocal tract resonance). If, a split second later, the event is reevaluated as a hoax, the state of the fast-responding SNS will change again, lowering pitch because of a decrease in muscle tone. The ANS is slower and it is likely that the effect of*

reduced salivation on vocal tract resonance will persist for some time.

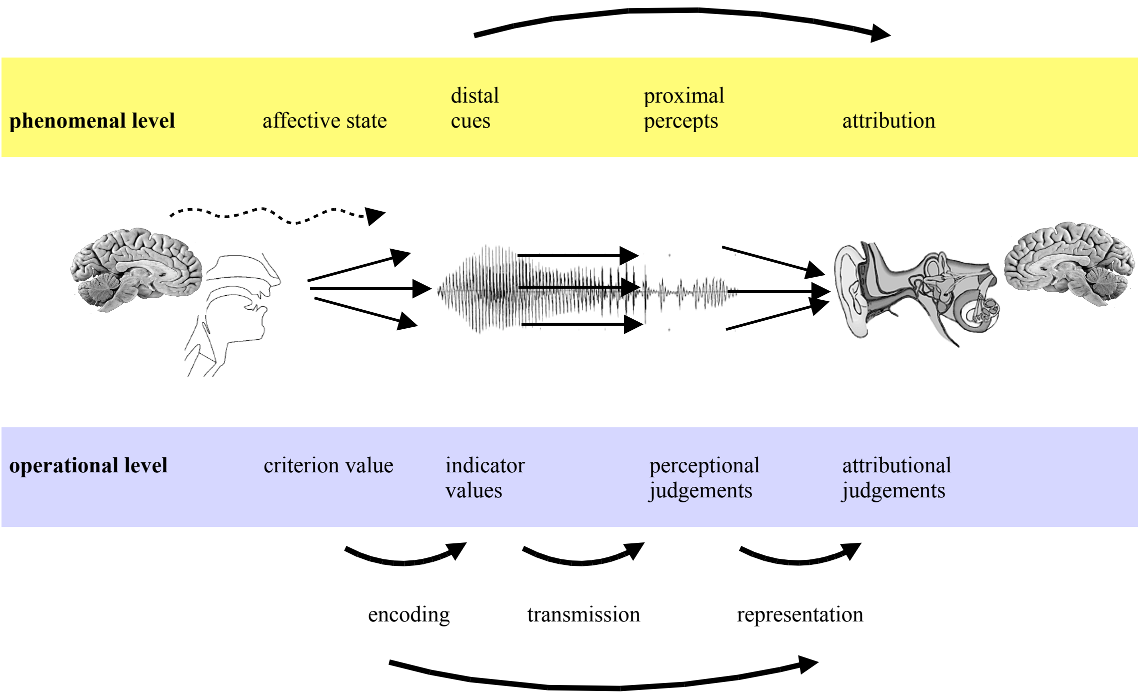
Certain vocal features can thus be seen as the overall result of a number of involuntary effects that consecutively modify the vocal settings and may all leave their traces in determining the vocal features at a particular point in this dynamic process. In addition, there might be voluntary effects that have their cause in, for example, affect control or self-presentation.

3.2.1 Tracing Vocal Affect Expression Through the Speech Chain: The Brunswikian Lens Model

To capture the whole process of vocal affect expression, Scherer (2003) suggests for research to employ the Brunswikian Lens Model (see Figure 3.1). This model is based on Brunswik's model for the study of visual perception (Brunswik, 1956) and is essentially an application of it to the speech chain models. Central to all speech chain models is the representation of the connection and interplay between speaker and listener along a temporal continuum, and of how a message is transmitted along this continuum passing through various, physically different stages, such as the speaker's vocal apparatus, the acoustic signal and the listener's organ of hearing (cf. for example Tillmann, 1980; Denes & Pinson, 1993).

In his adaptation of the Brunswikian Lens Model Scherer (2003) incorporates two levels, the 'phenomenal level' and the 'operational level'. The rationale is that, on what Scherer calls the phenomenal level, the emotional state affects vocal settings in the speaker (as outlined on pp. 22) which is represented in acoustic cues. These vocal correlates are

Figure 3.1: Schematic of the Brunswikian Lens Model as adapted by Scherer (1986)



initially remote from the observer (*distal cues*). The distal cues get transmitted as part of the speech signal and are then perceived as *proximal cues*, i.e. close to the observer and via the perceptual organ resulting finally in an attribution of the speaker's affective state by the listener. An example for distal and proximal cues are F_0 and pitch, respectively. To verify the model fit between stages, i.e. from affective state to distal cues, from distal cues to proximal cues, and from proximal cues to attribution in the listener, each element can be measured on an 'operational level'. This level is the level on which the phenomenon of affect transmission is studied, and constitutes the second layer of the Brunswikian Lens

Model as suggested by Scherer.

The model is a useful means for describing the impact of vocal affect on speech communication, as it accounts for and, thus, allows to model each step of the communication process. It is also useful to describe how faulty or missing links in the chain can distort the process, and how this, in turn, makes the encoding and decoding of vocal affect less reliable (Scherer, 2003). This important aspect of vocal affect expression will be further discussed in Section 3.2.4. Before that, the next two sections will provide an overview over the features that have been identified to be crucial for the production and perception of positive affect expression.

3.2.2 Production: Vocal Correlates of Positive Affect

Positive affect can have a number of effects on vocal production, the distal cues in Scherer's adaptation of the Brunswikian Lens Model. These effects can be evoked by physiological changes such as changes in breathing, posture and muscle tension. Changes in facial expression might in addition change settings of the oral cavity. Smiling is a prime example for such a change.

There is, however, one particular problem with positive affect. Ekman (1992) uses an overall category – which in this thesis is termed 'happiness' – for a whole catalogue of emotions such as amusement, contentment, relief, sensory pleasure, pride in achievement, the thrill of excitement and satisfaction, and argues that more than for any other family of emotions there is a many-to-one mapping in that there are no distinctive signals to distinguish each of these positive emotions. The reason for this is that different types of enjoyment have very different arousal patterns thus resulting in qualitatively very different physiological changes.

An insufficient sub-categorisation might thus be one explanation for why there are often inconclusive data for acoustic correlates of happiness. While this does not mean that research should abandon a general term like happiness, correlates found in the process of studying 'happiness', particularly if they are contradictory to results from the literature, should be carefully evaluated as belonging to either elation/joy or to enjoyment or indeed the overall category happiness.

Vocal features that have been reported, or often just predicted, to correlate with the expression of happiness in the sense of 'elation/joy' are, for example, fast speech rate, high pitch and increased pitch range, greater pitch variability, increased mean intensity, greater variability in intensity, higher amplitude and lower voice perturbations[2] (Bachorowski & Owren, 1995, for an overview see Murray & Arnott, 1993; Scherer, 2003), as a result of lip spreading in smiling a raise in the resonant frequencies of the vocal tract, namely F2[3] and F3, and to a certain extent F1 (Tartter, 1980; Robson & Mackenzie Beck, 1999), as well as a tense or harsh voice quality (Gobl & Ní Chasaide, 2003). The increase of speech rate in speakers who experience elation/joy seems to be in line with the increase of speech rate reported in connection with increased verbal working memory capacity (cf. Section 3.1.1). This example highlights once more the complex interplay of cognitive and physiological factors in the impact of emotional states on speech production.

[2]For a definition of voice perturbations, refer to the terms 'jitter' and 'shimmer' in the Glossary (Appendix E)

[3]For a definition of F1, F2, F3, refer to the term 'formant' in the Glossary (Appendix E)

Changes associated with 'elation/joy' might also alter the perceptibility of parts of speech. Peaks in pitch and intensity might emphasise important parts of utterances, while raises in F2 and F3 might have similar effects like formant changes that occur in the context of hyperarticulation, the exaggerated production of vowels and consonants (Erickson, Fujimura & Dang, 1999; Burnham, Kitamura & Vollmer-Conna, 2002; de Jong, 1993; de Jong, Beckman & Edwards, 1993). These potential effects are further aspects of positive affect facilitating the communication process.

The expression of happiness in the subdued sense of enjoyment resembles much more neutral states of affect, and the correlates reported here are often opposite to what is reported for elation/joy, particularly with respect to parameters connected to F_0, intensity and speech rate (Scherer, 2003).

For all parameters, it should be kept in mind, however, that hardly any of them have been confirmed reliably, across different studies. There is still not enough basic research on vocal affect expression overall, and there is in particular a lack of systematic and ecologically valid experiments that could corroborate parameters of positive affect based on genuine emotions rather than acted emotions (for a discussion see Scherer, 1986; Scherer, 2003, as well as p. 35).

3.2.3 Perception: Positive Affect Recognition in Speech Communication

The perception of facial expressions of emotional states has been shown to be robust, even across cultures (Ekman, 1989). However, the perception of vocal expressions of emotional states is less straightforward. Scherer (1986) attributes the better recognition of facial expressions to the notion that changes in muscle activity can be comparatively more directly observed in the face, whereas listeners have to infer the vocal expression of affect from the acoustic signal, the proximal cues.

Within the domain of vocal affect expression, recognition rates are generally much higher in non-verbal than in verbal stimuli, which might be due to the often more emblematic nature of non-verbal expressions of affect (Scott et al., 1997). Furthermore, affect in verbal stimuli is best recognised when it is acted rather than spontaneous, suggesting that listeners are more readily prepared to recognise prototypical displays of affect (Banse & Scherer, 1996). Interrater agreement is, moreover, markedly higher if the listeners' task is to attribute emotional states from a fixed list of choices rather than to label emotional states freely (Russell et al., 2003).

Acoustic parameters attributed to positive affect in perception studies are a broader pitch range, higher pitch, increased speech rate, modal or tense voice and, to a certain degree 'vowel precision' or hyperarticulation (Burkhardt & Sendlmeier, 2000; Scherer, 2003). It has to be again kept in mind, however, that these findings need more corroboration through ecologically valid experiments.

3.2.4 The Link Between Production and Perception

The link between production and perception is a notoriously difficult issue for research on vocal affect expression (Scherer, 1986).

To date, research on the vocal expression of emotional states has mainly focused on the relationship between speaker affect and vocal cues on the one hand, or vocal cues and perceived affect on the other hand (Banse & Scherer, 1996; Barrett & Paus, 2002; Gobl & Ní Chasaide, 2003; Russell et al., 2003; Scherer, 2003), implying a close agreement between what is produced and what is perceived. However, the notion that senders express their internal state through a range of reliable facial or vocal cues, and that receivers decode internal states of senders using exactly those cues, has been consistently challenged (Mozziconacci, 2002; Scherer, 1986).

The potential problems arising on the path between the speaker and the listener can be observed at each stage along the Brunswikian Lens Model (see Section 3.2.1). First, there are a number of reasons why affective state and distal cues might not match. Most importantly, there is, as discussed earlier, strong empirical evidence for the existence of incongruities between affective states and the way they are displayed (Ekman, 1992; see Section 2.2). Secondly, one and the same affective state can be expressed by different means, not least due to individual differences (Scherer, 2003). This effect can be amplified by the fact that the emotional state in itself is not a categorical entity but rather can be a mixed form and can occur at different grades of intensity, for example due to affect control by the speaker (Gross, 2002). It can also be expected to make a qualitative difference whether the distal cues are leaked as a physiological by-product, or signalled, that means consciously encoded (see Section 3.1.2). Thirdly, there can be mismatches on the transmission path between distal cues and proximal cues through, for example, distortion in the transmission channel (distance, noise, etc.) and in the perceptual organ[4] (Scherer,

[4]This, of course, is not a phenomenon specific to vocal affect expression but is true for all speech

2003). Fourthly, there can be mismatches between the proximal percepts and what the listener eventually attributes as perceived speaker affect. One factor that can cause such a mismatch are discrepant stereotypes that might lead to attributions different to what certain vocal cues might signal otherwise (Scherer, 2003). Fifthly, as for many aspects of speech communication, it is ultimately not known which cues listeners are actually making use of, and how these cues are weighted against each other.

Thus, the vocal expression of affect is indeed not a matter of a one-to-one mapping. Each potential point of mismatch along the path between speaker and listener provides an explanation for why recognition rates for vocal expressions of affect are usually low. It also becomes clear why listeners are more inclined to identify acted emotions (for both cf. Section 3.2.3), as acted emotions are not just more categorical (usually one specific emotion is acted out rather than a mixture and it is usually a full-blown display of that emotion rather than an attenuated version) but also more reflective of sociocultural norms or expectations (Scherer, 2003). One can speculate that such affect 'portrayals' are less prone to interference in the transmission process, simply because they are prototypical patterns yielding reliable behaviour in the speaker and the listener. That this might often not be the case with naturally occurring affect has to be kept in mind for studies that attempt to examine a whole pathway of transmission of genuine vocal affect from the speaker to the listener, as will be the case in Study 1 (Chapter 4).

The last sections have established how positive affect is encoded and decoded, respectively. The next section will return to the functionality of positive affect in speech com-

communication.

munication, with the aim of deriving predictions for the two studies of this thesis.

3.3 Predictions for the Studies

Previous sections have identified numerous ways in which positive affect might facilitate verbal communication. From these, one can derive a number of predictions. 'Happy' speakers could be, for example, more cooperative, and engage more readily in communication (see Section 3.1.2). 'Happy' speakers could be encoding messages more efficiently, for example due to increased verbal working memory capacity (see Section 3.1.1). Listeners, on the other hand, could be more inclined to listen to speakers who signal positive affect (see Section 3.1.2). Listeners could also be better able to comprehend parts of speech that are altered by by-products of positive affect expression (see Section 3.2.2). To enable a systematic exploration of the impact of positive affect on communication these possibilities will be described in form of three different models.

3.3.1 Vocal Correlates of Affect Expression and Communication

There are at least three ways in which positive affect could facilitate communication. Vocal cues associated with positive affect expression could:

1. directly support the comprehension of the message (henceforth *Direct Route*).
2. signal speaker affect through para-linguistic features which may elicit attention and facilitate communication (henceforth *Mediated Route*).

3. signal speaker affect through para-linguistic features which may elicit attention and signal speaker affect in addition to directly supporting comprehension of the message (henceforth *Dual Route*).

Direct Route:

Vocal correlates affect comprehension

First, it is possible that affect has a direct effect on the content and structure of the message, i.e. on its pragmatic, semantic, morpho-syntactic, or prosodic aspects. Some preliminary evidence for such effects on the pragmatic level comes from social-psychological research, where it has been, for example, shown that happy speakers formulate requests more directly, while sad speakers interact more politely (Forgas, 1999a; see Section 3.1.2). Furthermore, it can be hypothesised that positive affect may enable speakers to encode messages in a more efficient way, thereby producing utterances optimised in terms of, for example, their semantic and morpho-syntactic structure. This latter idea is in line with the study by Gray (2001; see Section 3.1.1) where emotional valence was conceptualised in terms of approach and withdrawal states and showed that participants experiencing approach states exhibited increased verbal working memory performance, while participants in withdrawal states showed increased spatial working memory performance. Such an effect of emotional valence on verbal working memory capacity may influence particularly taxing aspects of language production (see Section 3.1.1). However, to date, there is no direct empirical evidence for this hypothesis.

Secondly, it is possible that positive affect expression is associated with a specific range of vocal correlates, and these vocal correlates may not just contribute to the listeners' inferences about the underlying emotional state of the speaker, but may themselves have a direct effect on communicative effectiveness. Positive affect expression may alter vocal cues that contribute to the prosodic structure of an utterance. To date, there is controversy as to whether linguistic and affective prosody are separate and independent (McRoberts, Studdert-Kennedy & Shankweiler, 1995; Seddoh, 2002; Wong, 2002). For example, increased pitch range, and salience of pitch peaks associated with positive affect expression may serve to elicit the listener's attention (cp. Werker & McLeod, 1989), but may also increase emphasis on relevant parts of the message (Gussenhoven, 2002), which, in turn, may facilitate listener comprehension (Trainor & Desjardins, 2002). At the same time, the increased pitch range may independently contribute to the impression of speaker happiness. A prime example of this mechanism is the increased pitch range in child-directed speech, a feature that has been shown to elicit attention in infants and young children (Fernald, 1994), to highlight relevant parts of the utterance (Morgan & Demuth, 1996), and to thereby facilitate learning (Thiessen et al., 2005; Kempe, Brooks & Gillis, 2005; Kempe, Brooks, Mironova & Fedorova, 2003).

Mediated Route:

Vocal correlates elicit attention and signal speaker affect

Another possibility is that speaker affect may aid the listener in comprehending the message via the perception of affect. It is possible that the perception of positive speaker affect induced independently of the content of the message by the accompanying vocal cues may help to establish a relationship of trust and rapport, perhaps by contributing to the impression of submissiveness and friendliness of the speaker (Gussenhoven, 2002). This, in turn, could predispose the listener to pay closer attention to the message. Again, child-directed speech provides an instructive example for this Mediated Route: It has been suggested that its typical vocal and prosodic features such as reduced speech rate, heightened pitch and wider pitch range have the effect of establishing and strengthening the emotional bond between caretaker and child (Fernald, 1994). In fact, some authors have proposed that the primary function of child-directed speech is to manipulate arousal in infants and small children, and that this emotional relationship is at the basis of communicative development (Locke, 2001). Indeed, there is evidence that emotional responsiveness as expressed by various para-linguistic devices plays a causal role in facilitating children's linguistic development (Nicely et al., 1999) as it suggests to the child that communication is primarily about sharing emotional states. Thus, the expression of speaker affect may influence the social and emotional interaction with the listener, thereby affecting communicative effectiveness indirectly.

Dual Route:

A combination of the Direct and the Mediated Route

A third possibility is that a combination of the Direct and the Mediated Route operates in a complementary fashion in human communication, where positive affect can, on the one hand, modulate the linguistic content and the vocal characteristics of the message, thereby directly affecting comprehension of the listener, and, on the other hand, influence the success of communication mediated through the listener's perception of the emotional state of the speaker and the social consequences of this perception.

3.4 Testing the Predictions

To examine the effect of positive affect on communication two studies were carried out. The first study is correlational in nature, relating the perceived emotional valence of speakers to their communicative effectiveness. Communicative effectiveness here is assessed as listener performance in a Map Task paradigm, with a measure based on overall message comprehension. The second study aims at a specific verbal communication problem, the disambiguation of ambiguous sentences by prosodic means. Communicative effectiveness here is assessed in terms of occurrence of relevant prosodic features, in combination with a comprehension test.

For both studies, the affective state of the speaker had to be controlled for. As extensively discussed above, the vocal expression of affect is subject to a number of subtle changes evoked by involuntary as well as voluntary response to emotions. In recent literature, doubt has been cast on whether studies based on acted emotions can provide valid insights into vocal affect expression (see also Section 3.2.4). Scherer (1986) argues that acting, rather than experiencing, emotions might be more of a mimicking process, drawing on stereotypes. These stereotypes might be what listeners reliably recognise, but it is questionable whether actors experience the cognitive and physiological changes accompanying emotions to the full extent. As the current studies attempt to explore the functionality of vocal expression of affect and not just its correlates, using acted emotions would be a problem as they might fail to reflect all of what is happening in the production process. The increase in verbal working memory capacity in approach states (see Section 3.1.1) is one example for potential shortcomings of that type. Ultimately, too little is known about the impact of emotions on the speaker to draw conclusions from studies based on acted emotions, other than conclusions about acted emotions.

To maximise ecological validity, the first study examined naturally occurring variation in speaker affect, which was assessed based on speakers' self-reports and listeners' affect ratings, respectively. This study was exploratory in nature. In the second study, positive affect was controlled using a speech register where positive affect is thought to play an essential role: child-directed speech. Positive speaker affect was thus in both studies as much as possible positive affect as it would occur in a natural environment.

Part II

Studies on the Impact of Vocal Affect Expression

Chapter 4

Study 1: Does Vocal Affect Enhance Communicative Effectiveness?

4.1 Introduction

Chapter 3 highlighted the fact that most studies on the relationship between affect and speech have focused on the vocal expression of affect emphasising the correlation between acoustic parameters and the emotional state of the speaker (Banse & Scherer, 1996; Barrett & Paus, 2002; Gobl & Ní Chasaide, 2003; Russell et al., 2003; Scherer, 2003). In analogy to research on facial expression of emotional states, this research has attempted to link underlying emotional states with specific acoustic profiles (Banse & Scherer, 1996; Scherer, 2003). However, little is known about how the emotional state of the speaker affects the way in which messages are encoded and decoded in verbal communication. In this study[1], the effects of affect on communicative effectiveness are explored, based on the different aspects of the interplay between emotion and verbal behaviour that have been discussed in previous chapters. What are the consequences of a positive emotional state of the speaker for the success of a verbal interaction?

Section 3.3.1 outlined that there are at least three possible ways in which the vocal expression of positive affect might affect communication: directly, mediated and in a dual fashion. To test which of these routes are at play in normal adult communication, this study attempts to quantify communicative effectiveness in order to explore how it is influenced by speaker affect. In this sense, this study has an exploratory character as the aim is to compare the three Route Models to determine which shows the best fit to the data.

The research strategy took into account that eliciting strong positive affect in participants under laboratory conditions is difficult if verbal emotional stimuli (e.g. stories) or verbal elicitation techniques (e.g. the Velten procedure; Velten, 1968) are to be avoided. Avoiding verbal material, however, is important to prevent a confound between independent and dependent variables when the impact of affect on linguistic structure and content, as well as para-linguistic features, is at issue. I therefore opted for obtaining self-reports of current affective states from a large sample of participants. This allowed to study affect within a more naturalistic and ecologically valid setting, even though the range of intensity of experienced affect is probably narrower than in mood induction studies or

[1]This chapter is an adapted version of Biersack & Kempe (2005a)

emotionally challenging situations. Participants rated their affective state after completing a referential communication task which required them to instruct a potential listener to re-draw a route on a simple map. Effectiveness of communication was assessed by measuring the performance of independent listeners who were asked to follow the speakers' instructions. I specifically opted for this dissociation of speaker and listener as the aim was not to test whether a dyad of one speaker and one listener was communicatively efficient, not least because the evaluation of such dyads could have been confounded by hard-to-control factors (such as how well dyads are matched, how good the listener's feedback is, etc.) while not really allowing for generalisations in terms of effectiveness of the speaker's instruction. Rather I was interested to see how efficient the instruction of a particular speaker was for a number of different listeners and compared to other speakers, with regards to choice of words, makeup of prosodic variation, etc. Perceived speaker affect was assessed by yet another group of listeners. In addition, a range of acoustic parameters was measured to determine how the emotional state of the speakers affected vocal cues, and how these vocal cues in turn affected communicative effectiveness.

4.2 Method

4.2.1 Participants

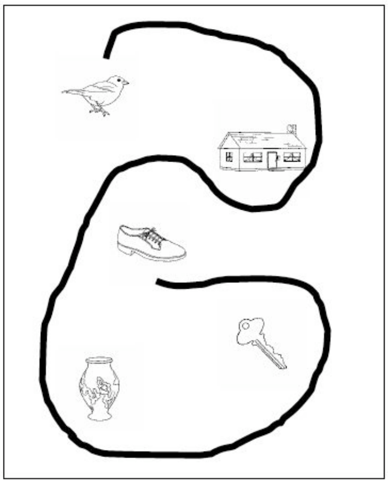
Eighty-eight men and 112 women, undergraduates at the University of Stirling, all native speakers of English, aged 16 to 43 years (mean age 23 years), received course credit or a £4 reimbursement for participation in the production task of the study. 600 men and women, aged 18 to 66, years participated in the comprehension task. These participants were recruited at the Glasgow Science Centre and were thus not familiar with any of the speakers. Another 37 men and 43 women, aged 18 to 51 years (mean age 32 years), staff and students from the University of Stirling, participated in the affect rating task.

