A CROSS-LINGUISTIC STUDY OF

AFFECTIVE PROSODY PRODUCTION
BY MONOLINGUAL AND BILINGUAL CHILDREN

Ioulia Grichkovtsova

Submitted for the Degree of

Doctor of Philosophy

at Queen Margaret University College

on completion of research in the

Speech and Hearing Sciences

September 2006.

I hereby declare that the work presented in this thesis was carried out by myself at Queen Margaret University College, Edinburgh, except where due acknowledgement is made, and has not been submitted for any other degree.

Ioulia Grichkovtsova (Candidate)

Date

To my sons,

Alexander Massinissa and Grégoire Jugurtha

Scientific communications

Peer-reviewed publications

1. **I. Grichkovtsova** and I. Mennen, "Child affective speech recording and its acoustic analysis in a cross-linguistic perspective", Proceedings of Measuring Behavior 2005, Wageningen, Netherlands, 30 August - 2 September 2005.

2. **I. Grichkovtsova** and I. Mennen, "Scottish English and French affective prosody: a cross-linguistic study of bilingual children", Proceedings of III Conference on Experimental Phonetics, University of Santiago de Compostela, 24-26 October 2005.

3. **I. Grichkovtsova** and I. Mennen, "Pitch range variation in child affective speech", Proceedings of Speech Prosody 3rd International Conference, Dresden, 2-5 May 2006.

4. **I. Grichkovtsova** and I. Mennen, "Child speech and emotions: a cross-linguistic prospective", book chapter; Editor: K. Izdebski "Emotions of the Human Voice" vol.2 , Plural Publishing, 2007.

Conferences

1. **I. Grichkovtsova** and I. Mennen "Affective prosody of bilingual children: a cross-linguistic study". *Oral presentation* at the XV Annual Pacific Voice Conference and First International Pacific Voice & Speech Foundation / PIXAR Animation Studios Conference on Voice Quality and Emotions. PIXAR Animations Studios, Emeryville, California, USA, 11-13 March 2005.

2. **I. Grichkovtsova** and I. Mennen "Child affective speech: A cross-linguistic study of bilinguals (Scottish English and French)". *Oral presentation* at the 5th International Symposium on Bilingualism (ISB5). Special Colloquium "Acquisition of phonological structure and phonetic detail in a bilingual context". Barcelona and 20-23 March 2005.

3. **I. Grichkovtsova** and I. Mennen "A cross-linguistic study of affective speech in

bilingual children". Poster presentation at the Xth International Congress For the Study of Child Language. Berlin, Germany, 25-29 July 2005.

4. **I. Grichkovtsova** and I. Mennen "Child affective speech recording and its acoustic analysis in a cross-linguistic perspective". Poster presentation at the 5th International Conference on Methods and Techniques in Behavioral Research. Wageningen, Netherlands, 30 August - 2 September 2005.

5. **I. Grichkovtsova**, "Child affective prosody in Scottish English and French". *Oral presentation* at the conference of the British Association of Academic Phoneticians - BAAP 2006, Edinburgh, Scotland, 10-12 April 2006.

6. **I. Grichkovtsova** and I. Mennen "Pitch range variation in child affective speech". Poster presentation at Speech Prosody 2006, Dresden, Germany, 2-5 May 2006.

7. **I. Grichkovtsova**, "Child affective prosody in Scottish English and French". Poster presentation at thematic school of CNRS "Language acquisition and corpus in linguistics and psychology", Moissac, France, 14-19 May 2006.

8. **I. Grichkovtsova**, "Affective prosody of bilingual children". *Oral presentation* at the 6th International Symposium on Bilingualism (ISB6), Hamburg, Germany, 30 May-2 June 2007.

Workshops

1. **I. Grichkovtsova** and I. Mennen, "Affective prosody of bilingual children". Poster presented at "Multidisciplinary approaches to bilingualism" Tutorials in behavioural and brain sciences for PhD students, Max Plank Institute for Psycholinguistics, Ehzerworld, Netherlands, 28 Juin-02 July 2004.

2. **I. Grichkovtsova** and I. Mennen, "Affective prosody of bilingual children". Poster presented at North Western Centre for Linguistics - Research Training for Postgraduate Linguistics Students, Manchester, UK, 6-10 September 2004.

Seminars

1. **I. Grichkovtsova**, "Affective Speech in Scottish English and French: monolingual vs. bilingual children". Seminar presented at Phonetics/Phonology Workshop, University of Edinburgh, 29th of November 2005.

Acknowledgements

A PhD thesis is usually regarded as a great challenge to take on. Giving birth to twin children and looking after them is also viewed as a particularly demanding task - probably a lot more than PhD work according to my experience. It so happened that after I applied for a PhD position at Queen Margaret University College, I fell pregnant with twins. My application accepted, I had to demonstrate that my motivation was such that I could take on the two greatest challenges of my life at the same time, and that was hard work (and I mean really hard!) It is therefore a great pleasure for me to take this opportunity to acknowledge and thank all the people who supported me and helped me, in one way or another, in going through the preparation my PhD thesis during these four years, partly in Edinburgh, partly in France.

I wish to thank first Dr. Ineke Mennen who accepted the task of being the main supervisor of my PhD project. I thank her not only for her supervision and guidance, but also for her strong support and trust in me all those years. I also thank Dr. Robin Lickley and Dr. Janet Beck for accepting to be the co-supervisors of this thesis. A special thank you goes to my examiners Pr Anne Wichmann and Dr Jocelynne Watson for their meticulous reading of my thesis and valuable recommendations.

I cannot thank my husband, Dr. Henni Ouerdane, enough, for his love, solid support as well as assistance with various technical aspects of my PhD work. This thesis would have been different without his sound knowledge of mathematics (useful to understand formulas of statistics), computing (I was introduced to the world of Linux and programming as well as typesetting with LaTeX etc...), and his ease to talk to people and persuade them to participate to my experiments. Being a multilingual and multicultural person, he also showed a great interest in my work which proved useful for discussions. I also warmly thank my friend Dr. Olga Gordeeva whose experience in data extraction was very useful in giving me a good start. My fellow

Abstract

The main objective of the research reported in the dissertation was to investigate the production of affective speech by monolingual and simultaneous bilingual children in Scottish English and French. The study was designed to address several important issues with respect to affective speech. First, possibility of identifying and comparing acoustic correlates of affective speech in productions of monolingual children was explored in a cross-linguistic perspective. Second, affective speech of bilingual children was examined in their two languages and compared to that of their monolingual peers. Third, vocal emotions encoded by monolingual and bilingual children were tested through the identification by French and Scottish monolingual adults. Five bilingual and twelve monolingual children were recorded for a cross-linguistically comparable corpus of affective speech. Children played four emotions (anger, fear, sadness and happiness) on one token utterance with the help of visual materials, which served as the reference of the expressed emotions, and as an affect inducing material. A large number of child speakers brings better understanding of cross-language and within-language variability in vocal affective expressions. The corpus was acoustically analysed and used in a cross-linguistic perception test with Scottish and French monolingual adults.