4.2.2 Procedure

Production: For the production task, I constructed a simple map containing five objects as landmarks as well as an inverted s-shaped route (see Figure 4.1), to be used in a Map Task (Brown, Anderson, Shillock & Yule, 1984). The 200 participants were then asked to describe the route so that potential listeners would be able to re-draw the route given just the landmarks. In order to obtain comparable speech samples for acoustic analyses, participants were instructed to start with two pre-specified 'calibration' sentences: "Okay, I'm now going to describe the route to you. Please draw the route exactly as I say." For all acoustic measurements, and for the ratings of perceived speaker affect in the perception task, the second calibration sentence was used. After completing the Map Task, which took on average 58 seconds, participants rated their current emotional state using the Brief Mood Introspection Scale (BMIS)[2] (Mayer & Gaschke, 1988) consisting of eight positive and eight negative 'mood adjectives' which participants rated on a scale from 1 (definitely do not feel) to 4 (definitely feel). The BMIS was administered after the Map Task since pilot studies which manipulated the timing of the self-reports had shown that self-reported emotional states tend to remain stable throughout this task. All tasks were

[2]For an example of the BMIS, as used in this study, refer to Appendix A.

Figure 4.1: Example for a simple map constructed for the Map Task



completed following instructions in a booklet, and in the absence of the experimenter so as not to influence the speakers' mood, and not to elicit any demand characteristics.

Comprehension: For the first perception task, 600 participants were asked to re-draw the route while listening to the audio-taped descriptions. Each description was played to three listeners via high-quality Beyerdynamic DT 250 high-quality headphones.

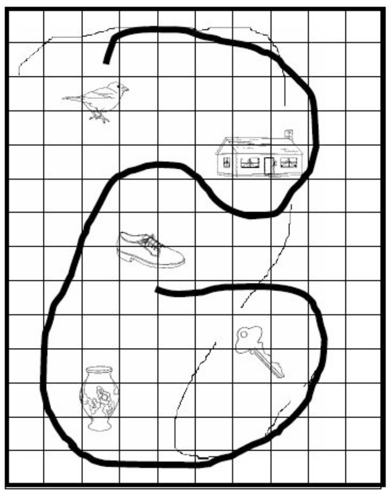
Affect ratings: For the second perception task, 80 participants were asked to listen to the second calibration sentence, and to rate expressed positive affect on on a 1 to 7 scale (1 = very unhappy, 2 = unhappy, 3 = somewhat unhappy, 4 = neither happy nor unhappy, 5 = somewhat happy, 6 = happy, 7 = very happy). The 200 calibration sentences were split into four sub-groups to reduce the amount of materials presented for listening, and played back to four groups of 20 listeners. The rating interface was created using a PsyScope script (Cohen, MacWhinney, Flatt & Provost, 1993) and run on a Macintosh G4 iBook.

4.2.3 Measurements

Communicative effectiveness was measured as a listener's ability to follow the speaker's instructions. To this end, I computed the deviation of the drawn route from the original route in cm^2 on a grid overlaid over the map (see Figure 4.2).[3] Each description was played to three listeners, and the median deviation of the three deviation scores from those three drawings served as my measure of communicative effectiveness.

Reported speaker affect was determined by computing the difference between the reported ratings for just the two mood adjectives 'happy' and 'sad' on the BMIS given by each speaker, which ranged from 1 (definitely do not feel) to 4 (definitely feel). These ratings reflect the certainty with which a speaker is experiencing these affective states.

Figure 4.2: Example for the grid with the original route overlaid over a map drawing



Perceived speaker affect was assessed by computing the means of the listeners' ratings of

[3]See Appendix B for examples of good and bad route drawings.

the speakers' happiness on the 1 to 7 scale in the second perception task.[4] Since the whole corpus of 200 utterances had been divided into 4 sub-sets and each sub-set was rated by a different group of listeners for reasons of feasibility, z-scores of the mean ratings within each sub-group of speakers served as a measure of perceived happiness.

Linguistic content: One of the main challenges of the current study was to find a suitable overall measure of linguistic content without introducing a large number of potentially arbitrary coding criteria. I therefore opted for a global similarity measure that would compare all descriptions to the best description, i.e. the description that was associated with the smallest route deviation (see Appendix B). This composite measure of linguistic content was obtained using Latent Semantic Analysis (LSA) (Landauer & Dumais, 1997), which computes the similarity of a text passage to a target passage. The analysis itself was performed using a web interface[5] where one text is put in as the required target passage followed by any number of texts that are to be compared to this target passage. While the program runs automatically, requiring no further input from the researcher, the process behind it is highly sophisticated. Rather than the researcher choosing which of the many possible factors might determine the goodness of a description (i.e. the frequency of landmark mention, or use of directional terms like 'left' and 'right' or 'north' and 'south' etc.), LSA makes it possible to operationalise linguistic content of a text passage (in my

[4]While it may often seem inappropriate to treat this kind of rating data as interval data I felt it was justified in this case because the rating categories were formulated in a way to suggest a linear progression, and because the means were normally distributed.

[5]freely available on the University of Colorado at Boulder website http://lsa.colorado.edu/

case the whole route description) as a vector in a multidimensional space. This multidimensional space is obtained by keeping track of the co-occurrences of words in passages, creating a vector representation in hyperspace with a dimensionality determined by the number of words and passages, and using single-value decomposition to reduce the dimensionality of this space. Similarity is then computed as the cosine of the angle between the vectors of each text passage and the target passage in reduced dimensionality space, which results in values between -1 and 1, and constitutes a single measure of the quality of the linguistic content for each description.

Acoustic parameters: I performed acoustic analyses measuring mean F_0, F_0 range, speech rate, jitter (irregularities in the fundamental frequency) and shimmer (irregularities in the intensity), as well as formant frequencies. All these parameters have been implicated in the expression of positive affect in previous studies (Bachorowski & Owren, 1995; Banse & Scherer, 1996; cf. Section 3.2.2). An acoustic parameter not included in the analysis was amplitude, due to technical problems with keeping the distance to the microphone constant in such a large sample of speakers. All acoustic parameters, except speech rate and F_0 range, were measured on the vowel /i:/ in the *please* of the second calibration sentence. This vowel was suitably sustained by all speakers due to its phrase position, and it was also the only vowel produced in a similar way in all of the dialects spoken by the participants. F_0 was measured over the steady state phase of the vowel. For the F_0 range measurements, the procedure of Patterson & Ladd (1999) was incorporated, which takes into account *"clearly defined linguistic targets in speech"* (p. 1169). Their suggested estimate of F_0 range is the difference between the second F_0 peak (M) and the penultimate F_0 valley (L) of an intonational phrase, converted into semitones. F_0, the formant frequencies, jitter and shimmer were measured using algorithms implemented in PRAAT (Boersma & Weenink, 2005). For measuring jitter and shimmer, two algorithms were chosen which were suitable for the relatively small number of pulses (30 on average) in the vowel /i:/ in please. To determine jitter, an algorithm was used measuring the relative average perturbation (rap), the average absolute difference between a period and the average of it and its two neighbours, divided by the average period. To determine shimmer, the three-point amplitude perturbation quotient (apq3) was used, the average absolute difference between the amplitude of a period and the average of the amplitude of its two neighbours, divided by the average amplitude. In addition, to avoid spurious measurements for F_0 and the formant frequencies, each recording was visually inspected by a trained phonetician. Pitch values that could be clearly identified as artefacts, e.g. octave jumps or measurements on voiceless portions of the speech signal, were excluded from further consideration. Instead, measurements were taken from adjacent, non-spurious portions of the speech signal. Formant values were verified by visual inspection of the sonagram. Formant tracker values that deviated from the formant pattern identified in the sonagram were discarded as well. Finally, the duration of the second calibration sentence was used as an estimate for speech rate. While this is a coarse measure it seemed justified due to the shortness of the sentence. Word repetitions, filled and unfilled pauses, which were extremely rare given the shortness of the utterance, were cut from the signal before duration was measured.

4.3 Results

Table 4.1 shows the means and standard deviations for reported and perceived happiness, median route deviation, LSA-based similarity to the best description, and the acoustic parameters for women and men separately, as well as the results of t-tests comparing all the variables across genders. The first important finding was that women women reported greater certainty in experiencing positive affect, $t(198) = 3.0$, $p < .001$, were judged as sounding happier, $t(198) = 4.1$, $p < .001$, and produced descriptions that resulted in more accurate route drawings than men, $t\,(198) = -2.1$, $p < .01$. As expected, gender differences were also found in F_0, F1, and F2 (all p's < .001).

Table 4.2 shows correlations between reported happiness, perceived happiness, linguistic content, and communicative effectiveness. Linguistic content as measured by the LSA-based similarity score was negatively associated with median route deviation. This correlation describes the finding that the better a description the easier it is for listeners to follow the instruction. Interestingly, there was no correlation between reported happiness and linguistic content. This does not necessarily imply that speaker affect does not influence how much speakers optimise the content of their descriptions; it merely shows that the relatively coarse-grained measure of linguistic content employed in this study did not capture such an association, if it exists. There was also no correlation between reported happiness and perceived happiness suggesting that there might not be a one-to-one mapping between experienced emotional state and the affect expressed vocally by the speaker, and that listeners may tend to misinterpret vocal cues as indicative of underlying

Table 4.1: Means and standard deviations (in parentheses) for reported and perceived happiness, median route deviation, LSA-based similarity and acoustic parameters for women (N=112) and men (N = 88)

	Women	Men
Reported happiness**	1.62 (1.19)	1.06 (1.45)
Perceived happiness (z-score)**	0.24 (0.93)	0.31 (0.99)
Median route deviation*, cm^2	158.7 (67.8)	179.4 (73.9)
LSA-based similarity to best description	0.73 (0.12)	0.70 (0.13)
F_0**, Hz	237 (33)	128 (24)
F_0 range, semitones	4.7 (2.7)	4.5 (2.0)
Speech rate duration, ms	2,037 (325)	1,989 (265)
Jitter, %	0.52 (0.63)	0.61 (0.77)
Shimmer, %	2.0 (1.5)	2.2 (2.0)
F1 /i:/ in *please***, Hz	370 (61)	300 (28)
F2 /i:/ in *please***, Hz	2,352 (274)	1,941 (148)

Asterisks represent significant gender differences as indicated by 2-tailed t-tests.

*p $<$ 0.01; **p $<$ 0.001.

emotional states (for a more detailed discussion see Biersack & Kempe, 2005b). This finding highlights once again that the link between production and perception of vocal affect is by no means a case of one-to-one mapping (see Section 3.2.4). The positive correlation between perceived mood and median route deviation with a p-value of .1 suggests a trend which fell short of significance for listeners to produce more accurate route drawings when speakers sounded happier.

Table 4.2: Correlation between reported happiness, perceived happiness, linguistic content (LSA-based similarity) and communicative effectiveness (median route deviation)

	Perceived happiness	LSA-based similarity	Median route deviation
Reported happiness	0.09	0.07	0.00
Perceived happiness		-0.05	-0.11
LSA-based similarity			-0.44*
*p < 0.01			

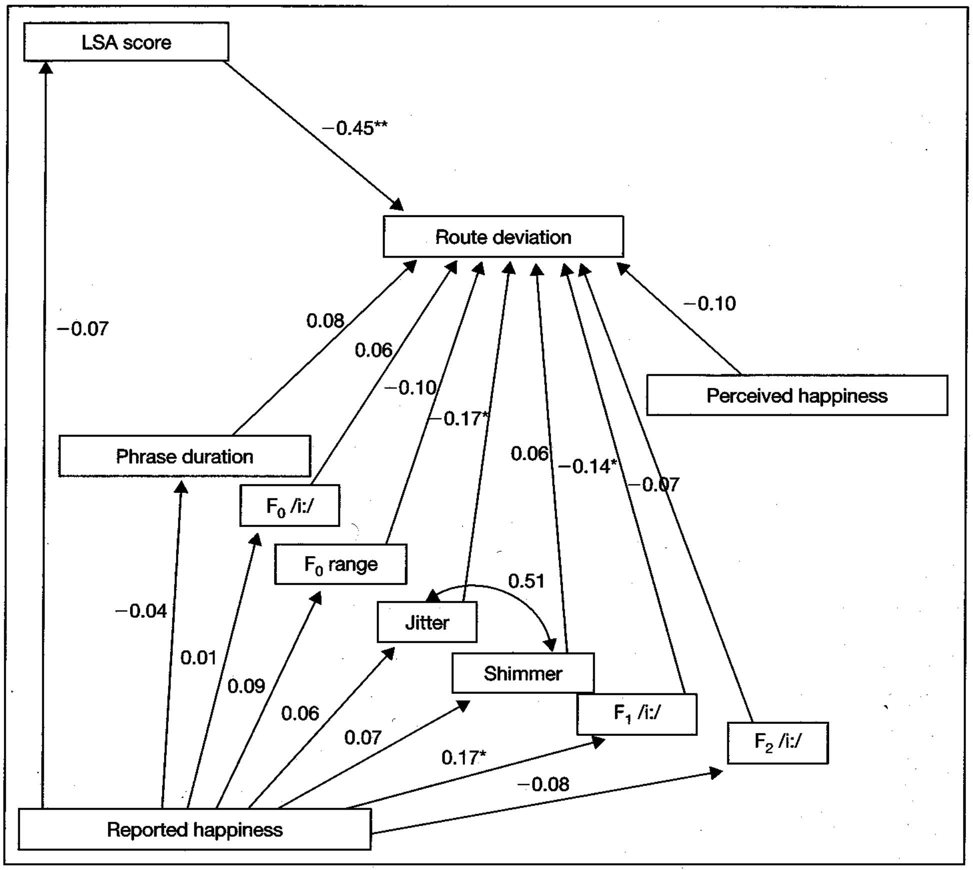
To determine whether perceived happiness had an effect on communicative effectiveness over and above linguistic content a stepwise regression analysis was performed with median route deviation as dependent variable. When gender, coded as a dummy variable, was entered at a first step, it accounted for 14.5% of variance ($F(1,198) = 4.2$, $p < .05$). Quality of linguistic content (measured as LSA-based similarity), entered at the next step, accounted for another 6.6% of variance ($F(2,197) = 27.6$, $p < .001$), confirming that linguistic content is crucial for successful comprehension. Most importantly, when perceived happiness was entered at the final step, it accounted for another 2% of variance over and above linguistic content ($F(3,195) = 19.2$, $p < .001$). This effect, albeit small, was significant, suggesting that perceived emotional valence of the speakers' utterances made an independent contribution to communicative effectiveness such that happy sounding speech facilitates listener comprehension.

Given that there was no relationship between linguistic content and reported happiness, the next set of analyses focused on the effects of the vocal cues associated with positive affect expression. As outlined in the introduction of this study, I set out to explore whether the various vocal cues had a direct effect on communicative effectiveness (Direct Route Model), whether their effect was mediated by the perception of speaker happiness (Mediated Route Model), or whether vocal cues affected both perceived happiness as well as communicative effectiveness independently (Dual Route Model) (see Section 3.3). To recapitulate briefly, the Direct Route Model assumes that reported happiness affects vocal cues which, in turn, directly impact communicative effectiveness as well as perceived affect. In this model, perceived positive affect has no independent effect on communicative effectiveness. The Mediated Route Model assumes that the degree of positive affect experienced by the speaker affects vocal cues, which, in turn, contribute to the impression of happiness. It is the impression of speaker happiness that is assumed to improve communicative effectiveness by eliciting listener attention, perhaps because happy sounding speakers seem more appealing and trustworthy. Finally, the Dual Route Model combines both pathways by assuming that vocal cues may have a direct effect on communicative effectiveness but also an effect mediated through the perception of speaker happiness.

To compare these three models, a series of path analyses was performed. Path analysis is an extension of multiple regression analysis, and is used to fit the correlation matrix of different causal models to the data to compare the goodness of fit of these models. The fit indices used were the χ^2 statistic, Akaike's Information Criterion (AIC), Bentler's Comparative Fit Index (CFI), and Bollen's Incremental Fit Index (IFI). The χ^2 statistic is a measure of 'badness of fit', and tests whether a model's covariance structure is significantly different from the observed covariance matrix. Thus, a significant χ^2 indicates bad model fit. The other indicators are 'goodness of fit' indicators. AIC takes into account the complexity of the model and penalises models with higher complexity, i.e. lower degrees of freedom due to more constraints in the model. Lower AIC values indicate better fit. CFI and IFI compare the existing model with a null model which assumes the variables in the model are uncorrelated (i.e. all covariances are set to 0). Higher CFI and IFI values indicate better fit. The CFI varies between 0 and 1 with values above .9 being considered a good fit. The IFI can exceed 1. The acoustic parameters included in these models were all obtained from the second calibration sentence as described above. In order to eliminate gender effects from the model, all measurements were transformed into z-scores within each gender.

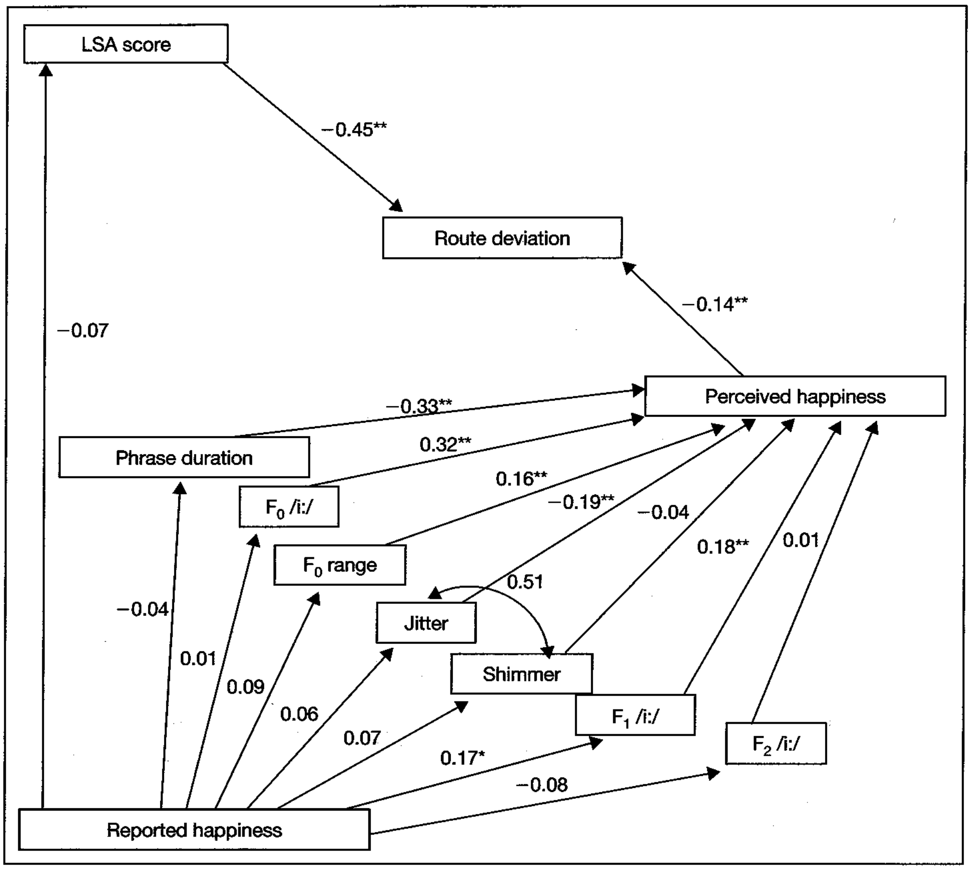
Figures 4.3, 4.4 and 4.5 show the Direct Route, the Mediated Route, and the Dual Route Models, respectively, including path coefficients. The fit indices of the models are listed in Table 4.3. As can be seen from this table, the χ^2 statistic was significant for the Direct Route Model indicating a poor fit. The Mediated Route and the Dual Route Model

*Figure 4.3: The estimated Direct Route Model. Coefficients shown are standardised path coefficients. *p < 0.05; **p < 0.01*



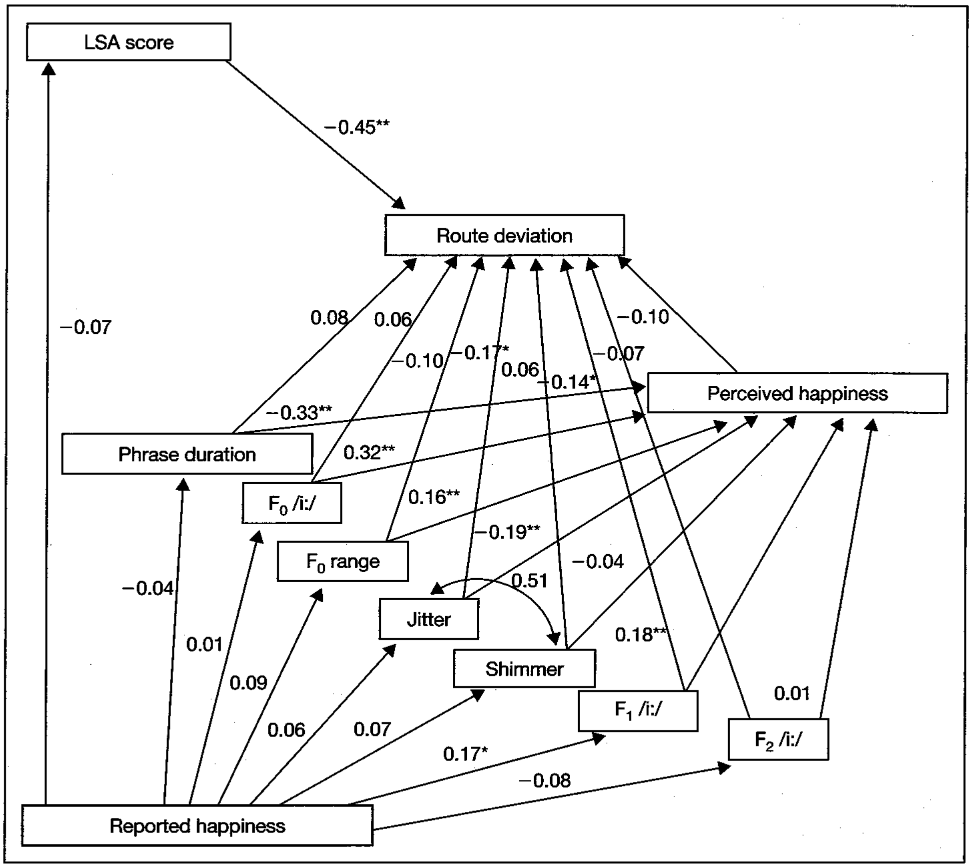
both produced non-significant χ^2 statistic indicating that the models did not significantly deviate from the observed covariance matrix (Mediated Route: χ^2 (df = 37, N = 200) = 45.5, p = .16; Dual Route: χ^2 (df = 30, N = 200) = 26.5, p = .65). A χ^2-difference test was performed to compare the Mediated and the Dual Route Model, which showed that the Dual Route Model fit was significantly better, χ^2_{diff} (df_{diff} = 7) = 18.9, p < .01. In addition, all other fit indices were higher for the Dual Route Model. This suggests that the Dual Route Model provides the best fit to the data.