The results of the perception test support the existing view in the cross-cultural research on emotions: even if people from different cultural groups could identify each others' emotions, an in-group advantage was generally present. Another important finding was that some affective states were more successfully identified in one of the languages by the two groups of listeners. Specifically, French anger, as expressed by bilingual and monolingual children, was identified more successfully by both French and Scottish listeners than anger, encoded by bilinguals and monolinguals in Scottish English, thus suggesting that children showed some emotions more in one of the languages. The joint analysis of production and perception data confirmed the association of the studied acoustic correlates with affective states, but

also showed the variability of different strategies in their usage. While some speakers used all the measured acoustic correlates to a significantly large extent, other speakers used only some of them. Apparently, the usage of all the possible acoustic correlates is not obligatory for successful identification. Moreover, one of the studied affective states (fear) was characterised by more variable usage of acoustic correlates than others. Cross-linguistic differences were attested in the usage of some acoustic correlates and in the preferred strategies for the realisation of affective states.

Simultaneous bilingual children could encode affective states in their two languages; moreover, on average, their affective states are identified even better than those of monolingual children. This ability to successfully encode vocal emotions can be interpreted as a signal of high social competence in bilingual children. Production results show that all bilingual children realise some cross-linguistic differences in their affective speech. Nevertheless, interaction between the languages in the affective speech was discovered both in the production and perception data for bilinguals. This finding comes in support of other studies which call language interaction as a characteristic feature of bilingual phonetic acquisition. The specific pattern of the affective speech realisation is individual for each bilingual child, depending on the affective state and on the used language. In this context, the theory of integrated continuum, developed by Cook (2003), is discussed for its possibility to describe the paralinguistic organisation in the bilingual mind.

This thesis thus contributes to a better understanding of phonetic learning by monolingual and bilingual children in the context of affective speech. It also gives a detailed analysis of cross-language and within-language variability present in affective speech. This new data will be of interest to the researchers working in speech sciences, psycholinguistics, developmental and cross-cultural psychology.

Keywords: affective prosody, speech production, speech perception, emotions, monolingual acquisition, bilingual acquisition, English, French.

Contents

List of Figures

xvii

List of Tables

List of Abbreviations

1. Speakers

   - Bj: jth bilingual child, $j = 1, \ldots, 5$.

   - BSE: bilingual children in Scottish English.

   - BjSE: jth bilingual child in Scottish English, $j = 1, \ldots, 5$.

   - BF: bilingual children in French.

   - BjF: jth bilingual child in French, $j = 1, \ldots, 5$.

   - MSE: Scottish English monolingual children.

   - MSEj: jth Scottish English monolingual child, $j = 1, \ldots, 6$.

   - MF: French monolingual children.

   - MFj: jth French monolingual child, $j = 1, \ldots, 6$.

   - ASE: Scottish English actress.

2. Measurements

   - M: Measure.

   - Span: Pitch span.

   - Level: Pitch level.

   - rINT: Raw Pairwise Variability Index for intervocalic intervals.

   - PVI: Normalised Pairwise Variability Index for vocalic intervals.

   - Syl: Final syllable length.

   - Rate: Speech rate.

   - VQ: Voice quality.

   - Id: Identification level.

   - F_0: Fundamental frequency.

3. Others

- Sp: Speaker.

- A: Anger.

- F: Fear.

- S: Sadness.

- H: Happiness.

- L2: Second language.

- L1: First language.

- HA: High arousal.

- LA: Low arousal.

Chapter 1

Introduction

The question of how humans express their emotions through voice draws the attention of numerous researchers at the crossroads of speech and language sciences and psychology. Interest in vocal emotions continues to grow, not surprisingly, since emotions play a vital role in human communication.

Human communication is realised by means of both linguistic and paralinguistic behaviour. Affective communication, the communication of the speaker's current affective state (such as sadness, excitement, cordiality, guilt ...) is realised by *paralinguistic behaviour*, which is not linguistic and is non-verbal (not involved in verbal/word identification), and which is coded in speech, gesture, posture, facial expression, gaze and eye contact behaviour (Laver 1994). The main difference between linguistic and vocal paralinguistic behaviour is that linguistic behaviour uses the dual-level code of spoken language made up of the phonological and grammatical units. Linguistic behaviour has a characteristic sequential structure - grammatical units receive their distinctive shape by various linear combinations of phonological units. In contrast, paralinguistic communication has no meaningful structure, the sequential relationships may only facilitate judgments about the relative degree of a used feature (Laver 1994).

Even if emotions are at the core of each human's life, the question remains about their universality. Large scale cross-cultural studies have been performed to test the universality of emotion recognition in facial expressions (Ekman & Friesen 1971, Ekman 1994) and in voice (Scherer et al. 2001). These studies show that the recognition of emotions is largely, but not entirely, universal. Different cultural groups recognise intended emotions with better than chance accuracy. However, recent research (Elfenbein & Ambady 2003) has also documented evidence for a within-group

advantage, meaning that perceivers are generally more accurate at judging emotions expressed by members of their own cultural group, than by members of a different cultural group.

Gussenhoven (2004) has developed a biologically motivated theory of universal para-linguistic intonation meaning, according to which there exist three biological codes associated with universal paralinguistic meanings in speech; the Frequency Code (based on the size of larynx), this code was first proposed by Ohala (1983), the Effort Code (based on the production effort) and the Production Code (based on the regulation of the breathing process). Nevertheless, Gussenhoven (2004) stresses that the universal biological codes are conventionalised within speech communities, and their implementation shows language-related and culture-related differences. For example, communities may choose to use a particular interpretation to a particular degree. Recently this theory has been tested in an experimental study by Chen (2005). This research study supports Gussenhoven's claims on the universality of intonational meaning, but it also brings to light significant differences between speech communities in interpreting paralinguistic pitch variation.

While there is some evidence of cross-linguistic differences in adult affective production and perception based on monolingual and adult bilingual speakers (Mejvaldova 2001, Chen 2005, Yanushevskaya et al. 2006, Shochi et al. 2006), the realisation of these differences in child speech remains to be addressed. Moreover, existing studies of child affective speech are mostly based on perception and cover only children from a monolingual acquisitional context (Rothman & Nowicki 2004, der Meulen et al. 1997, Clément 1999). To the best of our knowledge, mastery of affective speech by bilingual children acquiring two languages from birth has not been examined.

Child affective speech production provides some of the most challenging and fascinating areas in affective speech research. The main challenges lie in finding an adequate and reliable methodology to study this area of affective speech, and the rewards lie in gaining understanding of how children express emotions in speech and

in deepening knowledge on how affective speech is realised in particular languages.

1.1   Research questions

The main objective of the present dissertation is to investigate the production of affective speech by monolingual and bilingual children in Scottish English and French cross-linguistically. Following the main objective, a cross-linguistic corpus of affective speech, encoded by early Scottish English and French bilingual children (i.e. who are bilinguals from birth) and their monolingual peers, was recorded. Based on the analysis of existing methods in adult affective speech research, the recording methodology was developed and adapted to the research of both monolingual and bilingual children, with special attention to ethical requirements, perception validity and cross-linguistic comparability of the recorded corpus. The corpus was acoustically analysed and validated by identification tests with monolingual Scottish and French adults. Based on the joint analysis of production and perception data, the following research questions were addressed:

- Is it possible to identify acoustic correlates of affective speech in child productions?

- Are there any cross-linguistic differences in the acoustic correlates of affective speech, used by monolingual children in Scottish English and French?

- How do bilingual children produce affective speech in their two languages? How do their choice and usage of acoustic correlates compare to that of monolingual children?

- How do monolingual French and Scottish adults identify affective speech, encoded by bilingual and monolingual children in the two languages?

The first two research questions of the dissertation address the possibility of identifying and comparing acoustic correlates of affective speech in productions of Scottish and French monolingual children. In this context, it is crucial to summarise

important methodological issues present in adult affective speech research. Research in adult affective speech aims at identifying acoustic correlates of specific affective states, so that the differentiation of emotions on the basis of the acoustic signal could be possible. This search for stable associations between acoustic correlates and emotions has not been very successful. A large number of acoustic parameters (e.g. pitch, voice quality, speech rate, loudness) have been identified as being involved in vocal affective expressions, but their usage appears flexible. It has been suggested (Murray & Arnott 1993) that emotion is communicated by a **subtle combination** of features both at segmental and suprasegmental levels of speech description.

The suprasegmental level considers all factors that can be potentially prolonged beyond the domain of the segment, e.g. voice quality, pitch, loudness, rhythm, rate (Laver 1994). However, this division between segmental and suprasegmental levels is not always evident, as some phonetic correlates, e.g. stress, intonation, are often observed at the segmental level. For example, vowel reduction is strongly correlated with word stress in English. Researchers (Pfitzinger 2006, Campbell & Mokhtari 2006) studying these phenomena, prefer to use *prosody* instead of *suprasegmentals*, as a more general term covering both segmental and suprasegmental realisations of intonational, metrical, temporal speech organisation and voice quality settings. The present study uses the term *prosody* in this perspective; *affective prosody* is used to define those prosodic features playing the role of vocal affective cues in speech. Note that this definition of prosody deviates from a more traditional usage of terms *prosody* and *prosodic analysis* only in relation to intonation (see reference Laver (1994)).

Three possible reasons were advanced by Johnstone & Scherer (2000) as to why acoustic parameters that differentiate emotions have not yet been consistently found. First, a small number of relatively simple acoustic parameters have been measured in most studies. For example, statistical approach to pitch measurement is widely used. Second, many studies do not specify the type of arousal in studied emotions, for example hot and cold anger, while emotions of different arousal type may have

specific vocal profiles. Finally, Johnstone & Scherer (2000) note that very little is known about the effects of individual and cultural differences on affective speech production. It is common to use a small number of speakers and to factor the effects of individual speakers out of the analysis as an unwanted source of variance. Very few cross-linguistic studies have been carried out on emotion encoding in the voice.

Assuming that affective speech is characterised by individual variability, a cross-linguistic comparison requires an increased number of speakers per language. This importance of distinguishing between individual and cross-linguistic variability motivated the recording of a higher number of speakers per language for this dissertation than is the usual practice in production studies, even if it required significantly more time for annotation and acoustic analysis. Moreover, it was decided to measure pitch on the basis of linguistically important variation in the utterance, such as accent peaks, post-accent valleys and sentence final lows, and not on statistical estimations, like F_0 mean and standard deviation, difference between maximum and minimum F_0, etc. It was argued by Johnstone & Scherer (2000) that the measure of F_0 mean does not give much information on pitch variation, as a high F_0 mean may result from a high F_0 level, or a wide F_0 range, or both. The results of Mozziconacci (1998) and Patterson (2000) show that a linguistically motivated analysis gives better and perceptually relevant results than the more widely used statistical approach. The great popularity of the statistical approach is due to it being less time and work consuming, as it is automatic. Linguistic analysis has to be performed manually. Not to preclude a finer level of emotion differentiation, it was decided to take the challenge and use the linguistic analysis for a large cross-linguistic corpus of affective speech.

The third research question concerns the production of affective speech by simultaneous bilingual children. There are several reasons for the inclusion of bilingual children. First, the usage of bilinguals contributes to the cross-linguistic analysis of affective speech production in the two languages. As recognised by some scientists (Nicoladis 2001, Paradis 2000, Genesee 1989), bilinguals may act as their own

controls on many variables, for example, such as personality, temperament and cognitive abilities, which are very important in affective expressions. Each monolingual child has his or her own personality, and it is not possible to achieve such a perfect matching as one bilingual in the two languages. Moreover, time consuming acoustic analysis puts limits on the number of speakers which is possible to include in the corpus. In this light, the usage of bilinguals looks particularly attractive in the context of the present cross-linguistic production study. Affective speech analysis of bilingual speakers could help to investigate differences between individual and cross-linguistic variability at a different angle.

Second, taking into account that research on bilinguals is still scarce, it could enable filling the existing gap and address important questions specific to bilingual speech acquisition and models of bilingual speech production. Bilingual research is preoccupied with the question of how two languages are organised in one mind - as one fused system or as two differentiated systems. Recent evidence comes in support of the differentiated representation of the two languages in syntax, pragmatics and phonology. Research also acknowledges that bilingual speech acquisition has its specific characteristics, such as the presence of language interaction and somewhat slower rate of development in both prosodic and segmental phonology in comparison to monolingual children. The simultaneous exposure to two languages apparently results in the interactions between two differentiated but not entirely autonomous language systems.

Prosody plays both linguistic and paralinguistic functions in speech. Thus, prosody contributes to the two parallel speech channels, paralinguistic and linguistic, which are tightly coordinated in time and realised on the same acoustic signal. While linguistic and vocal paralinguistic cues are carried on the same acoustic signal, paralinguistic cues are regarded as *'modifications of the way in which phonological categories are realised'* (Ladd 1996). There exist studies on the linguistic prosody of bilinguals (Whitworth 2003, Gordeeva 2006), but to the best of our knowledge, affective prosody of simultaneous bilinguals has not yet been addressed. The ques-

tion remains whether bilingual children differentiate affective speech in their two languages. Moreover, it is important to investigate whether language interaction is also present in the acquisition of affective prosody by bilingual children. The answers to these questions may be specifically important for theories of phonetic learning, and may have implications for second language acquisition theories in general.

The fourth research question concerns the identification of child affective speech by monolingual French and Scottish adults. The identification test was designed to validate the recording methodology and to test if adults are able to identify emotions enacted by children with the help of visual materials. Moreover, the results of the identification test also help to address the second and the third research questions regarding the cross-linguistic differences in child affective speech production and the differentiation of affective speech by bilinguals. Perception results can allow us to test whether monolingual adults are more successful in the identification of emotions encoded by children in their own language than in an unknown language, thus showing an in-group advantage. Perception results can also help to see whether there is any difference in the perception of monolingual and bilingual children of the same language. The identification test gives a better insight into how bilingual children integrate/separate their affective speech in the two languages. Moreover, the joint analysis of production and perception data can help to distinguish between individual and cross-linguistic variability, as encoding strategies used by individual speakers may be analysed for the ease of identification within and across the two languages.