*Figure 4.4: The estimated Mediated Route Model. Coefficients shown are standardised path coefficients. *p < 0.05; **p < 0.01*



The significant path coefficients suggest the following picture: Degree of positive emotion experienced by the speaker is related to an increase in F1 which is the only vocal cue in this study to directly reflect underlying speaker affect. F1 together with F_0, F_0 range, speech rate, and jitter contribute to the perceivers' impression of speaker happiness. There was a positive relationship between F1, F_0, F_0 range, and speech rate and perceived speaker happiness. The higher F1 and F_0, the wider the F_0 range, and the faster the speech rate (i.e., the shorter the duration of the calibration sentences), the more were listeners

*Figure 4.5: The estimated Dual Route Model. Coefficients shown are standardised path coefficients. *p < 0.05; **p < 0.01*



inclined to attribute happiness to the speaker. These relationships are in line with what has been described in previous research on the vocal expression of happiness in acted speech (Banse & Scherer, 1996; Scherer, 2003). There was also a negative relationship between jitter and perceived happiness, a finding that corroborates results obtained by Bachorowski & Owren (1995). Note that listeners made their judgments about speaker happiness based on many more vocal cues than are actually related to the speakers' self-reports of happiness. I will return to this point in the discussion.

Table 4.3: Fit indices for the three models

Model	d.f.	χ^2	AIC	CFI	IFI
Direct Route	37	99.6*	179.6	0.63	0.67
Mediated Route	37	45.5	125.4	0.95	0.96
Dual Route	30	26.5	120.5	1.0	1.02

*$p < 0.001$.

In the Dual Route Model, as in the two other models, higher LSA-based similarity scores were associated with smaller route deviations confirming that, not surprisingly, better linguistic content improves communicative effectiveness. Moreover, all the models suggest that perceived emotional valence affects communicative effectiveness independently of linguistic content, although the path coefficient of .14 linking perceived happiness and route deviation was only significant in the Mediated Route Model, and fell short of significance in the other two models ($p = .16$). In addition, F1 and jitter had a direct effect on communicative effectiveness: The lower F1 and the smaller the perturbations in the fundamental frequency the greater the route deviation. Thus, in the best fitting model, vocal cues affect communicative effectiveness independently of linguistic content, both directly and mediated through perceived happiness.

4.4 Discussion

This study explored whether and how speaker affect contributes to the success of communication. I investigated a large sample of speakers engaged in a simple Map Task, and measured communicative effectiveness as the accuracy of the route drawings produced by participants listening to the speakers' route descriptions. To assess the current emotional state of the speakers, a short mood questionnaire, the BMIS (Mayer & Gaschke, 1988) was administered. Finally, invariant parts of the speakers' descriptions were submitted to acoustic analysis, and rated with respect to perceived happiness, which served as an indicator for the vocal expression of speaker affect.

The results showed that the vocal cues linked to positive affect expression made a small but significant independent contribution to communicative effectiveness over and above the effect of the linguistic content of the message. Thus, it appears that affect-related acoustic-phonetic variation can modulate how well a message is received regardless of the content of the message. In this study, there was no direct effect of speaker affect on the linguistic content of the message. However, this null effect may be due to the limited linguistic variation in the relatively simple route descriptions, or due to the coarse measure of content which was obtained through Latent Semantic Analysis. In future research, a range of more sensitive referential communication tasks should be employed, and more detailed measures of linguistic structure and content should be explored to capture potential links between speaker affect and linguistic content.

The observed effect of vocal cues on communicative effectiveness may be due to two mechanisms operating in conjunction: First, the vocal cues associated with positive affect expression may reflect acoustic-phonetic modifications which influence the listener directly, by facilitating comprehension or capturing attention, e.g. through more accurate articulation or a stronger emphasis on important parts of the message (cf. Section 3.2.2). This model, termed the Direct Route Model showed poor fit to the data. Secondly, the vocal cues may contribute to the listeners' perception of happiness of the speaker, which, in turn, may contribute to better comprehension of the message, perhaps by signalling trustworthiness of the speaker. This model was termed the Mediated Route Model, and showed much better fit to the data. It suggests that happier sounding speakers may be better able to elicit listener attention or create rapport, in short, to affect social attitudes towards the speaker in a way that is beneficial for communication (cf. Section 3.1.2). The best model fit was obtained for the Dual Route Model, a model in which vocal cues exert both a direct effect on communicative effectiveness as well as one mediated through perceived happiness of the speaker. Thus, vocal cues associated with positive affect expression can facilitate comprehension directly, but also contribute to the listeners' attitudes towards the speaker. The two cues that were most prominent in this regard were F1 and jitter. Note also that F1 was the only cue in this study linking reported happiness of the speaker with perceived happiness as well as communicative effectiveness. Raise in F1 is, among other things, associated with the opening of the oral cavity and a spreading of the lips, which, in turn, can be related to increased vocal effort, i.e. loudness (Traunmüller, 1988; Geumann, 2001), or smiling (Tartter, 1980; Robson & Mackenzie Beck, 1999). Thus, the path from reported happiness to F1 suggests that happier speakers may have produced higher first formants because they tended to speak louder. Geumann (2001) has shown that articulatory and acoustic modification of loud speech is associated with higher F1 values either because amplified jaw movements and an increase in fundamental frequency require raised F1 to preserve vowel quality or because a raised F1 may simply be a consequence of the wider opening of the oral cavity and the resulting lower jaw position. It is therefore possible that the relationship between reported speaker happiness, F1, and communicative effectiveness is just an indirect reflection of increased loudness[6], which, in turn, may enhance perceptual salience thereby facilitating comprehension. Happier speakers may also have produced higher first formants because they tended to smile more. Again, to the extent that smiling is associated with an opening of the oral cavity it will result in raised formants (Tartter, 1980; Tartter & Brown, 1994; Robson & Mackenzie Beck, 1999), even though an increase in F2 was not found in this study. F1, then, appears to have a two-fold effect: It may reflect increased loudness which facilitates comprehension and communicative effectiveness directly, and it may contribute to the impression of speaker happiness by inducing the perception of a smiling speaker who may be perceived as a more attention-eliciting and trusted source of information.

The second vocal cue that had a direct effect on communicative effectiveness was jitter, a measure of the degree of perturbation in fundamental frequency. Interestingly, the effect of jitter was the opposite for communicative effectiveness and for perceived happiness: While jitter was positively correlated with communicative effectiveness (as assessed by route deviation), it was negatively correlated with perceived happiness. Thus,

[6]Note that due to technical difficulties, it was not possible to measure amplitude directly.

greater perturbations in frequency were associated with higher communicative effectiveness and lower perceived speaker happiness. There was no relationship between reported happiness and jitter. It is not quite clear why higher jitter contributes directly to improved comprehension. The negative correlation between jitter and perceived happiness is in accordance with findings by Bachorowski & Owren (1995), who showed that low perturbations were linked to positive affect in a laboratory mood-induction procedure. At the same time, Johnstone & Scherer (1999) who also employed mood induction, report a positive correlation between jitter and speaker happiness, and a negative one with anxiousness and tension. It is possible that the happiness examined by Bachorowski & Owren (1995) and the current study is qualitatively different to the happiness examined by Johnstone & Scherer (1999), resulting in the type of mismatches in reported correlations addressed in Section 3.2.2. At this point, a reconciliation of these conflicting findings, as well as a better understanding of the relationship between perturbations and comprehension will have to await further studies.

Interestingly, in all three models, F1 was the only reliable indicator of speaker affect, i.e. the only parameter linked to both reported and perceived happiness. The other vocal cues used by the listeners for the perception of speaker happiness were not indicative of the self-reported affective state of the speaker. This suggests that there is only a weak relationship between vocal cues of affect encoded by speakers, and vocal cues used for decoding of affect by listeners (see Section 3.2.4 and cf. Biersack & Kempe, 2005b). As discussed in Section 3.1.2, vocal cues are not just epiphenomena of physiological changes associated with certain emotions, and, thus, not necessarily veridical indicators of speaker states (Russell et al., 2003). Apparently, listeners are not very successful in inferring underlying affect from the voice of unfamiliar speakers, and may tend to over-interpret vocal cues which they believe to express a certain emotional state.

Another finding worth mentioning was that women exhibited higher reported and perceived happiness than men, as well as higher communicative effectiveness. These gender differences, which were not related to the main focus of the study, should be corroborated in clinical and social-psychological research before speculating on their origin. Still, I would like to suggest that they may, in part, be due to lower social desirability to report and express positive affect in the male participants, mainly young Scottish undergraduates.

Given that this study was exploratory in nature, the conclusions drawn here are tentative. Further experimental studies will have to replicate the observed associations, specifically the contribution of formant frequencies and perturbations to communicative effectiveness. It will, moreover, be important to employ mood induction procedures in order to directly manipulate speaker affect in verbal communication.

Three avenues of future research on the relationship between affect expression and communicative effectiveness seem most promising in the context of the present results: First, further studies could explore the prosodic effects of vocal cues of positive affect expression to determine whether such cues can indeed amplify useful prosodic information. This will be attempted in the second study of this thesis (Chapter 5). Secondly, direct effects of speaker affect on language production should be further studied, particularly in the area of semantic and pragmatic aspects, for example, in view of whether 'happy' speakers are more cooperative by avoiding ambiguity and taking the listener's perspective into account. Thirdly, the social-psychological mechanisms that link vocal affect expression with impression formation could be investigated, specifically with respect to perceived speaker attributes like competence, likeability and trustworthiness.

Despite the many open questions, the approach taken in this interdisciplinary study constitutes an ecologically valid first step towards a fuller exploration of the relationship between affect and communication, as it was attempted to take into consideration the entire chain, from the emotional state experienced by the speaker to the communicative effect on the listener. This approach is very much in line with the Brunswikian Lens Model suggested by Scherer (2003) for the exploration of vocal affect expression (see Section 3.2.1), with the important addition of effects of the linguistic content of the message. The findings suggest that the relationship between affect and communication is a complex one, and that there are multiple routes by which speaker affect can influence the effectiveness of verbal communication.

This study was exploratory in nature. The next study will continue to examine the impact of positive affect in a more controlled fashion with respect to affect elicitation and the criteria for cues deemed crucial for the informativeness of the message.

Chapter 5

Study 2: Vocal Affect Expression in Child-Directed Speech – Does Vocal Affect Facilitate Prosodic Disambiguation?

5.1 Introduction

After examining communicative effectiveness on a relatively broad scale in the first study, namely as level of performance of listeners in a Map Task, this study examines communicative effectiveness in terms of a more specific communication problem: the disambiguation of syntactically ambiguous sentences. I specifically wanted to explore whether genuine affect expression can amplify relevant prosodic information. This approach is a continuation of the exploratory research undertaken in Study 1 (cf. p. 60). However, while in Study 1 the emphasis was on aspects of comprehension, Study 2 examines more carefully aspects of production in the communication process.

The question of whether vocal correlates of positive affect that manifest themselves in prosodic modifications have an effect on other prosodic variation touches on issues of the interaction between affective and linguistic prosody, about which there is an ongoing debate in the literature. While some believe that affective and linguistic prosody are intimately linked (cf. for example, Seddoh, 2002), others have tried to show that they are independent systems (cf. Baum & Pell, 1999; Wong, 2002; Ziegler, 2003; McRoberts et al., 1995). Linguistic prosody is generally described as a means of clarifying the structure and content of a message by an adequate placement of intonation, stresses and pauses Price, Ostendorf, Shattuck-Hufnagel & Fong, 1991. Affective prosody, on the other hand, can be seen as the nonverbal means of expressing emotions (cf., for example, Bostanov & Kotchoubey, 2004). Study 1 has demonstrated effects of emotion on affective prosody. However, since both aspects of prosody utilise the same vocal means they are inevitably interlinked. This study can further illuminate the nature of this interaction by helping us understand if and in what ways the emotional effects on affective prosody also affect linguistic prosody.

To control positive affect in a within-subjects design, child-directed speech (henceforth CDS) was contrasted with adult-directed speech (henceforth ADS) in this study.

This took advantage of the fact that interaction with their own baby serves as a powerful inducer of positive mood in mothers (see Section 5.1.1). A referential communication task with four conditions was created: mothers addressing their child, mothers addressing an adult, non-mothers addressing an imaginary child and non-mothers addressing an imaginary adult. It was hypothesised that prosodic disambiguation should be most efficient in the highly emotionally charged CDS register when produced by mothers interacting with their child, less efficient in speech addressed to an imaginary child, and least efficient in the adult-directed conditions. Before outlining the design of the study, a brief overview over research on CDS as an affective speech register is given.

5.1.1 Child-Directed Speech (CDS)

Child-directed speech denotes the speech register that speakers and in particular parents use to address a child of language acquisition age. To emphasise the age of the addressed child, and with that qualitative differences in linguistic abilities, it can be appropriate to distinguish between CDS for speech addressed to children with substantial linguistic skills (over 24 months of age) and IDS (infant-directed speech) for speech addressed to pre-linguistic children (under 24 months of age), although this distinction has been blurred in the literature. The current study deals with CDS, as only children from the age of 24 months could plausibly take part in a referential communication task.

To date, research on CDS has mainly focussed on the speech of mothers to their children. CDS is characterised by raised pitch, wider pitch range, exaggerated prosody, hyperarticulation, slower speech rate, and reduced linguistic complexity, and is assumed to constitute an adaptation that permits mothers to control the child's arousal, to elicit the child's attention (Fernald, 1994), and to maintain interaction over physical distances (Falk, 2004). There are a number of studies which have demonstrated that CDS contains many features that facilitate the task of language learning by providing cues for word boundaries (Batchelder, 2002; Kempe et al., 2005) and for syntactic constituents (Morgan & Demuth, 1996), as well as by simplifying and regularising morpho-syntax (Kempe et al., 2003; Dabrowska, 2006). Raised pitch has been interpreted as an adjustment to the immature auditory system of the child (Fernald, 1994); hyperarticulation has been shown to facilitate phoneme discrimination (Liu, Kuhl & Tsao, 2003). Slower speech rate as well as reduced linguistic complexity are generally assumed to facilitate comprehension on the side of the child. CDS can even facilitate the acquisition of aspects of style and group membership and has, for example, been shown to serve as a template for the sociolect the child is growing up to learn, e.g. emphasising certain phonological features representative of the ambient sociolect (Foulkes, Docherty & Watt, 2005).

Two recent studies (Singh et al., 2002; Trainor, Austin & Desjardins, 2000) have suggested that the prosodic characteristics of CDS, and even more so the characteristics of IDS, resemble those associated with the vocal expression of positive affect, namely higher pitch, wider pitch range, as well as higher acoustic energy, with only speech rate being described as slower in CDS compared to speech expressing positive affect (cf. Scherer, 2003). Trainor et al. (2000) examined the acoustic characteristics of affective speech directed to adults and to children, and found very little differences between the registers. Singh et al. (2002) demonstrated that when vocal affect is controlled for, infants prefer ‘happy’ speech regardless of speech register. Based on these findings, Singh et al. suggest that infants’ preference for IDS may, in essence, be a preference for ‘happy’ speech. The observation that IDS is speech expressing positive affect is in line with views that communication with young infants is predominantly affective in nature. There is growing evidence that affective communication is crucial in early child development (Locke, 2001). For example, IDS of depressed mothers, which exhibits attenuated prosodic features as well as sub-optimal tuning of affect-salient speech to the infant’s needs (Herrera, Reissland & Shepherd, 2004), tends to weaken associative learning in infants (Kaplan et al., 1999). Affective responsiveness of mothers has been shown to be crucial for children’s linguistic and cognitive development (Nicely et al., 1999).

With respect to mothers, it is reasonable to assume that vocal affect expression in IDS and CDS reflects positive emotions experienced towards the child. Indeed, Nitschke, Nelson, Rusch, Fox, Oakes & Davidson (2004) have shown that when mothers were exposed to pictures of their own children, these elicited positive feelings the strength of which correlated with levels of activation in the orbito-frontal cortex, and surpassed the positive feelings elicited by pictures of non-kin infants. A similar study by Bartels & Zeki (2004) showed the same results and compared neural patterns of maternal love to those of romantic love, concluding that maternal love might indeed be the predecessor of romantic love or at the very least occupy similar neural networks. Thus, if infants are powerful positive mood inducers for their mothers, the resulting strong positive affect might be responsible for the production of the salient prosodic features of CDS.

Less is known about CDS features of non-kin interlocutors who may not have an affective bond with the child. Jacobson, Boersma, Fields & Olson (1983) explored the speech of parents and non-parents to an imaginary child, a non-kin infant, and a non-kin toddler, and found that pitch and pitch range increased when speaking to an imaginary child suggesting that CDS of non-kin interlocutors is similar to parental CDS. Pitch and pitch range increased even more when the adults were speaking to a real child. Jacobson et al. (1983) attributed this additional increase to the effect of feedback from the child. This feedback might involve an emotional component which promotes vocal affect expression in the adult speakers thereby amplifying the vocal characteristics of CDS. For example, facial features of the child, such as cuteness (or 'babyfacedness'), have been shown to intensify CDS (Zebrowitz, Brownlow & Olson, 1992).

The current study will make use of two aspects of CDS that have been established in the literature: first, that children serve as a strong mood inducer in their mothers, and secondly that CDS can be elicited in non-kins and, with the actual child-interlocutor removed, as the same speech register but without the affective component. Interaction with a small child will thus be treated as a means of inducing mood, while the task of addressing an imaginary child serves as a control condition to examine potential didactic aspects of CDS that are not a by-product of positive affect but rather reflective of the implicit knowledge of what CDS should accomplish for an interlocutor with limited linguistic capacity.

The features of CDS described in the literature are crucial to the central hypothesis of this study: If CDS is assumed to be particularly beneficial for the child's comprehension and acquisition of language, the ambiguous sentences spoken by mothers to their children should be the ones best disambiguated by the prosodic means available. This effect should be stronger than in non-mothers addressing an imaginary child if the affective component is responsible for more informative prosody, and certainly stronger than in mothers and non-mothers addressing adults.

Furthermore, an examination of CDS in terms of its prosodic functionality could have important theoretical implications. The notion of prosodic cues playing a crucial role in CDS as a means for the child to assess syntactic structure has been advanced by the Prosodic Bootstrapping Hypothesis. The term 'Prosodic Bootstrapping' (Fisher & Tokura, 1996; Morgan & Demuth, 1996) refers to the assumption that since language is organised on many different levels (syntax, semantics, phonology, etc), a learner of this language has to acquire knowledge of each level to master the other levels, and that this learning in one domain (i.e. syntax) is facilitated by information from another domain (i.e. prosody). Findings in this line of research suggest that the vocal cues of CDS contribute crucially to the acquisition of this knowledge by, for example, highlighting syntactic structure (Morgan et al 1987), word meaning (Fisher 1996, Gleitman 1990) and word-level phonology (Gafos & Brent, 1994).

However, opponents of the Prosodic Bootstrapping Hypothesis, most notably Fernald & McRoberts (1996), criticise that most findings brought forward as a corroboration of the Bootstrapping Hypothesis are methodically flawed or at the very least not sufficiently replicated. In their view, the crux of many studies is an overestimation of the reliability of prosodic cues. A very instructive example they give is that studies presenting putative evidence for pause structure as a cue to syntax often fail to take into account all pauses occurring in relevant speech samples, and instead only test whether or not pauses coincide with the syntactic boundaries in question. Fernald & McRoberts (1996) argue that this procedure does not acknowledge the fact that if pauses appeared also frequently elsewhere, as a number of studies suggest (Boomer & Dittmann, 1962; Hawkins, 1971; Henderson, Goldman-Eisler & Skarbek, 1966), this would trigger a substantial number of false alarms, which in turn would cast doubt on the instructional value of those pauses in the first place.

Even though most of the relevant research is based upon determining co-occurrences and alignments of cues in speech samples, the Prosodic Bootstrapping Hypothesis is essentially a learner's theory, with very little empirical evidence for the motivation on the part of the speaker. Most notably, there is a lack of within-subjects studies exploring behavioural differences in prosodic cuing that might occur with different types of listener demands, namely when addressing a child as opposed to an adult.

With a concrete communication problem under examination, namely the disambiguation of ambiguous sentences, as proposed for this study, it will be possible to scrutinise the central ideas of the Prosodic Bootstrapping Hypothesis with respect to production more thoroughly: Are prosodic cues generally more pronounced in CDS? Is there a greater tendency in CDS to align prosody and syntax compared to ADS? And finally: Is positive affect responsible for prosodic bootstrapping? This leads to the design of this study.

5.1.2 Study Design

In a recent paper, Snedeker & Trueswell (2003) demonstrated that when the referential context supports various interpretations of a syntactically ambiguous sentence such as *Tap the frog with the flower*, speakers are able to disambiguate employing prosody. In their study, the most prominent prosodic cue was the duration of the pause before the prepositional phrase (henceforth noun pause) in relation to the duration of the pause after the verb (henceforth verb pause), indicating whether the prepositional phrase beginning with ‘with’ is attached to the verb ‘tap’ or the definite noun phrase ‘the frog’. The former would correspond to an interpretation of ‘the flower’ as an instrument, the latter would render ‘the flower’ a modifier (for example, to distinguish between two otherwise identical frogs). The pause structure serves to conjoin or separate sentence constituents, respectively, a mechanism which, if found to be in place, would be in close agreement with the mechanisms predicted by the Prosodic Bootstrapping Hypothesis. Snedeker & Trueswell (2003) embedded such sentences in a referential communication task where the objects and tools of the test sentences were on physical display.

Here, this paradigm was adapted to explore whether the described prosodic disambiguation is more pronounced in speakers experiencing positive affect. To this end, positive affect was induced by introducing mothers’ own children as interlocutors. As children between the age of 2 and 3.5 years were to be involved, Snedeker & Trueswell’s procedure had to be further simplified: Speaker and listener were not visually separated, only very easily recognisable objects were employed and only one type of action was used, the latter resulting in all sentences being of the type *‘Touch the ... with the ...’*. This also maximised comparability for subsequent acoustic measurements.

5.2 Part 1: Mothers Talking to Their Child and an Adult

5.2.1 Method

Production: Prosodic Disambiguation Task

Participants:

Twenty-four mothers, mean age 35 years (ranging from 23 years to 46 years) and their infants (N=24), aged between 2;0 and 3;8, took part in the task. The mothers, all native speakers of English, received course credit or a reimbursement of GBP 10.00.

Materials:

Two arrays with soft toys were set up, one array with a pig, a frog, a duck with a flower, a cat, a cat with a spoon, and a spoon (see Figure 5.1a), the second array with a snake, a fish, a horse with a spoon, a dog, a dog with a flower, and a flower (see Figure 5.1b). The presence of two spoons and two flowers, respectively, provided the necessary ambiguous context. For the instruction sentences to be used by the speakers, four booklets were created, with one sentence and an accompanying picture per page. The pictures indicated the actions listeners, in this case children, should be asked to perform. Each booklet contained 12 pages; five pages for one array, and five pages for the other. The two remaining pages, page one and six, only contained a picture of the array, with no action indicated (cf.

Figure 5.1: The two toys arrays

(a) The cat array (b) The dog array

Figure 5.1). This was to familiarise participants with the contents of the relevant array and to mark a divide between the first and second array. For each array there was one critical, i.e. ambiguous, sentence and four ‘filler’ sentences. Critical sentences were sentences that were ambiguous in terms of their syntactic structure, as described above, and that at the same time were ambiguous in terms of the context. Given the setup of the objects only the cat and the dog appeared in critical sentences. The relevant sentences were:

1. *Touch the cat with the spoon.*

   (ambiguous context, instrument or modifier interpretation)

2. *Touch the dog with the flower.*

   (ambiguous context, instrument or modifier interpretation)

Out of context, these sentences could either instruct the listener which soft toy to touch (henceforth modifier interpretation, resolving the prepositional phrase as a nominal phrase attachment, see Figure 5.2a), or instruct the listener to use the flower/spoon as an instrument (henceforth instrument interpretation, resolving the prepositional phrase as a verb phrase attachment, see Figure 5.2b). As the arrays contained two flowers and two spoons respectively, not only the syntactic structure but also the context was ambiguous. The target interpretation was only indicated to the speaker by the corresponding picture in the booklet. The filler sentences were either ambiguous in syntactic structure but not ambiguous in context, or unambiguous in both syntactic structure and context. Two filler

Figure 5.2: Two interpretations for production: "Touch the cat with the spoon"

(a) Modifier interpretation

(b) Instrument interpretation

sentences per booklet were ambiguous in syntactic structure but not ambiguous in context, thus serving as unambiguous controls.

3. *Touch the fish with the flower.*

   (unambiguous context, instrument interpretation)

4. *Touch the duck with the flower.*

   (unambiguous context, modifier interpretation)

5. *Touch the horse with the spoon.*

   (unambiguous context, modifier interpretation)

6. *Touch the frog with the spoon.*

   (unambiguous context, instrument interpretation)

Their occurrence was counterbalanced across booklets. Another six sentences per booklet were unambiguous in both syntactic structure and context. These filler sentences were the same for all booklets (for the combination of the sentences in each booklet see Appendix D).

Procedure:

Mothers were told that they were taking part in a pilot study aimed at exploring how children follow instructions and from which age they are able to do so. The two arrays with the toys were introduced as part of a game where the mother would give instructions and the child would manipulate the toys according to the instructions.

For each trial one experimenter was present who oversaw the recordings and instructed the participants as well as one confederate who acted as the adult addressee. Both the experimenter and the confederate were unknown to the mothers. Mothers were fitted with a JHS MUD-805 uni-directional headset microphone which was connected to an iRiver iHP-120, a multi-purpose mp3-player that allows uncompressed wave-format recordings. Sound files were recorded at a sampling rate of 44.1 kHz.

The order of addressee was counterbalanced. Half of the mothers first addressed their child and then the adult confederate, the other half first addressed the adult confederate and then their child. Mothers were randomly assigned to one of those groups. The booklet was the same for both conditions. Mothers who first addressed the adult were told that this was a trial round to get used to the sentences and the set up of the game. They were also told that there was no need to 'pretend' that they were talking to their child. Mothers who first addressed their child were told that playing the same game with the adult was only necessary to verify that the sentences were not too difficult in general. They, too, were told that there was no need to 'pretend' that they were talking to their child.

When mothers addressed the adult, the child was taken to an adjacent room, so as to keep the child out of sight and to avoid dual-tasking or even child-directed speech. When talking to their child, mothers were asked to sit them on the lap, and within easy reach of the array of toys. Having the children sit on their mothers' lap ensured, on the one hand, a relatively relaxed atmosphere for both mother and child and avoided, on the other hand, non-verbal communication, in particular use of gaze. Mothers were asked to not use their hands to indicate actions or to point out toys. However, once mothers had tried to instruct the child using just the pre-defined sentence, they were encouraged to carry out the action together with the child if the child had not managed to do so alone. This was to minimise frustration for both the mother and the child.

To avoid any disambiguation of the critical sentences by syntactic means and to prevent paraphrasing, mothers were asked to always produce the instruction sentences exactly as printed, and to not alter them. The contextual ambiguity of some of the sentences was not pointed out to the mothers. Mothers were, however, instructed to carefully study each picture before producing the accompanying sentence. The whole session took approximately 15 minutes.

<u>*Acoustic Measurements:*</u>

To ensure comparability, only sentences of the type *'Touch the* [N1] *with the* [N2]*'* were analysed (in bold print in Appendix D). Thus, eight sentences of each mother were analysed, four in each condition (CDS vs. ADS), and, with 24 mothers, 192 sentences altogether. All of these sentences were syntactically ambiguous but only *'Touch the cat with the spoon'* and *'Touch the dog with the flower'* were ambiguous also in context. Similarly to Snedeker & Trueswell (2003), a series of acoustic analyses was performed, measuring:

- the verb pause duration in ms

  (*Touch* <P_1> *the cat with the flower*)

- the noun pause duration in ms

  (*Touch the cat* <P_2> *with the flower*)

- F_0 (in Hz) on the steady-state portion of the first noun (N1) vowel

  (*Touch the* ***dog*** *with the spoon*)

- F_0 (in Hz) on the steady-state portion of the vowel in 'with'

- F_0 (in Hz) on the steady-state portion of the second noun (N2) vowel

  (*Touch the dog with the* ***spoon***)

Pauses were identified through visual inspection of the sonagram and the acoustic signal. The F_0 values were gathered using algorithms implemented in PRAAT (Boersma &

Weenink, 2005), with the settings recommended for female voices in the PRAAT manual. The for the measurements relevant portions of vowels and diphthongs were determined through auditory as well as visual inspection. For vowels the criterion was the presence of a distinct formant pattern in the sonagram with at least the first two formants clearly identifiable, as well as a quasi-stationary behaviour of formant values. For diphthongs only the strength of the formants was used as a criterion. To avoid spurious measurements of F_0, each recording was visually inspected by a trained phonetician, and in case of spurious values that appeared to be the result of artefacts, measurements were taken from adjacent, non-spurious portions of the speech signal.

Duration and pitch measurements should reveal whether speakers made use of the prosodic means available for disambiguating modifier and instrument interpretations. A longer noun pause was expected to mark a separation between the nominal phrase and a subsequent prepositional phrase favouring an instrument interpretation (cf. Price et al., 1991; Snedeker & Trueswell, 2003). A longer first pause, on the other hand, was expected to close ranks between the nominal phrase and a subsequent prepositional attachment favouring a modifier interpretation (cf. Price et al., 1991; Snedeker & Trueswell, 2003). In the data of Snedeker & Trueswell (2003) an instrument interpretation was also favoured by listeners if ‘with’ was emphasised by means of a pitch accent. While this is not a pattern commonly found in the literature it is important to keep in mind that most findings on prosodic disambiguation are based on forced-choiced perception tests with stimuli either created from synthetic speech (Lehiste, 1973) or from material where speakers have been specifically instructed to use certain prosodic patterns (for a discussion see Allbritton, McKoon & Ratcliff, 1996; Fox Tree & Meijer, 2000) or have been re-recorded until they did so by themselves (Price et al., 1991). Snedeker & Trueswell (2003) provide evidence from data of uninstructed speaker-listener dyads engaging in communication where genuine ambiguous contexts occurred. They carried out an analysis of a large number of prosodic parameters which was not hypothesis-driven with respect to relevant prosodic cues, and emphasis on 'with' was one of the most prominent prosodic cues they found. Thus, this pattern, which might not occur in other contexts, may be part of the strategies speakers adopt when confronted with a communication problem, that normally they could resolve in other ways, for example, by means of syntactic rephrasing. Finally, pitch measurements should also help to determine the degree of 'child-directedness' of speech in different conditions, with higher pitch values indicating more child-directedness (cf. Kitamura & Burnham, 1996, 2003; Kitamura, Thanavvishuth, Burnham & Luksaneeyanawin, 2002; Ryan, 1991; Trainor & Desjardins, 2002).

Perception: Comprehension Task

Participants

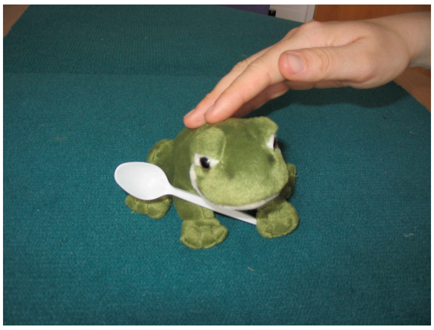
Twenty-four undergraduate students, 8 men and 16 women, aged 19 to 50 years (mean age 27 years) took part in the comprehension tasks, a forced choice test to determine whether listeners inferred an instrument or a modifier interpretation for each of the 192 sentences.

Materials

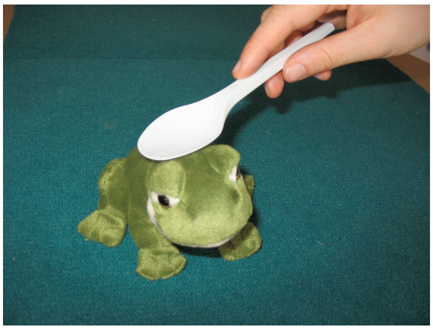
A forced choice task to test listeners' sentence interpretation was created using Eprime

(Psychology Software Tools, Pittsburgh). For each of 192 trials, two pictures were prepared, both showing the same soft toy but one time in a modifier (see Figure 5.3a) and one time in an instrument (see Figure 5.3b) context. Unlike in the production task only the relevant soft toy was pictured, not the whole array.

Figure 5.3: Two choices for comprehension: "Touch the frog with the spoon"



(a) Modifier interpretation



(b) Instrument interpretation

Procedure

The wave format sound files with the recordings of the mothers were transferred onto a Highgrade PC laptop on which the comprehension task was run. Stimuli were presented and responses were recorded via the Eprime interface. For each trial one sentence was played back via Beyerdynamic DT 250 high-quality headphones while simultaneously two pictures were displayed on the laptop screen (see Figure 5.3). Listeners were asked to indicate via keyboard feedback which interpretation they thought was appropriate for the sentence they just had heard. The items were the same for each participant, with the order of items randomised. It was also randomised for each item whether the modifier picture occurred on the left or right side of the screen. Each participant completed a short trial session and then performed the forced choice for the 192 sentences. The task took 20 to 25 minutes in total with two breaks at equal intervals.

Measurements

Listeners' sentence interpretation was assessed as the percentage of instrument choices in both the instrument and modifier conditions. The higher the instrument choice in the instrument condition, and the lower the instrument choice in the modifier condition, the better the disambiguation of a given sentence is.

Perception: Affect Ratings

Participants

Three men and 7 women, aged 21-49 years (mean age 29 years) participated in the second perception task designed to obtain perceived affect ratings.

Materials

For the affect rating interface, a PsyScope script was created (Cohen et al., 1993), containing all 192 sentences.

Procedure

The rating interface was presented on a Macintosh G4 iBook and stimuli were played back via Beyerdynamic DT 250 high-quality headphones. Listeners were asked to rate "how happy each speaker sounded" on a 1 to 7 scale (1 = very unhappy, 2 = unhappy, 3 = somewhat unhappy, 4 = neither happy nor unhappy, 5 = somewhat happy, 6 = happy, 7 = very happy) using the keyboard. Responses were recorded as part of the PsyScope script.

Measurements

Perceived speaker affect was assessed by computing the means of the listeners' ratings of speakers' positive affect on the 1 to 7 scale. Means were calculated across three conditions: addressee (child-directed vs. adult-directed), sentence type (instrument vs. modifier) and context ambiguity (ambiguous vs. unambiguous).

5.2.2 Results

Affect Ratings

First, to test whether mothers showed the expected positive affect when talking to their children it was examined how mothers' display of affect had been rated by listeners. Table 5.1 shows the mean affect ratings per condition.[1]

The mean affect ratings were submitted to a 2 (ambiguity: ambiguous vs. unambiguous) x 2 (addressee: adult vs. child) x 2 (sentence type: instrument vs. modifier) ANOVA, which yielded a main effect of addressee, $F(1,23) = 25.1$, $p < .001$. This clearly shows that the child-directed speech was rated as happier (mean rating 4.4) than the adult-directed speech (mean rating 3.8). No other effects reached significance.

A multiple regression analysis was performed with the verb and noun pause durations

[1] The means were again normally distributed (cf. footnote on p. 44).

Table 5.1: Mean affect ratings and standard deviations as a function of context ambiguity, addressee, and sentence type for sentences produced by the mothers

	Adult-directed		Child-directed	
	instrument	modifier	instrument	modifier
ambiguous	3.9 (0.8)	3.9 (0.8)	4.4 (0.8)	4.5 (0.9)
unambiguous	3.7 (0.9)	3.7 (0.9)	4.2 (0.9)	4.4 (1.0)

and the F_0 measures on the first noun, the 'with' and the second noun as predictor variables to explore which of these cues contributed to the perception of positive affect of the speaker. The model accounted for 53.6 % of the variance ($F(5,179) = 41.3$, $p < .001$) and indicated that F_0 on the vowel of the 'with' and the second noun contributed significantly to the perception of positive affect (F_0 'with': $\beta = .19$, $p < .01$; F_0 N2: $\beta = .68$, $p < .001$). This indicates that higher F_0, in particular at the end of the sentence, contributes mainly to the impression of expressed happiness of a speaker. Interestingly, durational measures, which might up to a certain degree reflect effects of speech rate, did not contribute to the impression of positive affect, presumably because their contribution was overridden by the substantial fluctuations in F_0.

Prosodic Disambiguation Cues

Pauses:

The analysis of the temporal prosodic pattern followed Snedeker & Trueswell (2003) who found that the most prominent prosodic cues used for disambiguating the sentences were the durations of the pause after the verb (verb pause) and the duration of the pause after the first noun (noun pause). In Experiment 1 of their study, in which speakers were aware of the context ambiguity, the average durations of the verb and the noun pause were 63 ms vs. 302 ms in the instrument condition, and 146 ms vs. 27 ms in the modifier condition, respectively, leading to reliable disambiguation of the two intended meanings by the listeners. The focus of my analyses of prosodic cues was therefore on these two pause durations. Two independent coders performed the pause duration measurements. The correlation between the two measurements was $r = .72$, $p < .001$ (N= 192) for the verb pauses, and $r = .96$, $p < .001$ (N=192) for the noun pauses. Given that these correlations are sufficiently high, the following analyses will present the measurements obtained by the first coder.

Table 5.2 presents the means and standard deviations for the verb and noun pause durations. Pauses over 500 ms were deemed hesitations and substituted by condition means. This affected only one value. A 2 (context ambiguity: ambiguous vs. unambiguous) x 2 (addressee: adult vs. child) x 2 (sentence type: instrument vs. modifier) within-subjects ANOVA was performed. Given that my adaptation of the procedure used by Snedeker & Trueswell (2003) to the use with children drastically reduced the variability in items to only two sentences for the ambiguous context, and four sentences for the unambiguous context, analyses by items were neither possible nor necessary. Therefore, only analyses by subjects will be reported. For the verb pauses, the ANOVA revealed a main effect of addressee, $F(1,23) = 5.2$, $p < .05$, which indicates that pause durations increased in the

Table 5.2: Means and standard deviations of the pause durations (in ms) produced by the mothers as a function of condition and addressee

	Ambiguous context			
	Adult-directed		Child-directed	
	instrument	modifier	instrument	modifier
Verb pause	56 (15)	58 (23)	62 (27)	58 (24)
Noun pause	95 (74)	60 (40)	132 (124)	97 (73)
	Unambiguous context			
Verb pause	54 (21)	55 (19)	69 (56)	65 (27)
Noun pause	103 (73)	111 (116)	150 (116)	179 (253)

child-directed condition, maybe due to an overall decrease of speech rate. However, pause duration can only be treated as an indicator of speech rate , not as a direct measure of it. No other effects were significant. For the noun pauses, the analyses also revealed a main effect of addressee, $F(1,23) = 6.4$, $p < .05$, again indicating longer pause durations in the child-directed condition. The effects of context ambiguity, $F(1,23) = 2.7$, $p = .1$, and the effect of sentence type, $F(1,23) = 2.9$, $p = .1$, fell short of significance suggesting that there was a trend for noun pauses to be shorter when the context was ambiguous compared to the unambiguous context, and when speakers intended a modifier interpretation of the sentence compared to an instrument condition. Thus, speakers showed the expected tendency towards producing shorter pause preceding the prepositional phrase in the modifier condition. So far, it is not clear why noun pause durations tended to be shorter in the ambiguous sentences.

These analyses do not indicate that the mothers amplified prosodic disambiguation cues in the child-directed condition, at least not with respect to pause durations. Neither were the verb pauses longer in the modifier sentences nor were the noun pauses longer in the instrument sentences addressed to a child. This null-effect was unexpected. However, it is possible that pause durations were not the relevant cue in a situation where mothers had to speak to their small children. It is possible that in this situation, speakers focus on pitch modifications because increased pitch and pitch excursions are among the most salient features of CDS. Indeed, in the Snedeker & Trueswell (2003) study, the other prominent prosodic cue was a pitch accent on the preposition 'with'. In order to check this possibility, and also to examine to what extent speakers increased pitch in CDS, the fundamental frequency was measured on selected parts of the sentence.

Fundamental Frequency (F_0):

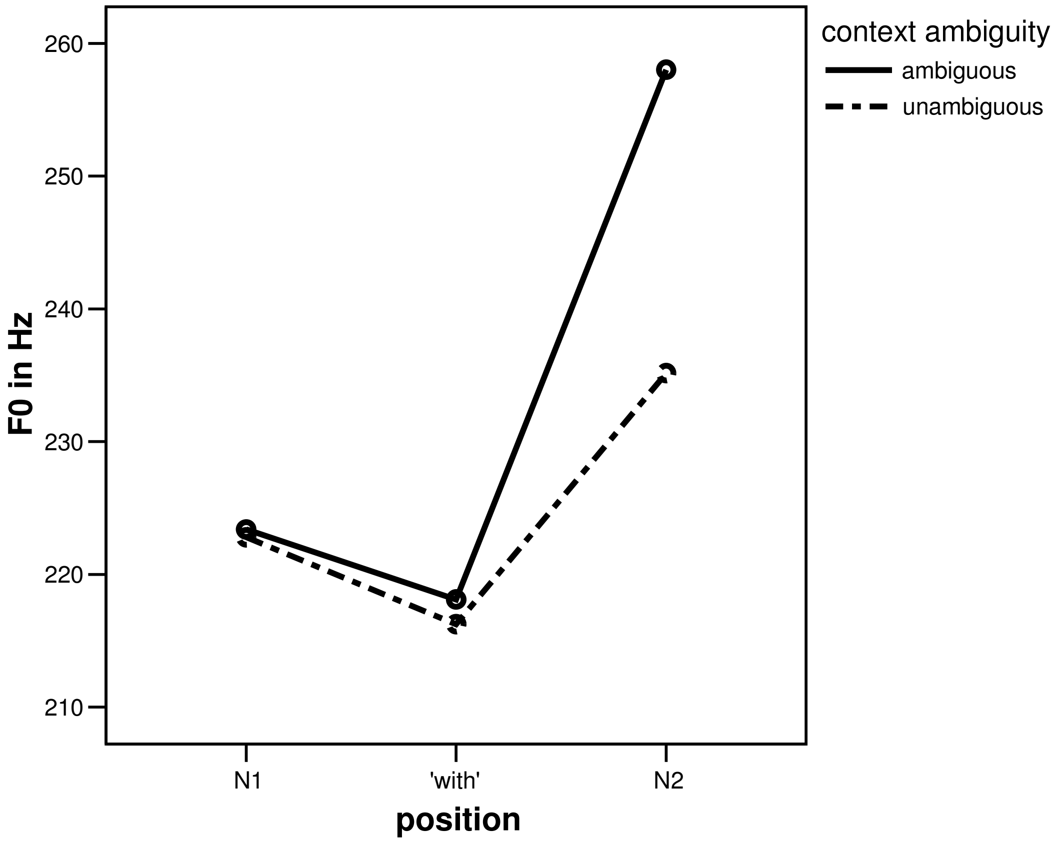
F_0 was measured on the stressed vowels of the first noun, the preposition 'with', and the second noun. If speakers produced a pitch accent on 'with', F_0 at this site would be expected to be higher compared with the preceding noun. In addition, the continuation of the frequency contour to the second noun will reveal whether the participants not only generally increased their pitch in the CDS condition, but also produced the CDS-typical final rise (Fernald, 1994).

A 3 (position: N1, 'with', N2) x 2 (context ambiguity: ambiguous vs. unambiguous) x

2 (addressee: adult vs. child) x 2 (sentence type: instrument vs. modifier) within-subjects ANOVA was performed on F_0, which was measured over the steady-state portion of the vowel in each word. For the bi-syllabic N2 'flower', the fundamental frequency was measured for the entire triphthong. Seven values were missing due to whispering or creakiness of the voice, and were replaced by condition means. The analysis revealed a main effect of register, $F(1,23) = 64.1$, $p < .001$ indicating that, as expected, F_0 was higher in the child-directed condition. There was also a main effect of position, $F(2,46) = 5.25$, $p < .05$. Posthoc tests using Fisher's LSD revealed a significant increase in F_0 from the preposition 'with' (217 Hz) to N2 (247 Hz). The difference between N1 and 'with' as well as between N1 and N2 failed to reach significance. This pattern indicates a final rise in pitch in the sentence. This main effect was specified by two 2-way interactions: the interaction between position and context ambiguity, $F(2,46) = 4.2$, $p < .05$, and the interaction between position and addressee, $F(2,46) = 4.5$, $p < .05$.

To qualify the interaction between position and context ambiguity, a 3 (position) x 2 (addressee) x 2 (sentence type) ANOVA was performed for the ambiguous and unambiguous sentences separately. For the ambiguous sentences, this analysis revealed a main effect of addressee, $F(1,23) = 42.3$, $p < .001$, again confirming the increase in F_0 in the child-directed condition. The effect of position was also significant, $F(2,46) = 10.5$, $p < .05$. Post-hoc tests revealed a significant increase in F_0 from N1 to N2, $p < .01$, and from 'with' to N2, $p < .01$. For the unambiguous sentences, the ANOVA revealed a main effect of addressee, $F(1,23) = 28.0$, $p < .001$. The interaction between position and addressee fell short of significance, $F(2,46) = 3.0$, $p = .08$. This suggests that in the unambiguous sentences, the tendency towards a final rise was somewhat stronger (see Figure 5.4).

Figure 5.4: F_0 produced by the mothers on the vowel of the first noun, the preposition 'with', and the second noun in ambiguous and unambiguous contexts.



To qualify the interaction between position and addressee, a 2 (context ambiguity) x 2 (position) x 2 (sentence type) ANOVA was performed for the adult-directed and the child-directed sentences separately. For the adult-directed sentences, the main effect of context ambiguity fell short of significance, $F(1,23) = 4.1$, $p = .06$, indicating a tendency towards slightly higher F_0 in the ambiguous sentences. For the child-directed sentences, this analysis revealed a main effect of position, $F(2,46) = 7.1$, $p < .05$. Post-hoc tests showed that there was a significant rise in F_0 from N1 to N2 ($p < .05$) and from 'with' to N2, $p < .01$. This indicates that the increase in F_0 in CDS was mainly due to a strong rise on the last noun. In addition, the interaction between position and sentence type fell short of significance, $F(2,46) = 3.3$, $p = .06$. A 3 (position) x 2 (context ambiguity) ANOVA for the instrument sentences revealed a main effect of position, $F(2,46) = 4.9$, $p < .05$, which was due to an F_0 increase from 'with' to N2, $p < .01$. For the modifier sentences, the significant effect of position, $F(2,46) = 7.2$, $p < .01$, was due to significant F_0 increases both from N1 to N2 ($p < .05$) as well as from 'with' to N2, $p < .01$. This pattern of results suggests that the final rise was slightly stronger in the modifier condition. Given that there were no F_0 differences between N1 and 'with' in any of the conditions, the occurrence of a pitch accent on the preposition 'with' can be firmly ruled out.

Taken together, the F_0 analyses indicated that mothers increased their F_0 when addressing their children, particularly on the second noun at the end of the sentence. This pattern corresponds to what would be expected in CDS (Fernald & Mazzie, 1991; Ryan, 1991). Interestingly, this final-rise intonation pattern was also more pronounced in the ambiguous sentences suggesting that the mothers were attempting to use increases in F_0 to disambiguate the two interpretations when the context was ambiguous. However, rather than placing a pitch accent on the preposition, as the speakers in the Snedeker & Trueswell (2003) study did, the mothers amplified the child-directed final-rise pattern. This may be due to a general tendency to emphasise content words as opposed to function words (Morgan, Shi & Allopena, 1996; Blanc, Dodane & Dominey, 2004). Combined with the pause analyses, the results suggest that, overall, mothers did not disambiguate more clearly using prosody when addressing their child compared to an adult. The following analyses will examine how the prosodic strategies adopted by the mothers affected comprehension.