1.2 Thesis layout

The thesis consists of eight chapters, and their description is briefly presented. Due to the multi-disciplinary and largely unresearched subject of the present dissertation, the theoretical background (Chapters 2-4) and research methodology (Chapter 5) were grounded in the more developed related research areas, such as child emotional development, adult affective speech production, and bilingual speech acquisition.

- Chapter 1 introduces the research area of the present study, the main objective of the study and its research questions.

- Chapter 2 presents existing theories of emotions in their variability. Theories applied in speech affective research receive a more detailed description. Special attention is given to the presentation of child emotional development in a cross-linguistic perspective. The milestones of child emotional development are briefly described in connection with cognitive development, and child individual differences in development are specified.

- Chapter 3 presents an overview of acoustic parameters involved in affective expressions through voice. The theory on universality and language specificity in pitch variation developed by Gussenhoven is presented. Production and perception studies on affective speech are reported, with particular emphasis on cross-linguistic and developmental studies.

- Chapter 4 presents theories of bilingual speech production and phonetic learning. Finally, monolingual and second language studies of affective phonetic acquisition are reported.

- Chapter 5 first presents existing methods of affective speech recording and their analysis in light of child affective speech, then the methodology, specifically developed for the present study, is described. Speech digitisation, annotation and data extraction are addressed in detail. All acoustic parameters measured in the study are given with their formulas.

- Chapter 6 describes the methodology and results of the perception experiment, serving to validate the recorded corpus and to answer the research question on the perception of monolingual and bilingual children by adults.

- Chapter 7 presents the production results. Group results are compared across monolingual and bilingual children. Then, individual results of each speaker

(monolinguals, bilinguals and adults) are reported and analysed. Finally, the joint analysis of production and perception data is reported.

- Chapter 8 summarises the results and gives general conclusions.

The thesis ends with Appendicies, including examples of annotation, statistical results and visual materials.

Chapter 2

Theoretical background on emotion

2.1 Introduction

This chapter presents general foundations in the theory of emotion. This overview is not intended to be exhaustive; its main purpose is to describe concepts in emotion research relevant to the present study. As the main objective of the dissertation is to investigate child affective speech production in a cross-linguistic perspective, it was felt vital to cover the following subjects; definition of emotion, child emotional development and theories of emotion applied in affective speech research. The description of milestones of child emotional development is relevant, as vocal emotions, the subject of the present study, are integrated in this process. Studies of child individual and cross-linguistic differences in emotional development are presented. The final emphasis is given to the theories applied in affective speech research. The methodology of the present study is based on the theory proposed by Scherer Scherer et al. (2001), thus its presentation is especially important.

2.2 Theories of emotions

Research into emotion is at the crossroads of many disciplines; psychology, philosophy, anthropology, neurology, sociology, culture and, of course, linguistics. Thorough overviews of theory and empirical research on emotion are provided by Strongman (1996), Lewis & Haviland (1993), Andersen & Guerrero (1998). The complexity of the subject explains the large diversity of theories and the difficulty of reuniting the existing knowledge on emotion into a complete theory. There is no consensus on the definition of emotion, as it can be approached from different directions, with different methods and within different disciplines. Generally, scientists give a definition of emotion in the framework they choose for their studies. A general description

of the main theoretical approaches in emotion research, relevant for the present work, is provided in this chapter. Special attention is paid to developmental and cross-cultural theories.

Behavioural theory: Theorists, such as Staats & Eifert (1990), usually regard emotion as a response or a large class of responses, basic to life and survival, rather than as a state of the organism. They define emotion in terms of operations believed necessary to bring it about. **Physiological theory:.** The major aim of this theory Plutchik (1993), Panksepp (1993), Scherer (1993) has been to find the substrate of emotion in the central nervous system, the peripheral nervous system and the endocrine system. The physiological theories of emotion rest on the beliefs that emotions have a biological base. At the complex human level, they might have socially constructed aspects, but even these are only additions to physical/biological foundation. **Cognitive theory:** While some cognitive theorists address profoundly the nature of the relationship between affective and mental processes, others study emotion from a broad perspective which still gives an integral part to cognition. The appraisal theorists assume that cognitions cause physiological and behavioural change, and that it is important to study the change in order to gain knowledge of the cognitions. Schechter, Frijda, Oatly and Johnson-Laird are among the most influential cognitive theorists Schachter & Singer (1962), Frijda (1993), Oatley & Johnson-Laird (1987). **Social theory:** Emotion is often conceptualised as a social phenomenon. For the most part the stimuli for emotional reactions come from other people and emotion occurs in the company of others. For many years social psychologists focused on emotional expression and its recognition. More recently social psychologists have turned their attention to other aspects of emotion, and have begun to explore emotion in relationships, in attitudes, in group settings, in social climates, and so on. For example, see references Heise & O'Brien (1993), Camras et al. (1993). **Cultural theory:** The study of human emotions is one of the priorities in cultural psychology. The cultural perspective assumes that psychological processes are not just 'influenced', but are thoroughly culturally constituted. Thus,

the characteristics, development, and functions of emotions are largely culturally determined Shweder (1993), Wierzbicka (1992, 1999), Russell (1991).

2.3   Child emotional development

Theories of emotional development deal primarily with both how emotions develop and how they influence the course of development. Lewis Lewis & Haviland (1993) has considerably contributed to our understanding of emotional development. He proposes a model of emotional development which is based on the conviction that most of the adult emotions have appeared by the age of three, even though others might emerge later, or the existing ones become elaborated. Lewis's view is that emotional development and cognitive development are integrally linked.

Piaget and Vygotsky were two of the most influential theorists of child cognitive development. Swiss child psychologist **Jean Piaget** Piaget (1968) proposed that cognitive development is a dynamic process that results from an individual's ability to adapt thinking to the needs and demands of a changing environment. See the following source for detailed reviews of his work Wadsworth (2004). While Piaget viewed cognitive development as an individual accomplishment shaped, in part, by environmental factors, he devoted little attention to the social or cultural context. This became the focus of another pioneer in cognitive development, the Soviet psychologist **Lev Vygotsky**. Vygotsky (1981) referred to his theory as cultural-historical, stressing that the factors determining the individual's life activity were produced by the historical development of culture:

> Any function in the child's cultural development appears twice, or on two planes. First it appears on the social plane, and then on the psychological plane. First it appears between people as an interpsychological category, and then within the child as an intrapsychological category.

He suggested that the process of cognitive development involved three major components: the use of language, the role played by culture, and a child's zone of

proximal development (ZPD). ZPD is the distance between a child's *'actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers'* Vygotsky (1978). The notion of ZPD emphasizes Vygotsky's view that social and cultural influences contribute significantly to children's development of cognitive abilities. For more details see Daniels (1996), Richardson (1998), Palmer (2001), Ratner (2004). Vygotsky's theory is receiving increased attention among developmental psychologists, especially those working in a cross-cultural perspective Gardiner (2001).