Comprehension Task

The final analysis examined listeners' interpretation of the 192 sentences. Table 5.3 presents the percent of instrument choices as a function of context ambiguity, addressee, and sentence type.

Table 5.3: Percent of instrument choices (with standard deviations in parentheses) as a function of context ambiguity, addressee, and sentence type

	Adult-directed		Child-directed	
	instrument	modifier	instrument	modifier
ambiguous	51.9 (14.3)	45.0 (11.5)	54.3 (18.2)	51.6 (18.4)
unambiguous	55.7 (10.3)	46.0 (16.2)	59.9 (17.2)	53.5 (14.9)

A 2 (context ambiguity: ambiguous vs. unambiguous) x 2 (addressee: adult vs. child) x 2 (sentence type : instrument vs. modifier) within-subjects ANOVA on the percent of instrument choices revealed a main effect of sentence type, $F(1,23) = 5.7$, $p < .05$, indicating that there were more instrument choices in the instrument sentences (55.5%) than in the modifier sentences (49.0%). There was also a main effect of context ambiguity, $F(1,23) = 10.2$, $p < .01$, indicating that instrument choices were more frequent in unambiguous (53.8%) than ambiguous (50.7%) sentences. Finally, the main effect of addressee, $F(1,23)$

= 4.4, $p < .05$, was due to more instrument choices in the child-directed condition (54.8%) than in the adult-directed condition (49%). Note that in general, comprehension of the interpretation intended by the speakers was quite poor with performance generally being around chance (50%). In order to see in which condition listeners were able to identify the intended interpretation reliably, one-sample t-tests against 50% were performed for all conditions. These tests showed that only in the unambiguous adult-directed and in the unambiguous child-directed instrument sentences were listeners able to reliably identify the intended instrument interpretation, and in the ambiguous adult-directed modifier sentences were they able to identify the intended modifier interpretation. In all other conditions performance was indeed at chance. This shows that the prosodic cues the mothers provided were generally not sufficiently clear to the listeners to enable them to identify the intended interpretation. Moreover, there was no indication for greater clarity in the disambiguating prosodic cues in the child-directed conditions, as this should have resulted in an interaction between addressee and sentence type.

5.2.3 Discussion of Part 1

This part of the study explored whether speakers, when in a state of positive affect, employ prosodic means of disambiguation more effectively. To this end, an ecologically valid means of controlling affect was introduced, namely interaction with a mother's own child as opposed to interaction with an unknown adult. The hypothesis was that mothers, when speaking affectively to their child, disambiguate better than when speaking to the adult, where an affective component can be ruled out.

The results showed that mothers sounded happier when talking to their child as opposed to when talking to the adult, confirming that CDS did indeed express positive affect. This effect was mainly carried by a higher F_0 on different parts of the examined sentences, most notably on the sentence-final second noun. This is in line with findings that associate higher F_0 with 'happy speech' (Bachorowski & Owren, 1995; Murray & Arnott, 1993; Scherer, 2003) as well as with findings that associate a final rise of F_0 with CDS (Fernald & Mazzie, 1991; Ryan, 1991). While F_0 is expected to behave similarly in CDS and in speech expressing positive affect, speech rate should change in the opposite direction: slower than normal in CDS (Fernald, 1994) and faster than normal in 'happy speech' (Scherer, 2003). With pauses as an indicator of speech rate, this was confirmed by overall longer pauses in the CDS condition. However, with respect to perceived happiness, the effect of increased F_0 apparently overrode the effect of speech rate.

Crucially, no evidence was found for exaggerated disambiguation cues in CDS. While pauses were overall longer in the CDS condition, pause durations were not efficiently adapted to disambiguate. Modifier sentences were not marked by longer verb pauses and neither were instrument sentences by longer noun pauses. While the essential interaction between sentence type and ambiguity did not reach significance, there was, however, a trend for an effect of sentence type. This suggests that mothers were aware of the ambiguity of the sentences. A similar picture emerged with F_0. While overall F_0 increased in CDS, it was not used as a means of emphasising the preposition 'with', which would have been an efficient way of expressing the instrument meaning. Instead, mothers amplified the CDS-typical sentence final F_0 rise in the crucial, context ambiguous sentences, not just when talking to their child, but, to some extent, also when talking to an adult. This again suggests that mothers may have been well aware of the ambiguity of some sentences, but did not use effective means to disambiguate, opting instead for more ‘child-directedness’. As a consequence of mothers’ failure to better disambiguate when talking to their child, listeners had difficulties reliably distinguishing the two interpretations in the child-directed condition.

It will now be interesting to see how non-mothers adapt their prosody to the different addressees. The initial hypothesis was that non-mothers should be less efficient in their attempt to disambiguate when addressing a child than mothers, as they will, without a real child to interact with, not experience the same positive affect. It is, however, also possible that while mothers’ communication strategy was mostly concerned with signalling affect and eliciting attention, non-mothers emphasise more didactic features, based on the implicit knowledge they have of CDS. To maximise comparability between the two parts of the study, the methodology and design were kept the same, wherever possible.

5.3 Part 2: Non-Mothers Talking to an Imaginary Child and Adult

5.3.1 Method

Production: Prosodic Disambiguation Task

Participants

Twenty-four women, mean age 27 years (ranging from 21 years to 42 years), all native speakers of English, participated in the production part of the second part of the study. None of these women were mothers.

Materials

The arrays and booklets were the same as in the previous part of the study.

Procedure

Participants were told that they were recorded giving instructions that would later be used for another study to test a ‘game’ in which listeners would have to follow instructions. They were shown the arrays and told how the game works. Like the mothers, they were fitted with a JHS MUD-805 uni-directional headset microphone which was again connected to an iRiver iHP-120. Sound files were recorded at a sampling rate of 44.1 kHz.

The order of imaginary addressee (adult vs. child) was counterbalanced. Participants were only told about the second type of imaginary addressee after they had finished the recordings for the first type of imaginary addressee. This had proven to be less confusing for speakers in pilot runs. When asked to address an imaginary adult, participants were instructed to imagine speaking to a friend or acquaintance. When asked to address an imaginary child, participants were instructed to imagine speaking to a two- to three-year old they might know or have seen in the department's toddler group.

To avoid any disambiguation of the critical sentences by paraphrasing, participants were asked in both conditions, child-directed and adult-directed, to produce the instruction sentences exactly as printed, and to not alter them. The contextual ambiguity of some of the sentences was not pointed out to participants. Participants were, however, instructed to carefully study each picture before producing the accompanying sentence. Each trial took approximately four minutes.

Measurements

The same acoustic parameters were measured for this part of the study as for the first part of the study, comprising of verb pause duration (in ms), noun pause duration (in ms), F_0 (in Hz) on the first noun, F_0 (in Hz) on 'with' and F_0 (in Hz) on the second noun. As for the mothers, 192 sentences were analysed.

Perception: Comprehension Task

Participants

Twenty-four undergraduate students, 7 men and 17 women, aged 18 to 42 years (mean age 22 years) took part in the comprehension task, a forced choice test to determine listeners' sentence interpretation.

Materials

The same Eprime interface was used for this part of the study as for the first part of the study. One-hundred ninety two trials were prepared in a similar fashion.

Procedure

As with the stimuli of the first part of the study, pictures for the 192 trials were presented on a Highgrade PC laptop, and the recordings of non-mothers' sentences were played back via Beyerdynamic DT 250 high-quality headphones. The task took 20 to 25 minutes in total with two breaks suggested on-screen.

Measurements

Listeners' sentence interpretation was assessed in the same way it was assessed in the first part of the study by means of determining listeners' percentage of instrument choice in relation to intended sentence type.

Perception: Affect Ratings

Participants

Five men and 5 women, aged 23-50 years (mean age 37 years) participated in the second perception task designed to obtain perceived affect ratings.

Materials

As for the first part of study, a PsyScope script was created, presenting recordings of all

192 sentences.

Procedure

The rating interface was again presented on a Macintosh G4 iBook and stimuli were played back via Beyerdynamic DT 250 high-quality headphones. Listeners were, similarly to those in the first part of the study, asked to rate "how happy speakers sounded" on a 1 to 7 scale.

Measurements

Perceived speaker affect was assessed in the same way it was assessed in the first part of the study by computing the means of the listeners' affect ratings on the 1 to 7 scale and for the three conditions addressee, sentence type and context ambiguity.

5.3.2 Results

Affect Ratings

First, it was examined how non-mothers' display of affect had been perceived by listeners. Table 5.4 shows the mean affect ratings for all conditions.

The 2 (ambiguity) x 2 (addressee) x 2 (sentence type) ANOVA on the mean affect ratings yielded a main effect of addressee, $F(1,23) = 4.9$, $p < .05$, which was qualified by an interaction between addressee and sentence type, $F(1,23) = 4.7$, $p < .05$, as well as a main effect of ambiguity, $F(1,23) = 4.3$, $p < .05$. The main effect of addressee indicated that child-directed sentences (mean rating: 4.2) sounded happier than adult-

Table 5.4: Mean affect ratings and standard deviations as a function of context ambiguity, addressee, and sentence type for sentences produced by the non-mothers

	Adult-directed		Child-directed	
	instrument	modifier	instrument	modifier
ambiguous	4.0 (0.7)	4.0 (0.8)	4.2 (1.0)	4.3 (0.8)
unambiguous	4.0 (0.9)	3.8 (0.7)	4.0 (0.7)	4.2 (0.8)

directed sentences (mean rating: 4.0). The interaction suggests that this effect was more pronounced in the modifier sentences. The main effect of context ambiguity was due to sentences in the ambiguous condition sounding slightly happier (mean rating: 4.1) than in the unambiguous condition (mean rating: 4.0).

As with the mothers' sentences, a multiple regression analysis was performed with the verb and noun pause durations and the F_0 measures on the first noun, the 'with' and the second noun as predictor variables to explore which of these cues contributed to the perception of positive affect of the speaker. The model accounted for 25.2% of the variance ($F(5,184) = 12.4$, $p < .001$) and indicated that duration of the two pauses as well as F_0 on the second noun contributed significantly to the perception of positive affect. Note that pause durations were negatively related to the rating of speaker happiness (verb pause $\beta = -.18$, $p < .001$; noun pause: $\beta = -.18$, $p < .001$) while second noun F_0 was positively related to the happiness ratings ($\beta = .35$, $p < .001$).

Shorter pauses are an indicator of higher speech rate, which suggests that when speech does not exhibit dramatic pitch excursions listeners rely more on speech rate when judging speakers' affective states. This may explain why the increase in perceived happiness was not as dramatic as in the mothers: The longer pauses in the CDS condition may have partially cancelled out perceived happiness due to increased F_0 on the last noun.

Prosodic Disambiguation Cues

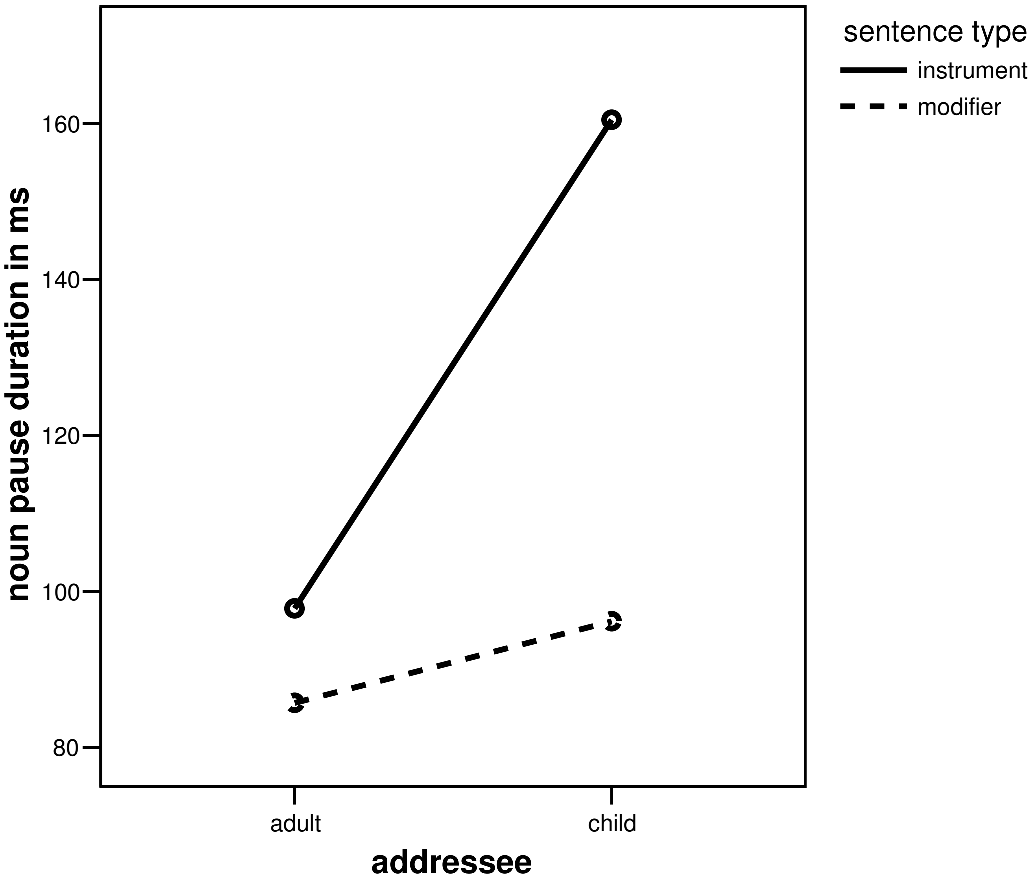
<u>*Pauses:*</u>

Table 5.5 presents the means and standard deviations for the verb and pause durations. One pause duration over 500 ms was deemed a hesitation, and substituted by the condition mean. The 2 (context ambiguity) x 2 (addressee) x 2 (sentence type) within-subjects

Table 5.5: Means and standard deviations of the pause durations (in ms) produced by the non-mothers as a function of condition and addressee

	Ambiguous context			
	Adult-directed		Child-directed	
	instrument	modifier	instrument	modifier
Verb pause	49 (30)	55 (35)	48 (26)	59 (24)
Noun pause	99 (85)	72 (69)	168 (116)	102 (72)
	Unambiguous context			
Verb pause	50 (24)	53 (33)	63 (86)	56 (24)
Noun pause	96 (74)	100 (77)	152 (126)	90 (76)

Figure 5.5: The noun pause duration in CDS and ADS produced by non-mothers in modifier and instrument sentences



ANOVA for the verb pauses did not show any significant effect indicating that the verb pauses did not differ systematically across conditions. For the noun pauses, the analyses revealed a main effect of addressee, $F(1,23) = 8.6$, $p < .01$, indicating longer pause durations in the child-directed condition. There was also a main effect of sentence type, $F(1,23) = 6.3$, $p < .05$, which was specified by an interaction between sentence type and addressee, $F(1,23) = 5.7$, $p < .05$. This interaction is depicted in Figure 5.5, and shows that the difference in noun pause duration between instrument and modifier sentences was significant in the child-directed condition, $F(1,23) = 8.3$, $p < .01$, but not in the adult-directed condition, $F < 1$.

Fundamental Frequency (F_0):

Two F_0 measures were missing due to whispering and creakiness of the voice, and were substituted by condition means. The 3 (position) x 2 (context ambiguity) x 2 (addressee) x 2 (sentence type) within-subjects ANOVA did not show any significant effects. The effect of position, however, fell short of significance, $F(2,46) = 3.4$, $p = .06$. Post-hoc tests using Fisher's LSD revealed a significant decrease in F_0 from N1 (224 Hz) to 'with' (212 Hz, $p < .001$) and N1 to N2 (206 Hz, $p < .05$). This indicates that the non-mothers produced regular, declining intonation contours without any pitch accents or rises.

Comprehension Task

The final analysis examined listeners' comprehension of sentences produced by the non-mothers. Table 5.6 presents the percent of instrument choices as a function of context ambiguity, addressee, and sentence type.

A 2 (context ambiguity: ambiguous vs. unambiguous) x 2 (addressee: adult vs. child) x 2 (sentence type: instrument vs. modifier) within-subjects ANOVA on the percent of instrument choices revealed a main effect of addressee, $F(1,23) = 11.4$, $p < .01$, indicating that there were more instrument choices in the child-directed sentences (57.0%) than in the adult-directed sentences (51.8%). There was also a significant interaction between

Table 5.6: Percent of instrument choices (with standard deviations in parentheses) as a function of context ambiguity, addressee, and sentence type

	Adult-directed		Child-directed	
	instrument	modifier	instrument	modifier
ambiguous	53.3 (10.9)	50.2 (12.8)	60.0 (13.0)	54.2 (11.0)
unambiguous	49.5 (10.4)	54.3 (13.1)	59.2 (11.8)	55.0 (14.7)

sentence type and addressee, $F(1,23) = 5.6$, $p < .05$, due to the fact that the difference between instrument sentences and modifier sentences was more pronounced in the child-directed condition (instrument: 59.4%; modifier: 54.6%) than in the adult-directed condition (instrument: 51.4%; modifier: 52.2%; $F(1,23) < 1$), even though the effect fell short of significance $F(1,23) = 3.1$, $p = .09$). One-sample t-tests against chance revealed that for instrument sentences, instrument interpretations were significantly above chance in the ambiguous and the unambiguous child-directed conditions (both p's < .01). This indicates that listeners were able to identify the intended instrument interpretations reliably in CDS suggesting that speakers provided clearer prosodic cues for the instrument interpretation when addressing an imaginary child.

5.3.3 Discussion of Part 2

This part of the study explored whether non-mothers, when addressing a child, as opposed to an adult, employ prosodic means of disambiguation more effectively, maybe due to tacit knowledge about the needs of an interlocutor with limited linguistic capacity, and despite the affective component of an interaction with a child missing. To remove any potential affect from the CDS of the non-mothers, the task was carried out with imaginary addressees.

The analysis of the affect ratings showed that non-mothers sounded happier in the CDS condition than in the ADS condition. Interestingly, the effect was carried by a negative correlation of the affect ratings with the duration of the two pauses, as well as by a positive correlation with F_0 on the second noun, even though the final rise of F_0 was less dramatic than that observed in the mothers. While the effect of F_0, to some extent, resembles the effect observed in the mothers, the effect of pause duration seems contradictory at first. As the subsequent analysis of the prosodic cues revealed, pauses were longer in the CDS of the non-mothers as well, and while the CDS of non-mothers seems to sound happier, it is shorter pauses that contribute to the perception of positive affect. On closer examination, this might, however, provide an explanation for the relative small increase in perceived happiness in non-mothers' CDS compared to non-mothers' ADS: both shorter pauses, as an indicator of faster speech rate, and increased pitch usually contribute to perception of positive affect (Bachorowski & Owren, 1995; Murray & Arnott, 1993; Scherer, 2003). Thus, even though their moderate pitch increase makes non-mothers sound somewhat happier in CDS, the variability in pause durations, and particularly the relatively strong increase in pause duration in CDS may cancel out the effect of pitch to a certain extent. If non-mothers' pitch increase had been more dramatic, like in the mothers, listeners would have probably disregarded pause durations altogether, as was the case in the mothers data.

The acoustic analysis of the prosodic cues revealed that non-mothers' disambiguation was more pronounced in CDS, which was unexpected in light of the initial hypothesis. While the verb pauses did not significantly differ between conditions, the noun pauses, crucial for the marking of instrument sentences, were used more appropriately in CDS. There was a main effect of sentence type as well as the essential interaction between sentence type and ambiguity. There was, however, no indication of employment of F_0 accents.

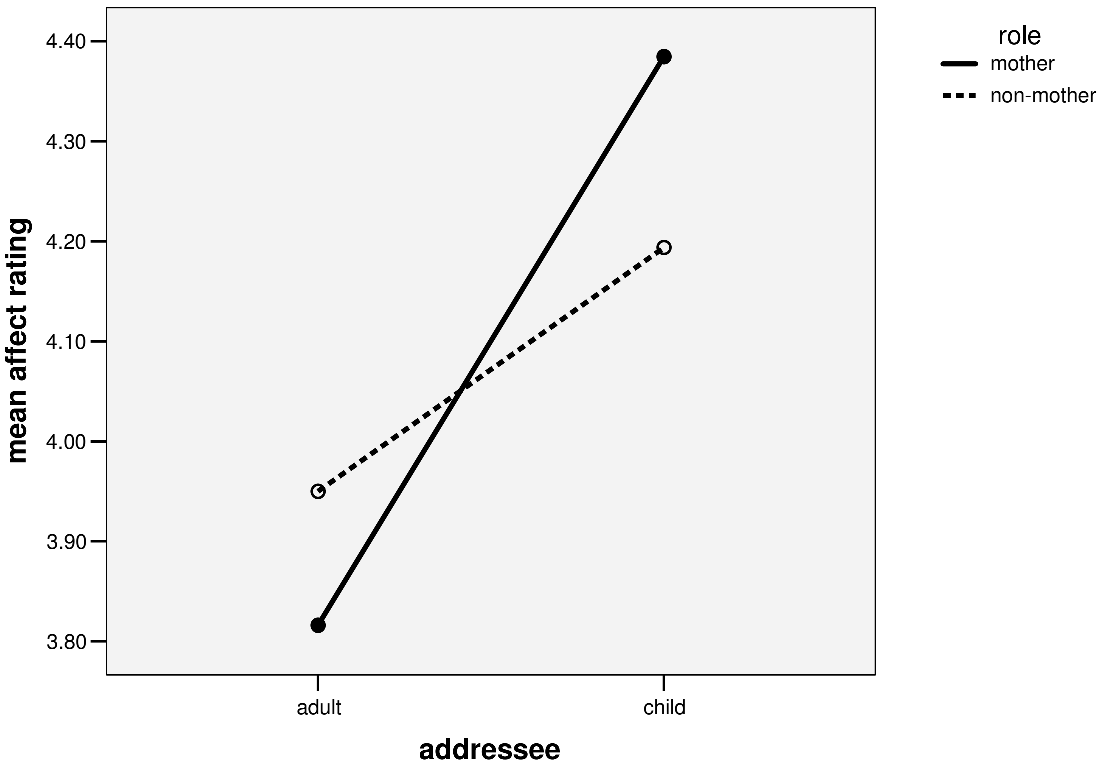
The analysis of the comprehension task revealed that non-mothers' attempts to disambiguate were indeed more successful in the CDS condition than in the ADS condition. Listeners could reliably identify instrument sentences, both in the ambiguous and unambiguous condition. They could, however, not reliably identify modifier sentences, which might be attributed to an insufficient employment of verb pauses by the speakers.

Contrary to the original hypothesis, it was the non-mothers who showed more pronounced disambiguation in CDS. This was evident in the pause durations and in the comprehension data. The overall discussion of Part 1 and 2 of this study will summarise these findings and attempt to find explanations for them.

5.4 Discussion of Part 1 and 2

Both mothers and non-mothers were rated as sounding more happy in the CDS condition. A joint analysis of the affect ratings was conducted to see how the increase in positive affect in the mothers compared to that in the non-mothers. This joint analysis showed a significant interaction between role (mother vs. non mother) and addressee (AD vs. CD) ($F(1,46) = 4.2$, $p < .05$) which is depicted in Figure 5.6 and shows that the mothers displayed more positive affect in their child-directed speech.

Figure 5.6: Mean affect ratings for mothers and non-mothers as a function of addressee



This confirms that mothers speaking to their children express more positive affect than non-mothers speaking to imaginary children.

The analysis of prosodic cues revealed that mothers and non-mothers adopted two very different communication strategies. Contrary to the original hypothesis, it was the non-mothers who showed more pronounced disambiguation in CDS. This was evident in the pause durations and in the comprehension data. In contrast, mothers tended to increase the noun pause irrespective of intended interpretation. While producing prosodic features typical for CDS, as evident in the increased pitch, they appear to fall back onto the instrument interpretation which would be the default interpretation in the absence of lexical or contextual constraints (Spivey-Knowlton & Sedivy, 1995).

The question is whether this makes mothers' CDS less adequate, or whether mothers' communication strategy, if it does not facilitate prosodic disambiguation, has other advantages. There might indeed be several good reasons why mothers did not disambiguate more clearly, when talking to their child than when talking to adults. Mothers might have had, for instance, implicit knowledge that the enhancement of certain prosodic features useful for prosodic disambiguation is inadequate for children. Mothers also had immediate feedback during the task and could adapt their behaviour to the child's reaction. They might have noticed that the task of interpreting the sentences was too difficult for their child, and opted to put more emphasis on attention eliciting features and the expression of positive affect than on disambiguating prosody to at least keep up the 'game'. Increased F_0 in the final part of the sentence rather than just on the preposition 'with' might be such an example, fitting the pattern of more 'child-directedness'.

Non-mothers, on the other hand, did not have immediate feedback from a real child.

One can speculate that throughout the whole task they could assume that clear prosodic disambiguation would help the child, resulting in an amplification of the relevant prosodic cues. This would confirm that speakers have implicit knowledge of the needs of interlocutors with limited linguistic capacity.

The picture that emerges is that speakers who have experience with CDS and/or a strong affective bond with the child may tend to emphasise affective prosody at the expense of linguistic prosody, while speakers without this experience may tend to emphasise prosodic cues to linguistic structure at the expense of affective prosody. The mothers' emphasis on affective prosody may, however, be a more beneficial communicative strategy given that pitch modifications in CDS attract attention and promote learning in children (Thiessen et al., 2005; Nicely et al., 1999). This is in line with the idea that the expression of positive affect as part of CDS benefits learning in children, and in line with studies that have demonstrated that children of depressed mothers, who do not display the same affective prosody in CDS, perform worse in associative learning tasks when maternal vocalisations serve as stimuli (Kaplan et al., 1999; Kaplan, Bachorowski, Smoski & Hudenko, 2002).

More generally, the observation of the different strategies adopted by mothers and non-mothers suggests that affective and linguistic prosody may be independent systems. This contradicts studies that have doubted this dissociation (Seddoh, 2002) and is in line with neurophysiological evidence that there are different neural pathways for producing affective versus linguistic prosody (Baum & Pell, 1999; Wong, 2002; Ziegler, 2003). While the neurophysiological evidence is still equivocal, this study has contributed to a corroboration of this idea with further evidence from speech production (see also McRoberts et al., 1995).

This study also set out to test the Prosodic Bootstrapping Hypothesis. The way the results present themselves here, the Prosodic Bootstrapping Hypothesis cannot be corroborated with respect to speech production. Mothers did show all of the prosodic features typical for CDS, and it seemed that exactly those features were counter-productive for prosodic disambiguation. This was the case for F_0 peaks that occurred sentence-final but not on ‘with’, as well as for pause durations that were overall longer, but not effectively adapted to disambiguate. Thus, it seems that Prosodic Bootstrapping only works well when affective prosody happens to coincide with linguistic prosody. This could be, for example, the case where sentence final peaks in F_0, coincide with focused nouns. Where, however, affective and linguistic prosody are pitched against each other, as was the case in this study, affective prosody, which is prevalent in CDS, fails to facilitate cues crucial for syntactic disambiguation. To my knowledge, the current study is the first study to have experimentally contrasted affective and linguistic prosody.

In future studies, it will be desirable to test two further conditions: mothers speaking to an imaginary child and adult, and non-mothers speaking to a real child and adult. This will help to determine whether mothers, without the affective reaction to their child, retreat to the same strategy the non-mothers displayed or whether they build on the experience they have and continue to emphasise the affective component of CDS. It will be also interesting to see whether non-mothers, when confronted with a real child, employ more affective prosody, maybe compromising on the linguistic prosody like the mothers, or whether they keep their emphasis on the didactic components of CDS. Another possible expansion of this study would be to test children's comprehension of the mothers' and non-mothers' sentences, which was not attempted in the current study due to too much off-task behaviour on part of the children and also because the participating children varied too much in age.

Even though disambiguation tasks are notoriously difficult for speakers as well as listeners (cf. Price et al., 1991), which is reflected in the overall low performance in this study, they should not be discarded as a paradigm for future research. The current study has shown that prosodic disambiguation tasks can be usefully employed to test variation in speaker behaviour as it might occur in connection with, for example, different listener needs and different affective states. Prosodic disambiguation is a clear-cut aspect of communicative effectiveness, and the respective performance criteria are easy to define.

In summary, this study has, in an ecologically valid set-up, shown that the affective components of mothers' CDS are given priority if in conflict with didactic components. This is in line with the notion that the emphasis of mother-child interaction is on creating and maintaining an affective bond and eliciting attention, and that this is what primarily promotes the child's cognitive development (Nicely et al., 1999). The study's results also suggested that affective prosody and linguistic prosody may be separate systems. At the same time the results cast doubt on the validity of the Bootstrapping Hypothesis, in the sense that prosody in CDS does not necessarily and always emphasise syntactic structure.

Taken together, these results can now be used to answer the questions formulated in the introduction of this study. The first question was whether prosodic cues are generally more pronounced in CDS. The results indicate that this is the case in terms of overall pause-duration and F_0 increase. The second question was whether there is a tendency in CDS to align prosody and syntax. The results suggest that this is not the case if affective prosody and linguistic prosody are pitched against each other. The last question was whether positive affect of the speaker may be responsible for prosodic bootstrapping. The results of this study did not find any evidence for this.

The following general discussion of this thesis will aim at reconciling the results of Study 1 and Study 2, put the findings into the context of emotion research and explain their significance.

Chapter 6

General Discussion: The Impact of Affect on Speech Communication

The aim of this thesis was to explore the impact of vocal affect expression on communicative effectiveness. Specifically, I was interested to explore the effect of genuine positive affect. Study 1 is the first study to directly investigate the communicative effect of positive affect expression on the listener. Study 2 is the first to examine experimentally the effect of positive affect expression on the informativeness of prosodic cues to syntactic structure.

A number of factors may lead to the assumption that the experience and display of positive affect can be advantageous in verbal communication. There is evidence that positive affect may lead to increased verbal working memory capacity (Gray, 2001), which, in turn, may improve ways in which speakers plan and produce messages. Furthermore, positive affect may improve accessibility of positive memories of social communication

(Bower, 1981) and help to weigh the cost and benefit of various communication strategies, thus encouraging the speaker to pursue his or her communicative goals (Forgas, 1998b, 1999a). Also, the vocal cues associated with the expression of positive affect may result in enhanced phonetic and prosodic features of the message, which, in turn, may facilitate communication and elicit listener attention. Finally, the signaling function of the vocal expression of positive affect may impact on the listener through the social perception of the speaker as being more trustworthy, reliable, appeasing, etc.

As previous research so far has not directly addressed these factors, the starting point was an exploratory study to test whether there is a correlation between positive affect in the speaker and communicative effectiveness. In this study, Study 1, a referential communication task allowed for a simple quantification of communicative effectiveness: Speakers were asked to provide instructions for reproducing a route on a simple map, and communicative effectiveness was operationalised as the accuracy of the route reproduced by the listeners. This study showed that the expression of positive affect had a small but significant positive effect on communicative effectiveness independent of message content. Thus, in addition to the informational content and the linguistic structure of the route descriptions, the para-linguistic aspects of the message had an effect on how listeners comprehended the message. Exploratory path analyses suggested that it was the perception of positive affect by the listener rather than the direct effect of the vocal cues alone that facilitated communicative effectiveness, although some vocal cues, most notably F1 and jitter, appeared to have a direct effect as well, as indicated by the slightly superior fit of a Dual Route Model compared to a Mediated Route Model. Crucially, this study showed that when listeners perceive speakers as expressing positive affect, regardless of whether it is actually experienced by the speakers or not, this in itself appears to have a beneficial effect on comprehension.

However, it would have seemed premature to discard a Direct Route Model, i.e. the possibility of a direct effect of positive speaker emotion, based on this first, exploratory study. After all, the LSA-based measure, which was used to capture quality of verbal descriptions without making arbitrary judgments about what features of the message to include, could have quite possibly been not sensitive enough to reflect direct effects of positive affect. Moreover, the vocal cues of positive affect expression were measured on very selective parts of a pre-determined sentence thus limiting the analysis of prosodic variability that could have enhanced important parts of the message. Therefore, Study 2 was aimed at exploring, in a more controlled fashion, whether speech expressing positive affect contains cues that enhance the informativeness of the message. The focus was on prosody, and the study employed a prosodic disambiguation task to see whether positive affect leads to more informative prosodic cues in syntactically ambiguous sentences. To find an ecologically valid way of inducing positive affect, I tested mothers addressing their small children, and compared them to a control group of non-mothers addressing an imaginary child. This was motivated by findings suggesting that interaction with their own child is a strong positive mood inducer for mothers (Nitschke et al., 2004; Bartels & Zeki, 2004), and that CDS resembles speech expressing positive affect (Singh et al., 2002; Trainor et al., 2000). The results of this study showed that only non-mothers, but not mothers, produced more pronounced prosodic disambiguation cues. At the same time, mothers demonstrated stronger pitch modifications associated with positive affect expression in CDS, and were judged as sounding happier. This result suggests that positive affect expression does not necessarily lead to improved prosodic disambiguation, and thereby confirms the findings from Study 1, namely, that the association between positive speaker affect and improved communicative effectiveness is not primarily a consequence of the direct effect of vocal cues on the informativeness of the message. Instead, the exaggerated prosody triggered by positive affect in the mothers seemed to mainly fulfill signaling purposes such as manipulating the arousal of the child and eliciting attention. The absence of positive affect in the non-mothers, on the other hand, resulted in an emphasis on linguistically meaningful prosody. This finding was taken as an indication that affective and linguistic prosody might be relatively independent systems, that can be conflicting when pitched against each other.

Two questions arise when comparing the results of Study 1 and Study 2. First, Study 1 showed a slightly better fit for a Dual Route Model, i.e. a model that incorporated both mediated and direct effects of vocal cues of positive affect expression on communicative effectiveness. Study 2, on the other hand, did not show such a direct link between positive affect expression in the mothers and improved comprehension. Here, it is important to remember that the cues that were responsible for the direct effect in Study 1 were F1 and jitter. These cues were not analysed in Study 2, which focused exclusively on the timing of prosodically relevant pauses and F_0 on the preposition *with*. As discussed previously, the direct effect of F1 on comprehension in Study 1 may have been due to increased vocal effort, i.e. loudness, for which F1 is an indirect indicator. Vocal effort, however, was not one of the relevant cues reported by Snedeker & Trueswell (2003) for the type of prosodic disambiguation tested in Study 2, and was, hence, not considered. The direct effect of jitter in Study 1 was difficult to interpret, and was also not considered in Study 2. Thus, the results of Study 2 do not exclude the possibility that other vocal cues associated with positive emotions may have a direct effect on communicative effectiveness; it merely shows that prosodic cues relevant for emphasising the syntactic structure of the message and improving its informativeness are not necessarily amplified under these conditions.

Second, Study 1 showed that perception of positive affect of the speaker by the listener had a positive effect on communicative effectiveness, while in Study 2, comprehension tended to be somewhat more successful when listening to the non-mothers, i.e. speakers who were expressing positive affect to a considerably lesser degree. In this context, it is important to note that the comprehension measures employed in the two studies are not directly comparable. In Study 1, a very coarse but effective measure was used, the median route deviation, which only gives an approximation for how well speakers may be able to understand a particular description, and was chosen in order to be able to examine a large sample of speakers. In Study 2, we presented the sentences to adult listeners in a forced-choice task. However, this measure was not an ecologically valid way of testing the communicative effectiveness of speech addressed to children. It was used just to obtain further evidence for the effect of the prosodic cues. It is not clear whether children would have been able to follow their mothers' instructions despite the fact that they did not contain informative and relevant prosodic cues that could have helped to distinguish between the two sentence interpretations. Unfortunately, the constraints of the laboratory setting and the variability in children's ages made it impossible to test their comprehension of the instructions. Moreover, presenting the recorded instructions of the non-mothers to children of comparable age did not seem to be a methodologically feasible possibility. A procedure in which children's comprehension can be evaluated directly is clearly a goal for future studies. Thus, Study 2 did not address the issue as to whether positive affect expression improves communicative effectiveness for the listener, but tested a more narrow question, namely whether positive affect expression is associated with more informative linguistic, in this case prosodic, cues in speech production.

The findings of Study 2 will also help to refine the understanding of Prosodic Bootstrapping in CDS. While the Prosodic Bootstrapping Hypothesis emphasises the positive role of enhanced prosodic cues to syntactic structure for language learning, it has never explicitly addressed the specific conditions under which speakers tend to align prosodic cues with the syntactic structure of utterances. A number of observational studies on naturalistic corpora have demonstrated that prosodic cues for sentence and clause boundaries do indeed tend to be exaggerated in CDS (Bernstein Ratner, 1986; Fernald et al., 1989). Study 2 of this thesis provides the first experimental test as to whether CDS invariably results in more informative prosodic cues. It demonstrated that when speakers are engaged in affective speech, as are mothers addressing their children, they emphasise affective prosody by, for instance, producing higher pitch, exaggerated pitch contours and slower speech. In doing so, they may fall back onto default strategies of syntactic processing, and fail to fine-tune their prosodic cues to clarify the intended syntactic structure. This does not invalidate the idea of Prosodic Bootstrapping in CDS but suggests that prosodic cues to linguistic structure are not necessarily a by-product of affective prosody.

In sum, the results of the two studies provide converging evidence for the idea that if positive affect expression by the speaker improves communicative effectiveness, it does so via the signaling of positive affect to the listener, and the social and attention eliciting effects this may have. The results presented in this thesis did not find evidence for the possibility that 'happy' speakers produce more informative messages.

How can this conclusion be reconciled with the findings from social-psychological studies which have shown that positive affect is associated with more cooperativeness in negotiation strategies (Forgas, 1998b), and more directness, less politeness and less evasiveness in request formulation (Forgas, 1998a, 1999b; Forgas & Cromer, 2004)? The Affect Infusion Model (Forgas, 1995) focuses on the effect of mood on information processing and communication strategies in the speaker, but does not directly address the issue as to whether messages are more informative for the listener. The fact that speakers adapt their sociopragmatic communication strategies dependent on their affective state does not necessarily imply that these effects result in optimisation of message formulation on all levels of linguistic processing. Still, when the effect on the listener was tested with respect to bargaining outcomes in negotiations, the findings suggested that happy speakers do indeed 'get their way' more often than sad speakers (Forgas, 1998b). Similarly, it is possible that less polite, less evasive, more direct and even rude requests of happy speakers still may result in greater communicative success, although this remains an empirical question for further research.

Thus, while the social-psychological research suggests a direct effect of speaker emotion on the level of communicative strategies, the results of this thesis suggest that the effect of speaker emotion is not necessarily related to increased informativeness of the message or optimised speech production, but to the impact of emotional signaling on the listener. Perception of positive speaker affect by the listener may have, in turn, an impact on perception of the message and the communicative situation by eliciting attention, and by signaling trustworthiness, appeasement or cooperativeness of the speaker.

A number of questions regarding the relationship between positive affect expression and communicative effectiveness remain unanswered and should be addressed in future research. The outcomes of Study 1 suggest that more precise measures are required to determine quality of message and communicative effectiveness. This would be important in order to further dissociate between the linguistic variation in content and structure of the message on the one hand, and the para-linguistic modifications that might reflect affective valence on the other hand. Moreover, it would be interesting to see how the findings of this study replicate if a broader variation in affective valence, and with that a broader range of para-linguistic variation, is examined. This could be, for example, achieved through effective non-verbal mood induction techniques.

Further research should also extend the findings of Study 2 to two additional conditions: mothers addressing an imaginary child, and non-mothers addressing a real child. This would clarify whether the mothers' failure to disambiguate prosodically in CDS is due to increased cognitive load associated with the interaction with the child or whether it is indeed a consequence of their emphasis on affect expression. Furthermore, as for Study 1, prosodic disambiguation should be tested in different affect conditions, employing mood induction techniques for stronger effects of happy vs. sad. The prediction from the findings of this thesis is that happy speakers would exhibit less prosodic disambiguation than speakers in a neutral or sad state. This would strengthen the interpretation of the present results as effects of positive affect expression.

In general, the thesis supports a view by which emotions exert an effect on communication via modifications of para-linguistic emotional signaling. This view is in line with research by Owren & Bachorowski (2003) who argue that non-linguistic vocalisations such as laughter serve the primary purpose of manipulating affect and, in doing so, the behaviour of the listener. Owren & Bachorowski (2003) interpret the mechanism of such manipulation in much the same way I interpreted the mechanisms underlying the vocal correlates of positive affect: They argue that the goal of such manipulation is to increase arousal and attention in the listener, and to *"promote a more favourable stance toward the laugher"* (p.183), similar to what I referred to as 'creating rapport'. They also suggest that direct affect manipulation is, in evolutionary terms, a much simpler and more straightforward function of non-linguistic vocalisations, than the truthful signalling of speaker affect. In other words, there is doubt that the primary function of the non-linguistic vocal signal is to encode information representative of the affective state of the signaller, and that it only influences affect after this information has bee decoded by the listener. It is important to note that Owren & Bachorowski (2003) do not suggest that affective vocalisations are not directly connected to the affective state of the signaller, or that listeners not also infer important information about the signaller, but these mechanisms are seen as by-products rather than the primary function of non-linguistics affective vocalisations. This could, for example, explain why in Study 1 of this thesis the effect of perceived speaker affect override the effect of reported speaker affect. If Owren & Bachorowski (2003) are correct in that the impact of affective vocalisations is immediate and not mediated by information about the speaker's affective state that has to be unequivocally deciphered, it would also leave room for the possibility that such signals occur in many acoustic realisations and across modalities. Laughter, which comes in so many forms and is manifested both in the voice and in the face, is, of course, a prime example. It is reasonable to suggest that laughter and other vocal correlates of positive affect are closely related in evolutionary terms, but more experimental evidence is needed to determine whether the effect of the vocal correlates of positive affect explored in this thesis is indeed as immediate and 'manipulative' as has been proposed for laughter by Owren & Bachorowski (2003).

Another interesting question related to the signalling function of affective vocalisations is whether this signalling is universal and whether its effects are universal as well. This thesis was not designed to address this question directly but some speculations can be made based on the obtained results.

A number of scholars have suggested that there is a certain consistency in the production of affective vocal cues due to universalities in emotion-specific physiological patterning influencing vocal production (for a discussion see Scherer, Banse & Wallbott, 2001). Conclusive evidence for this position is still wanting, as cross-cultural studies are difficult to conduct and therefore rare, but it is often taken as supportive evidence that some of the acoustic features of affective signals have been found to be fairly similar in a wide variety of mammals, including humans (Morton, 1977; Scherer, 1985, 1991, 1994;

Scherer & Kappas, 1988). If vocal affect expression was indeed universal, this should result in greater cross-cultural similarity of affective, compared to linguistic, prosody. In the context of the current thesis this would, in turn, raise the question whether the effect of affective signalling on communicative effectiveness is also universal. However, there are at least two major factors that could modify such effects.

First, cultures and individuals differ with respect to display rules which can modify vocal affect expression (Ekman, 1977; Banse & Scherer, 1996). These display rules might attenuate or enhance affective signalling. For example, a study by Tickle (2000) showed that English speakers' affect was better recognised than Japanese speaker's affect by both English and Japanese listeners. Tickle (2000) put this result down to a more restricted display of affect in the Japanese speakers. Future research will have to explore what potential impact such differences in degree of affect display have on communicative effectiveness.

Second, the effect of affective signaling on communicative effectiveness may depend on the specific interaction between affective and linguistic prosody in any given language. It could be, for example, that tone languages which systematically employ pitch modifications to mark linguistic properties provide less opportunity for affective prosody to have an effect than languages that do not already feature pitch modifications of that type. At the same time, it has been shown that speakers of tone languages reliably make perceptual distinctions between affective and linguistic prosody (Gandour, Wong, Dzemidzic, Lowe, Tong & Li, 2003). Again, it will take further research to uncover how exactly the mechanisms of affective and linguistic prosody interact, and to decide whether this changes the relationship of speaker affect and communicative effectiveness in these languages.

As stated above, these considerations are so far of a speculative nature. Future research will have to account for the cross-cultural differences in emotion expression and its effect on communicative effectiveness. In sum, this thesis has opened up a number of interesting avenues in which to study the communicative effects of speaker affect expression.

Bibliography

Allbritton, D., McKoon, G., & Ratcliff, R. (1996). Reliability of Prosodic Cues for Resolving Syntactic Ambiguity. *Experimental Psychology: Learning, Memory, and Cognition*, *22*, 714–735.

Alpert, M. & Rosen, A. (1990). A Semantic Analsysis of the Various Ways that the Terms "Affect", "Emotion", and "Mood" are Used. *Journal of Communication Disorder*, *23*, 237–246.

American Psychiatric Association (1980). *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed.). Washington, DC: American Psychiatric Association.

Averill, J. R. (1980). A Constructivist View of Emotion. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, Research and Experience* (pp. 305–339). New York: Academic Press.

Bachorowski, J.-A. & Owren, M. (1995). Vocal Expression of Emotion: Acoustic Properties of Speech are Associated with Emotional Intensity and Context. *Psychological Science*, *6*(4), 219–224.

Bachorowski, J.-A. & Owren, M. (2001). Not All Laughs are Alike: Voiced but not Unvoiced Laughter Readily Elicits Positive Affect. *Psychological Science*, *12*, 252–257.

Baddeley, A. D. (1986). *Working Memory*. Oxford University Press.

Banse, R. & Scherer, K. (1996). Acoustic Profiles in Vocal Emotion Expression. *Journal of Personality and Social Psychology*, *70*(3), 614–636.

Barrett, J. & Paus, T. (2002). Affect-Induced Changes In Speech Production. *Experimental Brain Research*, *146*, 531–537.

Bartels, A. & Zeki, S. (2004). The Neural Correlates of Maternal and Romantic Love. *NeuroImage*, *21*, 1155–1166.

Batchelder, E. (2002). Bootstrapping the Lexicon: A Computational Model of Infant Speech Segmentation. *Cognition*, *83*, 167–206.

Baum, S. & Pell, M. (1999). The Neural Bases of Prosody: Insights from Lesion Studies and Neuroimaging. *Aphasiology*, *13*(8), 581–608.

Bernstein Ratner, N. (1986). Durational Cues Which Mark Clause Boundaries in Mother-Child Speech. *Journal of Phonetics*, *14*, 1303–1309.

Biersack, S. & Kempe, V. (2005a). Exploring the Influence of Vocal Emotion Expression on Communicative Effectiveness. *Phonetica*, *62*, 106–119.

Biersack, S. & Kempe, V. (2005b). Tracing Vocal Emotion Expression Through the Speech Chain: Do Listeners Perceive what Speakers Feel? In *Proceedings of the ISCA Workshop on Plasticity in Speech Perception*.

Blanc, J. M., Dodane, C., & Dominey, P. F. (2004). Temporal Processing for Syntax Acquisition: A Simulation Study. *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*.

Boersma, P. & Weenink, D. (2005). Praat: Doing Phonetics by Computer (Version 4.3.04) [Computer Program].