Holodynski & Friedlmeier (2006) suggest that the sociohistorical approach of Vygotsky (1978) can serve as a developmental framework for cross-cultural studies in the domain of emotions. They argue that central components of emotions - expression, feeling, and emotion regulation - are to a large extent culturally defined, and their knowledge is transmitted in the daily interactions between the child and its caregivers. Culture determines the frame of the specific course and content of the interactions as a learning field for the child. They believe that growing capacities and abilities of a child are culturally-based capacities, rather than inborn capacities. The effects of culture can be derived from the understanding of specific interactive processes and the child's related internalisations/externalisations of socially mediated knowledge and messages.

Holodynski & Friedlmeier (2006) also suggest that **emotional expression** is part of human symbolic activity, analogous to human language. They say that expressive signs are used as symbols because they do not possess naturally inborn signal character, but can be removed from the referent. It is possible to see the symbol function of emotional expression in the variety of interpretations of expressive signs, and especially in the ability to dissociate expression and experience (the expression does not necessarily correspond with the experience). The reaction of another person to one's emotional expression primarily depends on the meaning that he/she gives to the expressive signs in a specific context. The symbol function is

also evident in the display rules, i.e. regulating expressive signs according to certain conventions. **Emotional experience** is described as a conscious process that requires a learning history for children. The person perceives the arousal of the vegetative nerve system (e.g. increase of heartbeat), as well as proprioceptive reactions (e.g. the feedback of the facial and body expression), and has to structure these processes. The child learns to recognise and structure the arousal in a categorical way and label it through interactions with the caregivers. **Emotional regulation** includes two hierarchical levels - instrumental and reflexive. Instrumental regulation suggests the action readiness that is inherent in emotions, like fleeing in fear. However, there are contexts where it is not possible to act according to this strategy, based on problem-focused regulation. The reflexive regulation may be used in its place, like self-soothing, distraction, reappraisal of the event. These strategies aim to change the internal state and can be seen as emotion-focused regulation.

The theory of Vygotsky, adapted and developed by Holodynski and Friedlmeier, looks very promising for cross-cultural developmental studies. Several experi-mental studies have recently been conducted on the emotional development in a cross-cultural perspective.

Cole et al. (2002) report cultural differences in children's emotion expression for two ethnic groups of school-age children in Nepal and the USA. The Nepali subjects were from two villages inhabited by two different social groups, Brahman and Tamang. Though the villages have a similar location and people all live in large family groups, there are also significant differences between them. Brahmans represent a very high caste associated with pride and status. There is a large difference with regard to the behaviour towards in- and out-group members. It is acceptable to show anger towards out-group members. Respect for authority is expected within the in-group, since dominance would threaten the harmony. In contrast, the Tamang, an indigenous group in Nepal of Mongolic origin, have what can be viewed as a low societal status. Equality is a central value. Material possession is shared, being unselfish is an ideal, and showing anger to either in- or out-group members is

generally avoided. The study shows that clear culture-specific patterns with regard to feeling and emotion regulation already occur in school-age children, reflecting cultural expectations of their cultural group. Tamang children displayed more shame and less control, Brahman children more anger and higher control. Both groups of Nepalese children showed a higher preference for emotion-focused regulation strategies that aimed at changing their own internal state, as compared with the US children, who reported more problem-focused justifications, and related regulation behaviour by changing the situation. In the long run, regulation strategies used by Nepalese children may lead to a decrease of negative feeling in future similar situations, a consequence that again fits in with cultural expectations.

Friedlmeier & Trommsdorff (2002) observed 2- and 5-year-old Japanese and German girls and their mothers, to test the extent of distress reactions in a disappointment situation, as well as the relationship between the course of emotions and interactive regulation with the mother. The course of negative emotion did not differ between the 2-year-old Japanese and German girls, but between the 5-year olds it did. It seems that different developmental paths concerning the expression and regulation become more apparent in preschool age. The German girls of sensitive mothers displayed a stronger negative expression and did not recover over the observed period as compared to the girls of less sensitive mothers, whereas the Japanese girls demonstrated an opposite pattern. Japanese mothers showed a preference for emotion-focused regulation. They comforted and distracted the child more often than the German mothers did. They ignored the cause and created a situation for the child 'as if nothing had happened', i.e. the child's focus was not oriented towards her own feeling. In contrast, German mothers showed strong problem-focused regulation. They commented on the cause directly and verbally, talked about solutions and the child's emotion, i.e., the child's focus was oriented towards her own inner experience. The results yielded culture-specific differences in regard to maternal sensitivity and its relation to children's negative emotional expression and regulation. In both cultures, the more sensitive mothers promoted their children's ability

to develop the culturally appropriate way to express and regulate negative (self- and other-focused) emotions.

The study of Tsai & Chentsova-Dutton (2003) presents interesting evidence in support of the theory developed by Holodynski and Friedlmeier. Tsai and Chentsova-Dutton compared the emotional facial expressions from two groups of adult European Americans (EA) several generations removed from their ancestors: EA originally from Scandinavian countries (EA-S), who value emotional control, and EA from Ireland (EA-I), who value emotional expression. The authors discovered that European Americans varied as a function of their countries of origin. EA-S were less expressive than EA-I while reliving various emotions, especially happiness and love, suggesting that in this domain, European Americans continue to be influenced by their cultural heritages. The study shows that emotion regulation specific to a particular culture may be transferred in the family context through several generations after immigration.

Holodynski & Friedlmeier (2006) claim that emotions and emotion regulation develop and become more and more integrated into higher levels of behaviour regulation. It means that the emotional behaviour goes from a universal starting point to an increasingly culturally mediated form. They underline that developmental perspective requires the assumption of immediacy of culture in order to describe, explain and predict the processes by which cultural knowledge is transmitted to the developing child.

2.3.1 Stages of emotional development

The research into emotional development starts from newborn babies. Special coding systems (Izard 1979, Izard et al. 1983) are used to analyse babies' facial expressions. A wide range of facial expressions is present in newborn babies (Camras et al. 1991), but their elicitation is not yet analogous to that of adults: disgust is observed as a reaction to bitter substances, first smiles are produced in a paradoxical sleep, thus these first facial expressions are primary linked to reflexes. Other facial expressions,

such as anger, fear and sadness, are produced mainly randomly, and a considerable number of negative facial expressions are not differentiated by adult listeners (Camras 1991).

The differentiation of negative emotions expressed by children becomes possible only in the second half of the first year. In the first eight or nine months of life, children's emotional behaviour reflects the emergence of the six early emotions; joy, surprise, sadness, disgust, anger and fear. In the second half of the second year, the child gains a particular cognitive ability, the development of a concept of self. It is only then that children both can be in a particular emotional state and can be said to experience that state. The production of certain states, e.g. embarrassment, empathy, envy, requires self-awareness, therefore these emotions are unlikely to occur until this cognitive process emerges (Camras 1992, Kopp 1993).