Boomer, D. & Dittmann, A. (1962). Hesitation Pauses and Juncture Pauses in Speech. *Language and Speech*, *5*, 215–220.

Bostanov, V. & Kotchoubey, B. (2004). Recognition of Affective Prosody: Continuous Wavelet Measures of Event-related Brain Potentials to Emotional Exclamations. *Psychophysiology*, *41*, 259–268.

Bower, G. (1981). Mood and Memory. *American Psychologist*, *36*, 129–148.

Brown, G., Anderson, A., Shillock, R., & Yule, G. (1984). *Teaching Talk*. Cambridge: Cambridge University Press.

Brunswik, E. (1956). *Perception and the Representative Design of Psychological Experiments*. Berkeley: University of California Press.

Burkhardt, F. & Sendlmeier, W. (2000). Verification of Acoustical Correlates of Emotional Speech Using Formant Synthesis. *Proceedings of the ISCA Workshop on Speech and Emotion*, 151–156.

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's New Pussycat? On Talking to Babies and Animals. *Science*, *296*, 1435.

Cahn, J. (1990). Generating Expression in Synthesized Speech. Technical report, MIT Media Lab, Boston.

Camras, L. (1992). Basic emotions and expressive development. *Cognition and Emotion*, *6*(3,4), 269–284.

Caplan, D. & Waters, G. (1999). Verbal Working Memory and Sentence Comprehension. *Behavioral Brain Science*, *22*(1), 77–94.

Chevallier-Skolnikoff, S. (1973). Facial Expression of Emotion in Nonhuman Primates. In P. Ekman (Ed.), *Darwin and Facial Expression* (pp. 11–83). New York: Academy Press.

Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A New Graphic Interactive Environment for Designing Psychology Experiments. *Behavioral Research Methods, Instruments, and Computers*, *25*(2), 257–271.

Cornelius, R. R. (1996). *The Science of Emotion. Research and Tradition in the Psychology of Emotion*. Upper Saddle River, NJ: Prentice-Hall.

Dabrowska, E. (2006). Low-level Schemas or Productive Rules: The Role of Diminutives in the Acquisition of Polish Case Inflections. *Language Sciences*, *28*, 120–135.

Daneman, M. & Carpenter, P. (1980). Individual Differences in Working Memory and Reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals* (1979, reprinted ed.). London: Julian Friedmann Publishers.

Davidson, R. (1992). Anterior Cerebral Assymmetry. *Brain and Cognition*, *20*, 125–151.

de Jong, K. (1993). The Supraglottal Articulation of Prominence in English: Linguistic Stress as Localized Hyperarticulation. *Journal of the Acoustical Society of America*, *97*(1), 491–504.

de Jong, K., Beckman, M., & Edwards, J. (1993). The Interplay Between Prosodic Structure and Coarticulation. *Language and Speech*, *36*, 197–212.

Denes, P. & Pinson, E. (1993). *The Speech Chain - The Physiology and Biology of Spoken Language*. New York: W.H. Freeman and Company.

Dimberg, U. & Öhman, A. (1996). Behold the Wrath: Psychophysiological Responses to Facial Stimuli. *Motivation and Emotion*, *20*, 149–182.

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious Facial Reactions To Emotional Facial Expressions. *Psychological Science*, *11*, 86–89.

Ekman, P. (1977). Biological and Cultural Contributions to Body and Facial Movement. In Blacking, J. (Ed.), *The Anthropology of the Body*, (pp. 39–84). London: Academic Press.

Ekman, P. (1989). The Argument and Evidence About Universals in Facial Expressions of Emotion. In H. Wagner & A. Manstead (Eds.), *Handbook of Social Psychophysiology* (pp. 143–164). Chichester: Wiley.

Ekman, P. (1992). An Argument for Basic Emotions. *Cognition and Emotion*, *6*, 169–200.

Ekman, P. (2001). Smiling. In C. Blakemore & S. Jennett (Eds.), *Oxford Companion to the Body*. London: Oxford University Press.

Erickson, D., Fujimura, O., & Dang, J. (1999). Articulatory and Acoustic Characteristics of Emphasized and Unemphasized Vowels. *Journal of the Acoustical Society of America*, *106*(4), 2241.

Falk, D. (2004). Prelinguistic Evolution in Early Hominins: Whence Motherese? *Behavioral and Brain Sciences*, *27*(4), 491–503.

Fernald, A. (1994). Human Maternal Vocalizations to Infants as Biologically Relevant Signals: An Evolutionary Perspective. In P. Bloom (Ed.), *Language Acquisition: Core Readings*. Cambridge, MA: MIT Press.

Fernald, A. & Mazzie, C. (1991). Prosody and Focus in Speech to Infants and Adults. *Infant Behavior and Development*, *27*, 209–221.

Fernald, A. & McRoberts, G. (1996). Prosodic Bootstrapping: A Critical Analysis of the Argument and Evidence. In J. Morgan & K. Demuth (Eds.), *Signal to Syntax : Bootstrapping from Speech to Grammar in Early Acquisition* (pp. 365–388). Hillsdale, NJ: Lawrence Erlbaum Associates.

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B., & Fukui, I. (1989). A Cross-Language Study of Prosodic Modifications in Mothers' and Fathers' Speech to Preverbal Infants . *Journal of Child Language*, *16*, 477–501.

Fisher, C. & Tokura, H. (1996). Acoustic Cues to Grammatical Structure in Infant-Directed Speech:Cross-Linguistic Evidence. *Child Development*, *67*(6), 3192–3218.

Forgas, J. (1995). Mood and Judgment: The Affect Infusion Model (AIM). *Psychological Bulletin*, *117*(1), 39–66.

Forgas, J. (1998a). Asking Nicely? The Effects of Mood on Responding to More or Less Polite Requests. *Personality and Social Psychology Bulletin*, *24*(2), 173–185.

Forgas, J. (1998b). On Feeling Good and Getting Your Way: Mood Effects on Negotiator Cognition and Bargaining Strategies. *Journal of Personality and Social Psychology*, *74*(3), 565–577.

Forgas, J. (1999a). Feeling And Speaking: Mood Effects on Verbal Communication Strategies. *Personality and Social Psychology Bulletin*, *25*(7), 850–863.

Forgas, J. (1999b). On Feeling Good and Being Rude: Affective Influences on Language Use and Request Formulations. *Journal of Personality and Social Psychology*, *76*(6), 928–939.

Forgas, J. & Cromer, M. (2004). On Being Sad and Evasive: Affective Influences on Verbal Communication Strategies in Conflict Situations. *Journal of Experimental Social Psychology*, *40*(4), 511–518.

Foulkes, P., Docherty, G., & Watt, D. (2005). Phonological Variation in Child Directed Speech. *Language*, *81*, 177–206.

Fox Tree, J. & Meijer, P. (2000). Untrained Speakers' Use of Prosody in Syntactic Disambiguation and Listeners' Interpretation. *Psychological Research*, *63*, 1–13.

Frank, M., Ekman, P., & Friesen, W. (1993). Behavioural Markers and Recognizability of the Smile of Enjoyment. *Journal of Personality and Social Psychology*, *64*, 83–93.

Fredrickson, B. (2003). The Value of Positive Emotions. *American Scientist*, *91*, 330–335.

Gandour, J., Wong, D., Dzemidzic, M., Lowe, M., Tong, Y., & Li, X. (2003). A Cross-Linguistic fMRI Study of Perception of Intonation and Emotion in Chinese. *Human Brain Mapping*, *18*(3), 149–157.

Geumann, A. (2001). Vocal Intensity: Acoustic and Articulatory Correlates. In *Proceedings of the 4th Conference on Motor Control*.

Gobl, C. & Ní Chasaide, A. (2003). The Role of Voice Quality in Communicating Emotion, Mood and Attitude. *Speech Communication*, *40*, 189–212.

Grammer, K. (1990). Strangers Meet: Laughter and Nonverbal Signs of Interest in Opposite-Sex Encounters. *Journal of Nonverbal Behavior*, *14*, 209–236.

Gray, J. R. (2001). Emotional Modulation of Cognitive Control: Approach-Withdrawl States Double-Dissociate Spatial From Verbal Two-Back Task Performance. *Journal of Experimental Psychology*, *130*(3), 436–452.

Gross, J. (2002). Emotion Regulation: Affective, Cognitive, and Social Consequences. *Psychophysiology*, *39*, 281–291.

Gussenhoven, C. (2002). Intonation and Interpretation: Phonetics and Phonology. In *Proceedings of the 1st Speech Prosody*.

Hatfield, E., Cacioppo, J., & Rapson, R. (1992). Primitive Emotional Contagion. In M. Clark (Ed.), *Emotion and Social Behaviour* (pp. 151–177). London: Sage.

Hawkins, P. (1971). The Syntactic Location of Hesitation Pauses. *Language and Speech*, *14*, 277–288.

Henderson, A., Goldman-Eisler, F., & Skarbek, A. (1966). Sequential temporal Patterns in Spontaneous Speech. *Language and Speech*, *9*, 207–216.

Herrera, E., Reissland, N., & Shepherd, J. (2004). Maternal Touch and Maternal Child-Directed Speech: Effects of Depressed Mood in the Postnatal Period. *Journal of Affective Disorders*, *81*, 29–39.

Isen, A. (1987). Positive Affect, Cognitive Processes and Social Behaviour by supporting compassionate acts to- positive affect, cognitive processes and social behavior. *Advances in Experimental Social Psychology*, 202–253.

Jacobson, J., Boersma, D., Fields, R., & Olson, K. (1983). Paralinguistic Features of Adult Speech to Infants and Small Children. *Child Development*, *54*, 436–442.

Johnson-Laird, P. & Oatley, K. (1992). Basic Emotions, Rationality, and Folk Theory. *Cognition and Emotion*, *6*, 201–223.

Johnstone, T. & Scherer, K. R. (1999). The Effects of Emotions on Voice Quality. *Proceedings of the 14th International Conference of Phonetic Sciences*.

Kaplan, P., Bachorowski, J.-A., Smoski, M. J., & Hudenko, W. J. (2002). Infants of Depressed Mothers, Although Competent Learners, Fail to Learn in Response to Their Own Mothers' Infant-Directed Speech. *Psychological Science*, *13*(3), 268–271.

Kaplan, P., Bachorowski, J.-A., & Zarlengo-Strouse, P. (1999). Child-Directed Speech Produced by Mothers with Symptoms of Depression Fails to Promote Associative Learning in 4-Month-Old Infants. *Child Development*, *70*(3), 560–570.

Kempe, V., Brooks, P. J., & Gillis, S. (2005). Diminutives in Child-Directed Speech Supplement Metric with Distributional Word Segmentation Cues. *Psychonomic Bulletin and Review*, *12*(1), 145–151.

Kempe, V., Brooks, P. J., Mironova, N., & Fedorova, O. (2003). Diminutivisation Supports Gender Acquisition in Russian Children. *Journal of Child Language*, *30*, 1–15.

King, J. & Just, M. (1991). Individual Differences in Syntactic Processing: The Role of Working Memory. *Journal of Memory and Language*, *30*, 580–602.

Kitamura, C. & Burnham, D. (1996). Pitch and Communicative Intent in Infant-Directed Speech: Longitudinal Data. In *Proceedings of the 6th Australian International Conference on Speech Science & Technology*, (pp. 325–330).

Kitamura, C. & Burnham, D. (1998). Acoustic and Affective Qualities of IDS in English. In *Proceedings of the 5th ICSLP*, (pp. 441–444).

Kitamura, C. & Burnham, D. (2003). Pitch and Communicative Intent in Mother's Speech: Adjustments for Age and Sex in the First Year. *Infancy*, *4*(1), 85–110.

Kitamura, C., Thanavvishuth, C., Burnham, D., & Luksaneeyanawin, S. (2002). Universality and Specificity in Infant-Directed Speech: Pitch Modifications as a Function of Infant Age and Sex in a Tonal and Non-Tonal Language. *Infant, Behavior and Development*, *24*, 372–392.

Landauer, T. & Dumais, S. (1997). A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Represantation of Knowledge. *Psychological Review*, *104*(2), 211–240.

Lehiste, I. (1973). Phonetic Disambiguation of Syntactic Ambiguity. *Glossa*, *7*, 102–122.

Levenson, R., Ekman, P., & Friesen, W. (1990). Voluntary Facial Expression Generates Emotion-Specific Nervous System Activity. *Psychophysiology*, *27*, 363–384.

Liu, H.-M., Kuhl, P., & Tsao, F.-M. (2003). An Association Between Mothers' Speech Clarity and Infants' Speech Discrimination Skillls. *Developmental Science*, *6*(3).

Locke, J. (2001). First Communication: The Emergence of Vocal Relationships. *Social Development*, *10*(3), 294–308.

Mayer, J. D. & Gaschke, Y. N. (1988). The Experience and Meta-Experience of Mood. *Journal of Personality and Social Psychology*, *55*, 102–111.

McRoberts, G., Studdert-Kennedy, M., & Shankweiler, D. (1995). The Role of Fundamental Frequency in Signalling Linguistic Stress and Affect: Evidence for a Dissociation. *Perception & Psychophysics*, (2), 159–174.

Mehrabian, A. & Russell, J. (1974). *An Approach to Environmental Psychology*. Cambridge: Cambridge University Press.

Morgan, J. & Demuth, C. (1996). *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acqusition*. Hillsdale: Lawrence Erlbaum Associates.

Morgan, J. L., Shi, R., & Allopena, P. (1996). Perceptual Bases of Rudimentary Grammatical Categories: Toward a Broader Conceptualization of Bootstrapping. In Morgan, J. & Demuth, K. (Eds.), *From Signal to Syntax*, (pp. 263–281)., Mahwah, NJ. Erlbaum.

Morton, E. (1977). On the Occurrence and Significance of Motivational-Structural Rules in some Bird and Mammal Sounds. *American Naturalist*, *111*, 855–869.

Mozziconacci, S. (2002). Prosody and Emotions. *Proceedings of the 1st Speech Prosody*.

Murray, I. & Arnott, J. (1993). Toward the Simulation of Emotion in Synthetic Speech: A Review of the Literature on Human Vocal Emotion. *Journal of the Acoustical Society of America*, *93*(2), 1097–1108.

Nicely, P., Tamis-LeMonda, C., & Bornstein, M. (1999). Mothers' Attuned Responses to Infant Affect Expressivity Promote Earlier Achievement of Language Milestones. *Infant Behavior and Development*, *22*, 557–568.

Nitschke, J., Nelson, E., Rusch, B., Fox, A., Oakes, T., & Davidson, R. (2004). Orbitofrontal Cortex Tracks Positive Mood in Mothers Viewing Pictures of Their Newborn Infants. *NeuroImage*, *21*, 583–592.

Oatley, K., Keltner, D., & Jenkins, J. (2006). *Understanding Emotions*. Cambridge, MA: Blackwell Publishing.

Ortony, A., Clore, G. L., & Collins, A. (1988). *The Cognitive Structure of Emotion*. Cambridge: Cambridge University Press.

Owren, M. & Bachorowski, J.-A. (2003). Reconsidering the Evoultion of Nonlinguistic Communication: The Case of Laughter. *Journal of Nonverbal Behaviour*, *27*(3), 183–200.

Panksepp, J. (1998). *Affective Neuroscience: The Foundations of Human and Animal Emotions*. New York: Oxford University Press.

Patterson, D. & Ladd, R. (1999). Pitch Range Modelling: Linguistic Dimension of Variation. In *Proceedings of the 14th International Conference of Phonetic Sciences*, (pp. 1169–1172).

Pfitzinger, H. (2001). Phonetische Analyse der Sprechgeschwindigkeit. *FIPKM*, *38*, 117–264.

Pompino-Marschall, B. (1995). *Einführung in die Phonetik*. Berlin, New York: Walter de Gruyter.

Price, P., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The Use of Prosody in Syntactic Disambiguation. *Journal of the Acoustical Society of America*, *90*(6), 2956–2970.

Raine, A., Hulme, C., Chadderton, H., & Bailey, P. (1991). Verbal Short-Term Memory Span in Speech-Disordered Children: Implications for Articulatory Coding in Short-Term Memory. *Child Development*, *62*, 415–423.

Redican, W. (1982). An Evolutionary Perspective on Human Facial Displays. In P. Ekman (Ed.), *Emotion in the Human Face* (pp. 212–280). Elmsford, New York: Pergamon.

Robson, J. & Mackenzie Beck, J. (1999). Hearing Smiles – Perceptual, Acoustic and Production Aspects of Labial Spreading. *Proceedings of the 14th International Conference of Phonetic Sciences*, 219–222.

Rolls, E. (2000). Précis of The Brain and Emotion. *Behavioral and Brain Sciences*, *23*, 177–234.

Russell, J. (2003). Core Affect and the Psychological Construction of Emotion. *Psychological Review*, *110*(1), 145–172.

Russell, J., Bachorowski, J.-A., & Fernández-Dols, J.-M. (2003). Facial and Vocal Expressions of Emotion. *Annual Review of Psychology*, *54*, 329–349.

Russell, J. M. & Fernández-Dols, J. M. (1997). What does a facial expression mean? In J. M. Russell & J. M. Fernández-Dols (Eds.), *The Psychology of Facial Expression* (pp. 3–30). Cambridge: Cambridge University Press.

Ryan, S. (1991). *Pitch Contours in Japanese Mothers' Speech to Infants*. Unpublished manuscript, Stanford University.

Scherer, K. (1986). Vocal Affect Expression: A Review and a Model for Future Research. *Psychological Bulletin*, *99*(2), 143–165.

Scherer, K. (1985). Vocal Affect Signalling: A Comparative Approach. In Rosenblatt, J., Beer, C., Busnel, M., & Slater, P. (Eds.), *Advances in the Study of Behavior*, (pp. 189–244). New York: Academic Press.

Scherer, K. (1991). Emotion Expression in Speech and Music. In Sundberg, J., Nord, L., & Carlson, R. (Eds.), *Music, Language, Speech, and Brain*, (pp. 146–156). Wenner-Gren Center International Symposium series. London: Macmillan.

Scherer, K. (1994). Affect Bursts. In van Goozen, S., van de Poll, & Sergeant, J. (Eds.), *Emotions: Essays on Emotion Theory*, (pp. 161–196). Hillsdale: Lawrence Erlbaum Associates.

Scherer, K. (2003). Vocal Communication of Emotion: A Review of Research Paradigms. *Speech Communication*, *40*, 227–256.

Scherer, K., Banse, R., & Wallbott, H. (2001). Emotion Inferences from Vocal Expression Correlate Across Languages and Cultures. *Journal of Cross-Cultural Psychology*, *32*(1), 76–92.

Scherer, K. & Kappas (1988). Primate Vocal Expression of Affective States. In Todt, D., Goedeking, P., & Newman, E. (Eds.), *Primate Vocal Comunication*, (pp. 171–194). Heidelberg: Springer.

Schröder, M. (2004). Speech and Emotion Research: An Overview of Research Frameworks and a Dimensional Approach to Emotional Speech Synthesis. *PHONUS 7, Research Report of the Institute of Phonetics, Saarland University*.

Scott, S., Young, A., Calder, A., Hellawell, D., Aggleton, J., & Johnson, M. (1997). Impaired Auditory Recognition of Fear and Anger Following Bilateral Amygdala Lesions. *Nature*, *385*, 254–257.

Seddoh, A. (2002). How Discrete or Independent are 'Affective Prosody' and 'Linguistic Prosody'. *Aphasiology*, *16*(7), 683–692.

Singh, L., Morgan, J., & Best, C. (2002). Infants' Listening Preferences: Baby Talk or Happy Talk? *Infancy*, *3*, 365–394.

Snedeker, J. & Trueswell, J. (2003). Using Prosody to Avoid Ambiguity: Effects of Speaker Awareness and Referential Context. *Journal of Memory and Language*, *48*, 103–130.

Spivey-Knowlton, M. & Sedivy, J. (1995). Resolving Attachment Ambiguities with Multiple Constraints. *Cognition*, *55*, 227–267.

Tartter, V. C. (1980). Happy Talk: Perceptual and Acoustic Effects of Smiling in Speech. *Perception and Psychophysics*, *27*(1), 24–27.

Tartter, V. C. & Brown, D. (1994). Hearing Smiles and Frowns in Normal and Whisper Registers. *Journal of the Acoustical Society of America*, *96*(4), 2102–2107.

The American Heritage Dictionary of the English Language (2004). Boston: Houghton Mifflin.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-Directed Speech Facilitates Word Segmentation. *Infancy*, *7*(1), 53–71.

Tickle, A. (2000). English and Japanese Speakers' Emotion Vocalisation and Recognition. In *Proceedings of the ISCA Workshop on Speech and Emotion*, (pp. 104–109). Northern Ireland.

Tillmann, H.-G. (1980). *Phonetik - Lautsprachliche Zeichen, Sprachsignale und lautsprachlicher Kommunikationsprozeß*. Stuttgart: Klett-Cotta.

Trainor, L., Austin, C., & Desjardins, R. (2000). Is Infant-Directed Speech Prosody a Result of he Vocal Expression of Emotion? *Psychological Science*, *11*(3).

Trainor, L. & Desjardins, R. (2002). Pitch Characteristics of Infant-Directed Speech Affect Infant's Ability to Discriminate Vowels. *Psychonomic Bulletin and Review*, *9*(2), 335–340.

Traunmüller, H. (1988). Paralinguistic Variation and Invariance in the Characteristic Frequencies of Vowels. *Phonetica*, *45*, 1–29.

Velten, E. (1968). A Laboratory Task for Induction of Mood States. *Behavior Research and Therapy*, *6*, 473–482.

Waters, G. & Caplan, D. (1996). Processing Resource Capacity and the Comprehension of Garden Path Sentences. *Memory and Cognition*, *24*, 342–355.

Werker, J. & McLeod, P. J. (1989). Infant Preference for both Male and Female Infant-Directed Talk: A Developmental Study of Attentional and Affective Responsiveness. *Canadian Journal of Psychology*, *43*(2), 230–246.

Wolfe, V., Fitch, J., & Cornell, R. (1995). Acoustic Prediction of Severity in Commonly Occurring Voice Problems. *Journal of Speech and Hearing Research*, *38*, 273–279.

Wong, P. (2002). Hemispheric Specialization of Linguistic Pitch Patterns. *Brain Research Bulletin*, *59*(2), 83–95.

Yumoto, E., Sasaki, Y., & Okamura, H. (1984). Harmonics-to-Noise Ratio and Psychophysical Measurement of the Degree of Hoarseness. *Journal of Speech and Hearing Research*, *27*, 2–6.

Zebrowitz, L. A., Brownlow, S., & Olson, K. (1992). Baby Talk to the Babyfaced. *Journal of Nonverbal Behavior*, *16*(3), 143–158.

Ziegler, W. (2003). Speech Motor Control is Task-Specific: Evidence from Dysarthria and Apraxia of Speech. *Aphasiology*, *17*(1), 3–36.