The next cognitive milestone occurs between 2 and 3 years of age, when children acquire a capacity to evaluate their behaviour against a standard. This ability gives rise to so-called 'self-conscious evaluative emotions', e.g. pride, shame, guilt. Thus complex social evaluative emotions make their appearance at about 3 years of age (Harris 1993). The emotional life of children continues to be elaborated and to expand, but the basic structures necessary for this expansion have already been formed.

The next developmental stage appears between 6 and 10 years old when children acquire metacognitive abilities. At this period, children become capable of hiding their emotions (Saarni 1984). Children learn to recognise, analyse and speak about their emotional experiences. This capacity is called emotional intelligence.

2.3.2 Individual differences in emotional development

Temperament is one of children's individual differences which remains stable in time and in various situations. This factor is important for our understanding of how emotional styles emerge. Temperament has a neurophysiological and partially hereditary basis (Emde et al. 1992, Goldsmith 1993). It is suggested by Goldsmith &

Campos (1982) that temperament is based on the innate structure which organises emotional expression.

The studies of temperament show that children are born with a predisposition to react in a particular way to events. For example, some children get angry quite easily, some are more frightened and others are generally happy. The research of Caspi et al. (1987) shows that emotional configurations change. It is also evident that the emotional bases people have from birth make some developmental paths more probable for one person than for another. These emotional bases influence the manner of reaction in relation to others and the opportunities a particular individual has in life.

Even if hereditary factors and temperament play their role, the capacity to identify and understand emotions is closely linked to the social experience of the child, i.e. to the interactions experienced in the family, especially with the mother, and outside the family with their peers and other adults (Denham 1997). A longitudinal study by Dunn & Brown (1991) shows that the child's capacity to understand basic emotions at the age of 3 depends on their family experience, and that the same variation is observed in their understanding of complex emotions at the age of 6.

2.4 Theories of emotions applied in affective speech research

Three emotion theories have strongly influenced affective speech research: discrete theory, dimensional theory and componential theory.

The first approach is based on the discrete emotion theory, proposed by Charles Darwin, and then developed by Ekman (1992) and Izard (1971). This approach suggests the existence of a number of basic or discrete emotions that are characterised by very specific response patterns in physiology, as well as in facial and vocal expression. Many studies in affective speech research () have followed this theoretical model and investigate the vocal characteristics of happiness, sadness, fear, anger and surprise. The application of the concept that there exist a few basic uniform emotions has not resulted in concrete predictions for specific vocal patterns of emo-

tions. Recently, the concept of emotion families (Ekman 1992) has integrated the fact that basic emotions are not completely homogeneous, there are different types of each specific emotion. For example, anger may be expressed as explosive (hot) anger or as controlled (cold) anger. Vocal patterns apparently depend on the type of expressed emotion. Nevertheless, many researchers still neglect this phenomenon.

The second approach is based on the dimensional theory of emotions. It describes emotions in a two- or three-dimensional space. The names used for the emotion dimensions are activation (or arousal), evaluation (also pleasure or valence), and power (or control). Researchers using this model (Bachorowski 1999, Cowie et al. 2001, Schröder 2004) investigate negative *versus* positive, active *versus* passive differences in the vocal expressions of affective states. This approach is often criticised for its simplification and reduced generalisation of affective states (Lazarus 1991).

The third approach, proposed by Scherer et al. (2001), presents a componential theory of emotion. This model describes different emotional states in their variability, as produced by different types of appraisal patterns. Scherer (2000) has suggested the distinction of affective states by the following components: the relative intensity and duration, the degree of synchronisation of organismic subsystems, the extent to which the reaction is focused on an eliciting event, the degree to which appraisal has played a role in the elicitation, the rapidity of state changes, and the extent to which the state affects open behaviour (see Table 2.1).

Based on these components, affective states are described in the following way by Scherer (2003):

- *Emotion: relatively brief episode of synchronised response of all or most organismic subsystems in response to the evaluation of an external or internal event as being of major significance (angry, sad, joyful, fearful, ashamed, proud, elated, desperate);*

- *Mood: diffuse affect state, most pronounced as change in subjective feeling, of low intensity but relatively long duration, often without apparent cause (cheer-*

ful, gloomy, irritable, listless, depressed, buoyant);

- *Interpersonal stances: affective stance taken toward another person in a specific interaction, colouring the interpersonal exchange in that situation (distant, cold, warm, supportive, contemptuous);*

- *Attitudes: relatively enduring, affectively coloured beliefs, preferences, and predispositions towards objects or persons (liking, loving, hating, valuing, desiring);*

- *Personality traits: emotionally laden, stable personality dispositions and behaviour tendencies that are typical for a person (nervous, anxious, reckless, morose, hostile, envious, jealous).*

Table 2.1: Design feature delimitation of different affective states. 0: low, $+$: medium, $++$: high, $+++$: very high, $-$: indicates a range. This table is adopted from Scherer Scherer (2000).

Components	Emotion	Mood	Interpersonal stances	Attitudes	Personality traits
Intensity	$++-+++$	$+-++$	$+-++$	$0-++$	$0-+$
Duration	$+$	$++$	$+-++$	$++-+++$	$+++$
Synchronization	$+++$	$+$	$+$	0	0
Event focus	$+++$	$+$	$++$	0	0
Appraisal elicitation	$+++$	$+$	$+$	$+$	0
Rapidity of change	$+++$	$++$	$+++$	$0-+$	0
Behavioural impact	$+++$	$+$	$++$	$+$	$+$

Emotions constitute a special group among various types of affective states. They are more intense and of shorter duration, as shown in the Table above 2.1. In particular, they are characterised by a high degree of synchronisation between the different subsystems. Furthermore, they are likely to be highly focused on eliciting events and produced through cognitive appraisal. In terms of their effect, they strongly impact behaviour and are likely to change rapidly. Another issue that must not be neglected is the existence of powerful regulation mechanisms as determined by display rules and feeling rules. Such regulation efforts are particularly salient in the case of very intense emotions, such as hot anger, despair and strong fear (Scherer 2000).

The componential model developed by Scherer et al. (2001) is gaining influence in affective speech research. It is much more open with respect to the kind of emotional states that are likely to occur, given that there is no limited number of relatively fixed neuromotor patterns, as in discrete emotion theory. Another advantage of the componential model is that it does not limit the description to two basic dimensions. It emphasises the variability of different emotional states, as produced by different types of appraisal patterns. Furthermore, it permits us to model the distinctions between members of the same emotion family (Scherer 2003), which is an important issue in affective research.

2.5 Conclusion

Based on the research in cross-cultural and developmental psychology, child emotional development has been presented with special attention to cross-cultural and individual differences. The dominant view on emotional development suggests that it is integrally linked with cognitive development. At the same time, the importance of cognition does not exclude physical/biological foundations of emotion: temperament also has its influence on emotional development and determines some individual variability in emotion expression and regulation. This review comes in support of the decision to use bilingual children in the present cross-linguistic study of child affective speech. As has been stated in Chapter 1 'Introduction', bilingual children may act as their own controls on such important variables for emotion expression as personality, temperament and cognitive abilities.

Cross-cultural studies of child emotional development have recently emerged (Cole et al. 2002, Friedlmeier & Trommsdorff 2002). They suggest that central components of emotions, i.e. expression, feeling and emotion regulation, are to a large extent culturally defined, and that they are acquired in daily interactions between the child and its caregivers. Moreover, it has been shown (Tsai & Chentsova-Dutton 2003) that some specific cultural patterns of emotion regulation may be passed even through several generations after immigration to a new country. The specific role

of culture in the development of vocal emotions remains to be investigated, but one may assume that it is crucial.

Three theoretical approaches, applied in the studies of vocal emotions, have been presented: the discrete emotion theory, the dimensional theory and the componential theory of emotion. The componential model of emotion proposed by Scherer et al. (2001) is adopted in this study and applied in the methodology, see Chapter 5 'Methodology' for further details. The following theoretical chapter will describe the existing research on vocal emotions.

Chapter 3

Vocal affective communication

3.1   Introduction

This chapter gives an overview of acoustic parameters involved in vocal affective expressions. Their universality and language specificity are discussed in the frames of existing theoretical approaches. Production and perception experimental studies on affective speech are reported with particular emphasis on cross-linguistic studies, both in first and second language contexts.

3.2   Acoustic correlates of vocal emotions

Emotion is communicated through voice by a variable combination of acoustic parameters. The search for vocal signatures of emotions, i.e. stable associations between a group of particular acoustic parameters and specific affective states, has not been successful. A large number of acoustic parameters (e.g. pitch, voice quality, speech rate, loudness) have been identified as being involved in vocal affective expressions, but their usage allows considerable flexibility for speakers. This variability in the affective production does not apparently interfere with the recognition of affective states which remains quite efficient even in unknown languages (Scherer et al. 2001).

Johnstone & Scherer (2000) have reviewed previous studies (see Murray & Arnott 1993, for a detailed analysis) with respect to the acoustic patterns that characterise the vocal expression of studied emotions. The authors link much of the consistency in the findings to differential levels of arousal or activation for the target emotions. Their theoretical approach (Scherer et al. 2001), describing different emotional states in relation to different types of appraisal patterns, has been presented in Chapter 1. Note that the summarised results are based on statistical methods of

analysis, i.e. F_0 mean, F_0 maximum and minimum, F_0 standard deviation, which are predominantly used in the existing studies. Methods of acoustic analysis will be discussed later in this chapter and in Chapter 4 'Methodology'.

- **Anger**: In general, an increase in mean F_0 and in mean intensity has been found in angry speech. Johnstone & Scherer (2000) also report that some studies find increased F_0 range and variability. They suggest that these studies may have measured hot anger or rage, while other studies, not finding these characteristics, may have measured cold anger or irritation. F0 contours tend to be directed downward and speech rate tends to be generally increased for hot anger.

- **Fear**: High arousal fear has been reported with increased intensity, mean F_0, and F_0 range, although Banse & Scherer (1996) found a decrease in F0 range. Speech rate is also reported to be high. Low arousal realisations, i.e. anxiety, also show faster articulation, but data on other varieties are less consistent: both an increase and decrease in mean F_0 and F_0 range have been reported. Lower intensity for anxious speech is also reported.

- **Sadness**: For sadness, decreases in mean F_0, F_0 range and intensity are usually found (although Banse & Scherer (1996) reported increased F_0 range), as well as downward-directed F_0 contours. Speech rate also decreases for sadness. Most studies reported in the literature seem to have studied low arousal forms of sadness. High arousal forms, such as desperation, are realised with increased intensity and F_0 level (or bottomline).

- **Joy**: This emotion is most often studied in the form of elation, rather than enjoyment or happiness. For high arousal happiness, increases in mean F_0, F_0 level and intensity are reported. F_0 range and variability are also found to increase. There is also inconclusive evidence for an increase in speech rate. Quieter forms of the emotion, such as contentment, seem to be characterised by relaxed vocal settings, leading to low mean F_0, low F_0 level, lower intensity

and slower articulation.

This brief review of acoustic patterns for four widely studied emotions, anger, fear, sadness and happiness, shows that contradictory results are reported for the usage of some acoustic correlates, like increased and decreased F_0 range for fear. Possible reasons for inconsistent, and sometimes even contradictory, results in the acoustic analysis of affective speech were suggested by Johnstone & Scherer (2000) and mentioned in the chapter 'Introduction'. To summarize, they say that few acoustic parameters are measured in most studies, and that these measurements tend to be relatively simple. Moreover, arousal types of recorded emotions, as well as individual and cultural variability, are rarely taken into account in production studies. The following subsections will present research on the usage of voice quality and intonation in production and perception of vocal emotions in a more detailed way. Special emphasis will be given to studies which report cross-linguistic and individual differences in affective speech.

3.2.1 Voice quality

Voice quality is attested as an important acoustic correlate of affective expressions by many researchers (Murray & Arnott 1993, Scherer 1986, Laver 1994). Nevertheless it is one of the most neglected parameters due to several difficulties, notably variability among researchers at the definition level, and difficulty of obtaining appropriate measures of the glottal source at the methodological level.

Laver proposed a classification system of voice quality, based on available physiological and acoustic data, and defined three major phonation types; voicelessness, whisper and voicing (Laver 1994). Voicelessness is divided into nil phonation and breathiness. Voicing is divided into normal voicing, creak and falsetto.

Normal voicing shows regular, efficient vibration of the true vocal folds without audible friction (Laver 1994).

Breath phonation describes the condition of slightly turbulent flow through the widely abducted glottis.

Whisper is characterised by a more turbulent airflow which is caused by a much smaller open area of the glottis. Strictly speaking, breath and whisper phonation are related to the same phenomenon of decreasing width of glottal opening and parallel constriction increasing. While the glottis adjustment for breathiness is typically set to between 60 and 95 % of the maximum possible glottal area, the adjustments for whisper are often less than 25 %.

Falsetto is characterised by longitudinally stretched vocal folds with particularly thin-edged glottal margins and by a slightly open glottis. As a result, the top of the fundamental frequency range is markedly higher than in normal voicing: the bottom end of the falsetto range overlaps with the top end of the voiced phonation range. Falsetto is also often accompanied by a slight whisperiness.

Creak phonation is characterised by a very low frequency, irregularly spaced in time. Its glottal adjustment varies across speakers, from loosely pressed vocal folds (Catford 1977) to relatively high level of adduction. Creak is able to occur in simultaneous combination with voicing, which gives the compound phonation of creaky voice.

The quality of overall laryngeal tension makes audible changes in phonation. When the whole laryngeal musculature is more relaxed than in modal voice, then a *lax setting* is observed. The level of relaxation goes from slightly *whispery voice* (small degree of relaxation) to *breathy voice* (with extreme relaxation). When the muscle laryngeal tension is increased, a *tense setting* is observed, the level of tension can go from *tense voice* (small degree of tension) to *harsh voice* (extreme hypertension).