Appendix A

The Brief Mood Introspection Scale (Study 1)

Figure A.1: The Brief Mood Introspection Scale, as used in Study 1

Participant Number ____________

Underline the response on the scale below that indicates how well each adjective describes your present mood.

tired	definitely do not feel	do not feel	slightly feel	definitely feel
jittery	definitely do not feel	do not feel	slightly feel	definitely feel
bubbly	definitely do not feel	do not feel	slightly feel	definitely feel
calm	definitely do not feel	do not feel	slightly feel	definitely feel
sad	definitely do not feel	do not feel	slightly feel	definitely feel
happy	definitely do not feel	do not feel	slightly feel	definitely feel
lively	definitely do not feel	do not feel	slightly feel	definitely feel
caring	definitely do not feel	do not feel	slightly feel	definitely feel
drowsy	definitely do not feel	do not feel	slightly feel	definitely feel
active	definitely do not feel	do not feel	slightly feel	definitely feel
gloomy	definitely do not feel	do not feel	slightly feel	definitely feel
content	definitely do not feel	do not feel	slightly feel	definitely feel
grouchy	definitely do not feel	do not feel	slightly feel	definitely feel
fed up	definitely do not feel	do not feel	slightly feel	definitely feel
loving	definitely do not feel	do not feel	slightly feel	definitely feel
nervous	definitely do not feel	do not feel	slightly feel	definitely feel

Appendix B

Examples for Good and Bad Route Descriptions,

Examples for Good and Bad Route Drawings

(Study 1)

The best route description

You start off at the cake at the top of the map and go up and to the right all over and go straight on right then you come down on the right hand side looking at the paper of the bus come under the bus and then keep going round to the left of the ball and around the ball go down to the bottom and underneath the sock and then come along to the right hand side of the paper come up and right to the flag and along the top and stop just below the ball. That's it.

This description yielded the lowest score of route deviation in cm^2 (57) and was thus taken as the target passage for the LSA measurements. Consequently, the LSA score for this description was 1.

The worst route description

It is like a figure of 8 but the end of it do not touch uh it is also like an s-shape and the end of the s-shape curves back in ehm it starts of with a big curve almost a circle but not quite and then the end of the circle goes up past the beginning and then round into half of an s-shape.

This description yielded one of the highest route deviations in cm^2 (269) and the overall lowest LSA score (0.17). It is important to note that this description does not contain any mentioning of the landmarks.

Figure B.1: Example of a good route drawing, with the original map in the bottom left corner

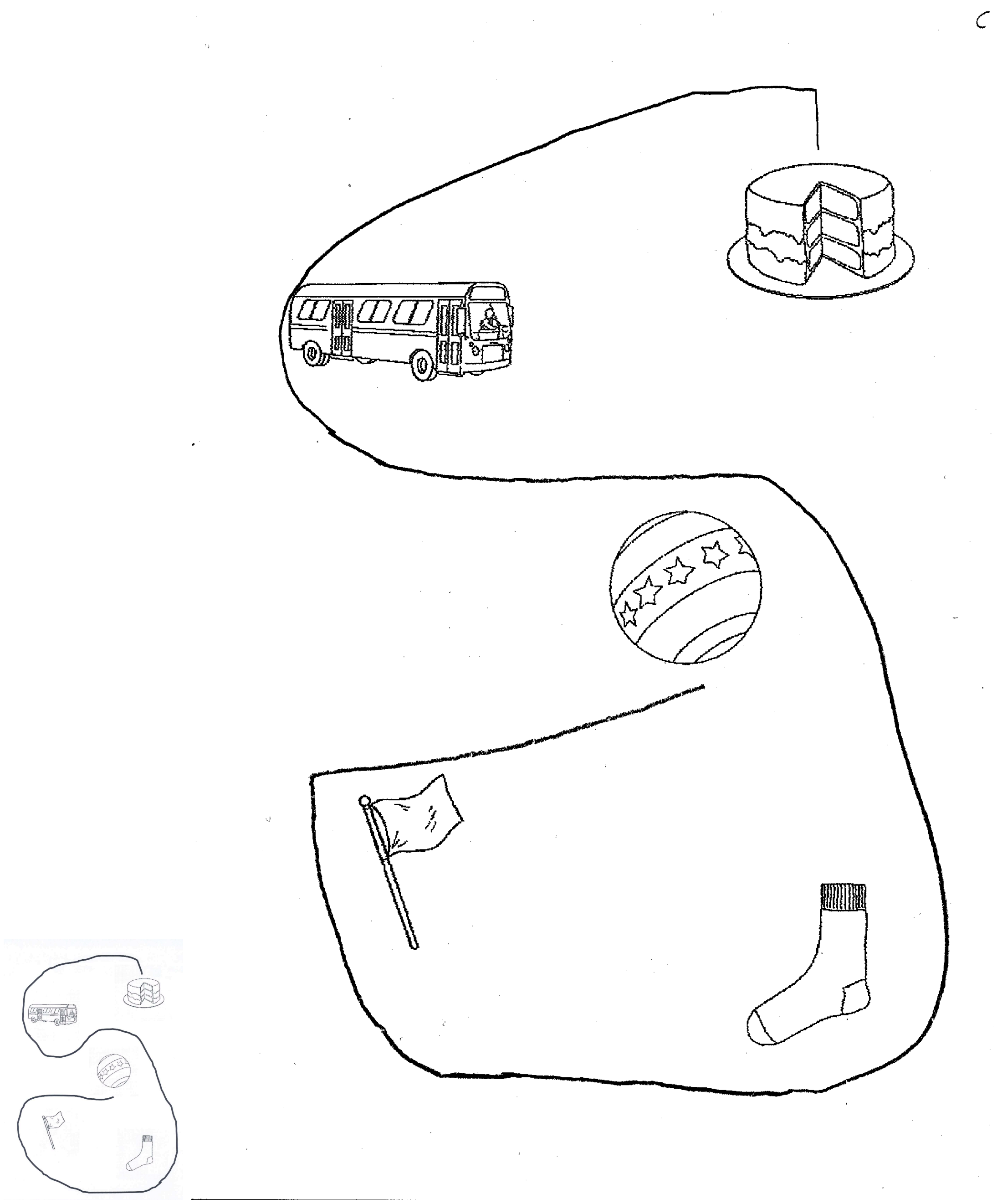
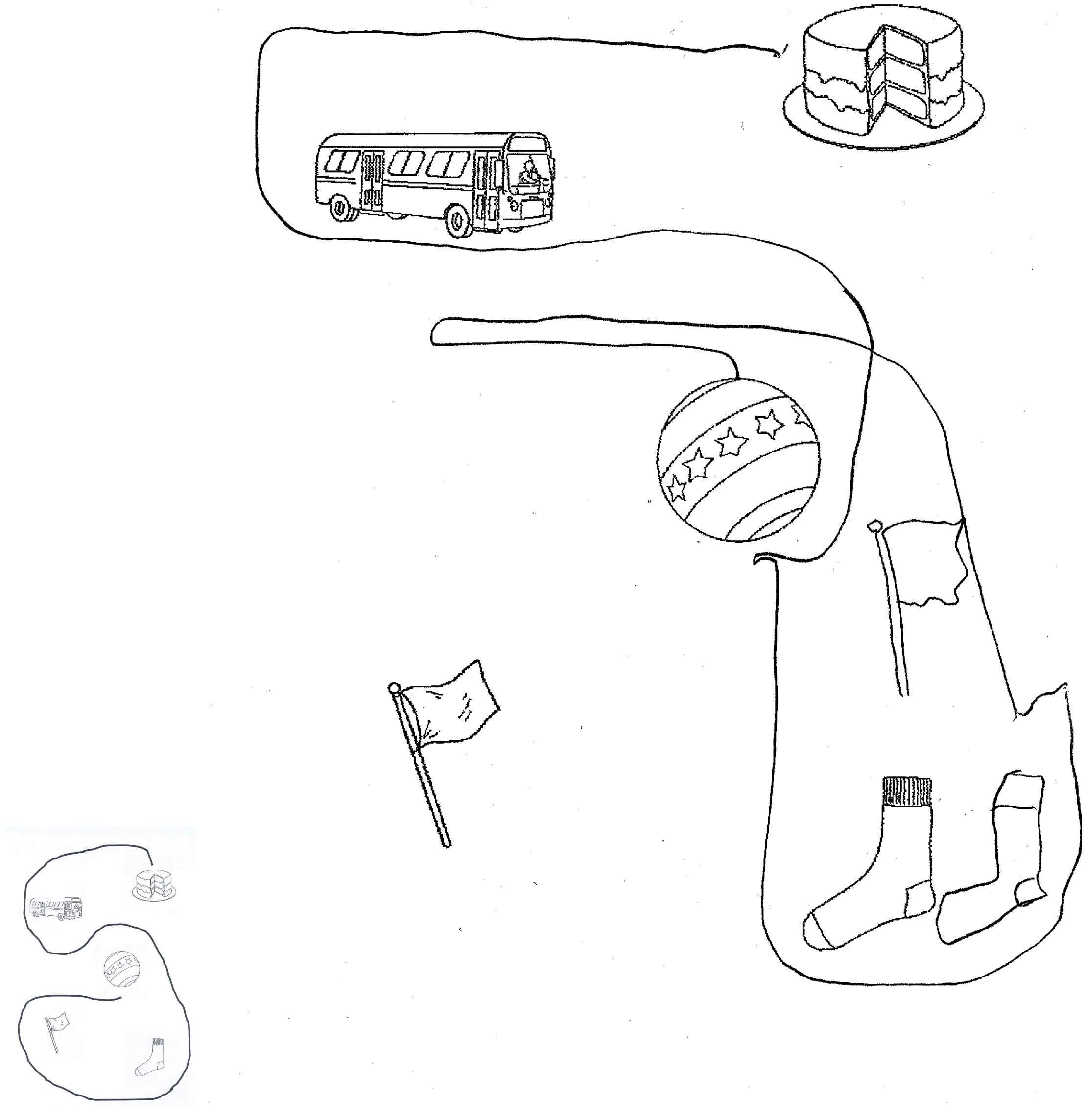


Figure B.2: Example of a bad route drawing, with the original map in the bottom left corner



Appendix C

Screenshots of Praat Analysis Windows

Figure C.1: Praat screenshot of "Please draw the route exactly as I say" (Study 1)

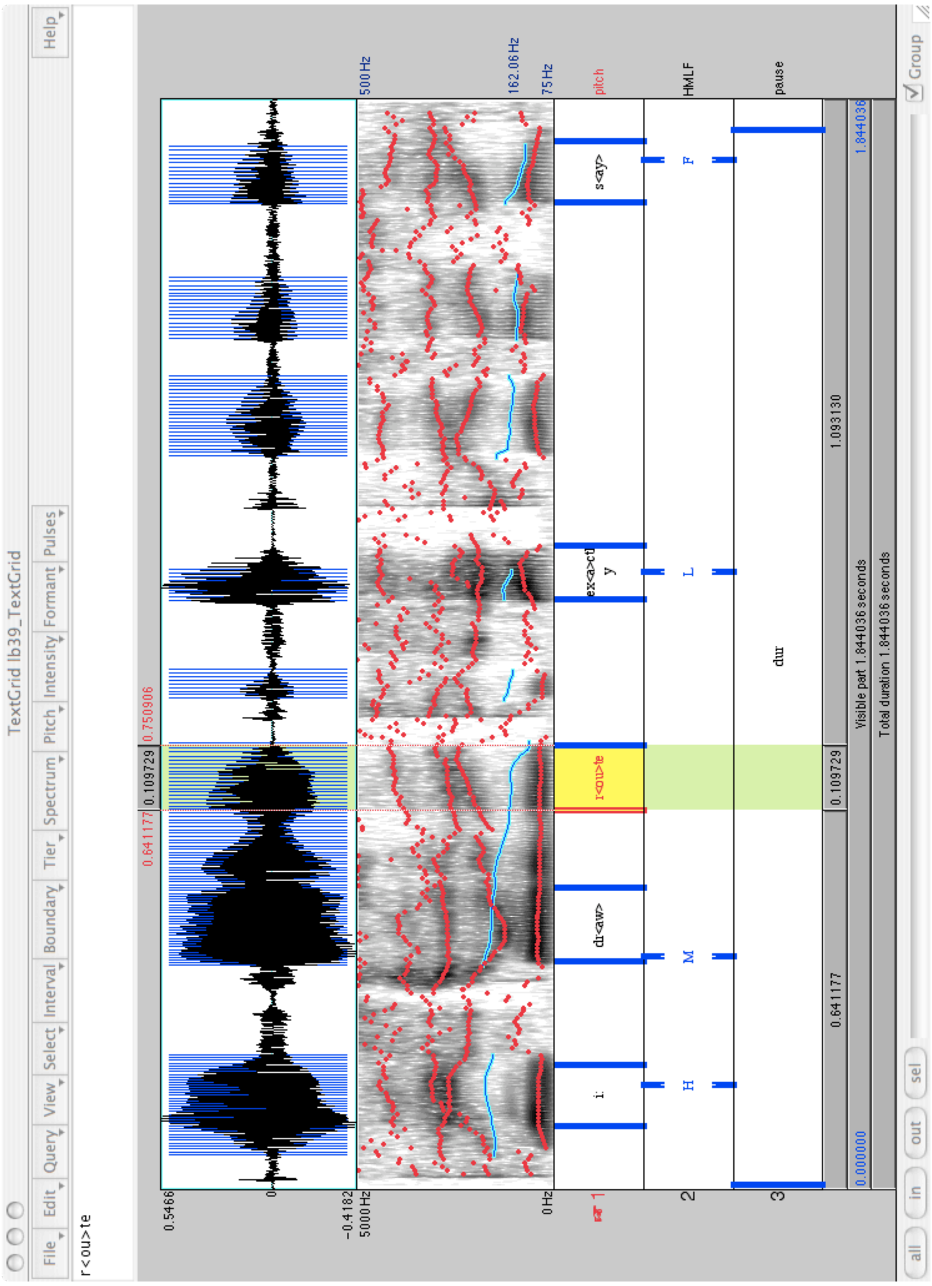
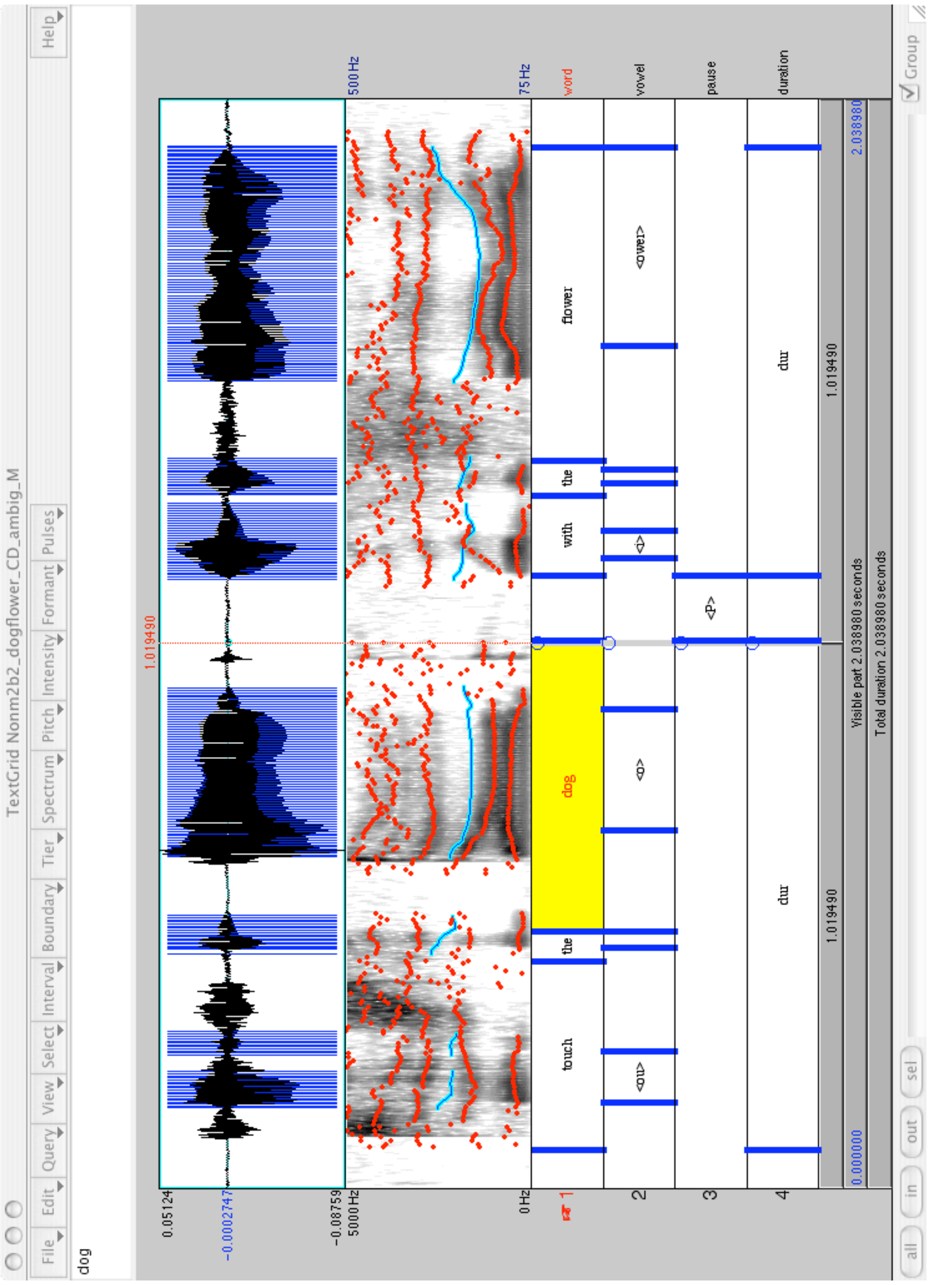


Figure C.2: Praat screenshot “Touch the dog with the flower” (Study 2)



Appendix D

Test Sentences for Study 2

Booklet 1 comprised the following sentences:

– *Touch the snake.*

– **Touch the dog with the flower.** MODIFIER (ambiguous context)

– *Touch the snake and the fish.*

– **Touch the fish with the flower.** INSTRUMENT (unambiguous context)

– *Touch the fish.*

————————- array changeover ————————-

– *Touch the pig.*

– **Touch the cat with the spoon.** INSTRUMENT (ambiguous context)

– *Touch the pig and the frog.*

– **Touch the duck with the flower.** MODIFIER (unambiguous context)

– *Touch the frog.*

Booklet 2 contained the same sentences but with arrays in reversed order.

Booklet 3 (and in reversed order Booklet 4) contained the following sentences:

- *Touch the snake.*
- **Touch the dog with the flower.** INSTRUMENT (ambiguous context)
- *Touch the snake and the fish.*
- **Touch the horse with the spoon.** MODIFIER (unambiguous context)
- *Touch the fish.*

————————- array changeover ————————-

- *Touch the pig.*
- **Touch the cat with the spoon.** MODIFIER (ambiguous context)
- *Touch the pig and the frog.*
- **Touch the frog with the spoon.** INSTRUMENT (unambiguous context)
- *Touch the frog.*

Appendix E

Glossary

Affect Term used in this thesis for emotion-related behaviour that can occur for a longer time span, as an extension of a certain emotion. → *Section 2.1*

Appraisal The link between an event and the emotion that this event elicits. Appraisal takes place in two stages: In the first stage the event is evaluated in terms of how good or bad it is, in the second stage the event is evaluated in terms of how it can be consolidated with the individual's current goals, needs and beliefs (Oatley et al., 2006). The type of emotion elicited depends on the appraisal. → *Section 2.2*

Approach state vs. **withdrawal state**; conceptualisations supporting the notion that emotional appraisal is intimately linked with distinctive activation patterns in the brain. EEG studies have shown that positive emotions like happiness, involving approach tendencies such as reaching out or smiling, are accompanied by more left-sided activation in the frontal brain region. Negative emotions like disgust, involving withdrawal tendencies such as wrinkling of the nose or pulling away, are accompanied by more right-sided activation in the frontal brain region (Davidson, 1992). → *Section 3.1.1*

CDS Acronym for child-directed speech. Child-directed speech denotes the speech register that speakers and in particular parents use to address a child of language acquisition age. The term can be used to distinguish CDS from **ADS** (adult-directed speech) and **IDS** (infant-directed speech, speech register used to address pre-linguistic infants < 24 months). → *Section 5.1.1*

Emotion Term used in this thesis only for **basic emotions** in Ekman's sense. Basic emotions are discrete states such as happiness, sadness, anger, fear and disgust and are defined by nine fundamental components: distinctive universal signals, presence in other primates, distinctive physiology, distinctive universals in antecedent events, coherence among emotional response, quick onset, brief duration, automatic appraisal, unbidden occurrence. → *Section 2.1 and 2.2*

Formants Term referring to resonant frequencies of the vocal tract. Formants provide a prime means of distinguishing between different vowels, because each vowel is characterised by its own distinctive formant pattern. Vowels are produced by vibration of the vocal folds, in combination with vowel-specific configurations of the overall shape of the vocal tract. The vocal tract acts as a resonator, selectively amplifying those sound frequencies that correspond to the characteristic resonances of a particular vocal tract shape. Each vowel requires a different vocal tract shape, and hence is characterised by different resonances (i.e. formants). Most vowels have four or more distinguishable formants referred to as F1, F2, F3, F4 (in order of increasing frequency), and measured in Hertz (Hz). Formants are particularly salient to human speech perception. The first two formants (F1, F2) provide sufficient information for listeners to identify a particular vowel. Perturbations of vocal tract shape triggered by displays of emotion (like smiling, wider opening of the mouth) also alter vowel formants, and can be sufficiently large to produce audible consequences.

F_0 or **fundamental frequency** refers to the rate at which the vocal folds vibrate. This oscillation rate leads to the perceptual sensation of **pitch**. F_0 is dependent on the amount of air pressure employed to excite vocal fold vibration, and on the length and tension of the vocal folds. In parts, changes to these factors are subject to voluntary control, and can be used for linguistic means, e.g. to mark distinctions in tone languages or to mark the sentence type or focus. But changes to vocal fold vibration can also occur due to involuntary factors such as changes in overall muscle tension. F_0 values are usually measured in Hertz (Hz), a linear scale, although this is not the best scale with regard to the perceptual sensation. → **Pitch**

Jitter A measure of irregularity in the voice. Jitter refers to irregularities in the fundamental frequency. Jitter can be caused by a variety of factors (e.g. vocal fold tension, imbalances between the vocal folds, imbalances in the amount of applied pressure etc.) (Pompino-Marschall, 1995), not all of which are well understood. Jitter measures have originally been used in the area of speech pathology, as increased jitter correlates with the impression of hoarseness or roughness of the voice (e.g. Yumoto, Sasaki & Okamura, 1984).

Mood Term referring to a sustained, and often diffuse affect state. Unlike affect and emotions, mood can occur without any external cause. → *Section 2.1*

Pitch The term pitch refers to the perceptual sensation evoked by the rate of vocal fold oscillation (**F_0**). Generally, raises in F_0 lead to a perception of raises in pitch. This connection is, however, logarithmic rather than linear, and a variety of scales has been suggested to describe this connection. The most well-known scale is the semitone scale, which reflects the fact that a duplication of F_0 always leads to a similar sensation of pitch change, i.e. a change from 100 to 200 Hz is perceived as about as large as a change from 1000 to 2000 Hz. The terms 'fundamental frequency' and 'pitch' are sometimes used synonymously in the literature.

Shimmer A measure of irregularity in the voice. Shimmer refers to irregularities in the intensity or amplitude of the vocal fold vibration. Shimmer is caused by similar factors as jitter (Pompino-Marschall, 1995), and is, like jitter, an important diagnostic measure in speech pathology (e.g. Wolfe, Fitch & Cornell, 1995).

Speech rate A feature of speech that can be measured in various ways[1], for example as phone rate (the number of 'sounds' per time unit) or syllable rate (the number of syllables per time unit). Speech rate can also be indirectly estimated based on pause durations and overall sentence duration if the aim is to compare speech rate across similar utterances. These estimates are, however, less reliable.

Valence Term used in theories that propose a dimensional representation of emotions. Valence refers to the dimensional scale of pleasantness, i.e. positive/negative, liking/disliking. → *Section 2.1*

Working memory Conceptualisation referring to the brain's structures and processes used for temporarily storing and manipulating information. → *Section 3.1.1*

[1]For an extensive account on speech rate measurements see Pfitzinger, 2001.