In the acoustic analysis, breathy voice is often accompanied by friction using a high rate of air flow and a decrease in overall acoustic intensity and noise in spectrum. A creaky voice has a characteristic low fundamental frequency with aperiodic glottal pulses. Harsh voice quality is also characterised by aperiodic noise, but a normal fundamental frequency distinguishes it from a creaky voice quality. Whisper is characterised by the absence of the fundamental frequency. Falsetto usually has an

extremely high fundamental frequency.

Gobl & Chasaide (2003) examined the role of voice quality in the communication of emotions, moods and attitudes. They tested Irish English listeners' reactions to an utterance synthesised with different voice qualities: harsh voice, tense voice, modal voice, breathy voice, whispery voice, creaky voice and lax-creaky voice. The seven stimuli were used in a set of perception tests to elicit listeners' responses in terms of pairs of opposing affective attributes. The differentiation appears to be in terms of activation, and, to a lesser extent, power, for the used voice quality stimuli. The attributes associated with the tense/harsh stimuli have high activation and/or high power, but include affects with positive (confident, interested, happy) and negative (angry, stressed) valence. The other, non-modal group of stimuli - the breathy, whispery, creaky and especially the lax-creaky voiced stimuli - are associated with attributes which have low activation, but both positive (relaxed, content, intimate, friendly) and negative (sad, bored) valence.

Burkhardt & Sendlmeier (2000) report that falsetto voice is associated with fear, tense voice is associated with anger, falsetto and breathy voice are linked with sadness. Other studies (Scherer 1986, Laukkanen et al. 1996, Gobl & Chasaide 2003) also show that breathy voice is associated with sadness. Gobl & Chasaide (2003) report on the association of lax-creaky voice with sadness. The association of tense voice with anger is strongly supported in many studies (Scherer 1986, Laukkanen et al. 1996, Burkhardt & Sendlmeier 2000, Gobl & Chasaide 2003). Laver (1994) suggests that harsh voice, extreme hypertension, is associated with anger and aggression.

Gobl & Chasaide (2003) stress that voice quality varies in a continuous way. The same type of voice quality may be present in various degrees, from slightly audible to extreme. The exploration of the full voice-quality continuum in communication of affect is still not known. Another important observation made by Gobl & Chasaide (2003) is that some affective states, like anger, are strongly associated with a particular voice quality for all perceivers, while others, like fear, show much more

individual variability in perception.

Research (Scherer 1986, Laukkanen et al. 1996, Gobl & Chasaide 2003) suggests that voice quality alone may evoke affective associations. Nevertheless, these associations do not exist on a one-to-one basis, specific voice qualities are apparently associated with a group of related, or even unrelated, affective states. Voice quality permits a considerable flexibility for its interpretation, but as it is only one of the acoustic cues for affective communication, it is not used in isolation.

3.2.2 Intonation

The auditory correlate of the frequency of vibration of the vocal folds (fundamental frequency or F_0) is the *pitch* of the voice. The pitch of the voice is present in speech in its continual variation, and constitutes the basis of **intonation** (Laver 1994). There exist several theoretical models developed for the linguistic analysis and description of intonation, for example, superpositional models and tone-sequence models. Superpositional models, i.e. the experimental-phonetic approach to intonation (IPO) developed for the Dutch language ('t Hart et al. 1990), consider F0 curves as the superposition of pitch movements on a declination line. Tone-sequence models, i.e. the Tones and Break Indices system (ToBI) (Pierrehumbert 1980), describe pitch as a sequence of tones on linguistically defined targets (accented syllables and intonative boundaries), ToBI is a widespread standard for labelling intonation. Linguistic models of intonation are developed to describe linguistic functions of intonation, and only a small number of studies have applied linguistic models to the description of affective intonation (Stibbard 2000, Mozziconacci 1998).

Intonation patterns

Mozziconacci (1998) studied the role of intonation patterns in conveying emotions and attitudes in Dutch. Intonation patterns produced by three speakers were described according to the experimental-phonetic approach to intonation (IPO) developed for Dutch ('t Hart et al. 1990). The IPO model of intonation is a two-

component model treating whole contours as combinations of straight-line segments, each segment corresponding to a single pitch movement. In this model, the end point of the declination line represents the pitch level, while the excursion size of the pitch movements represents the pitch range.

The distribution analysis of intonational patterns over emotions showed that identified patterns were not equally distributed over the seven affective states studied (neutrality, joy, boredom, anger, sadness, fear and indignation). A prominence-lending rise-fall pattern (i.e. 1&A or 'pointed-hat' pattern in IPO terminology) was used widely and regularly in all affective states. However, it also appeared that many utterances were produced with other intonation patterns, and some intonation patterns seemed to be more characteristic for some emotions than for others. In particular, it was noticed that the pattern of a late rise and a very late fall was never used in final position in utterances expressing neutrality. Moreover, it is the second most often used intonation pattern in final position, just after the prominence-lending rise-fall pattern.

In the perception part of the study, a series of different intonation patterns was synthesised and transferred onto neutral utterances. The main objective of this experiment was to investigate whether specific intonation patterns contribute to the perception of particular emotions. It appeared that for all intonation patterns, the identification of boredom and neutrality was very high, and the identification of sadness and anger was low. The patterns 1A and 1EA showed the best results for the perception of anger. Although the identification was poor for sadness and mediocre for fear, the best results for these affective states were obtained using a final pattern of a late rise and a very late fall. The use of the pattern of a rise followed by a very late rise in the final position resulted in a strong perceptual bias towards indignation.

The choice of intonation patterns appears to be relevant for the perception of some affective states in speech. Some patterns appear to be more suited to convey certain emotions than others. Specific patterns introduce a perceptual bias towards

the perception of particular emotions, as in the case of indignation. The association between intonation patterns and some emotions, i.e. sadness and anger, is less successful, which suggests that other prosodic cues may be more important for their vocal expression.

The influence of emotions on intonation patterns is also addressed by Bänziger & Scherer (2005) on the corpus of acted affective speech with arousal control in the German language. They measured pitch points and pitch range, as suggested by Patterson & Ladd (1999): they represented an intonation contour in relation to linguistically meaningful intonation peaks and valleys of the utterance. Their results also showed comparatively little evidence for qualitatively different contour shapes for different emotions.

Stibbard (2000) analysed natural English emotional speech taken from television programmes using the ToBI labelling system. The analysis of total occurrences of annotated pitch accents in five emotions (anger, fear, disgust, happiness and sadness) show that H^* pitch accents (H^* stands for a high tone on accented syllable in ToBI annotation) predominate greatly across all the emotions. Regardless of emotion, $!H^*$, the downstepped equivalent of H^*, is the second commonest pitch accent type after H^*, and all other types occur much less frequently. Sadness has a larger number of L^* pitch accents (low tones) than the other emotions. The results for boundary tones show that $L - L\%$ boundary tones outnumber all other types put together across all emotions. Results for terminal tone contours, which consist of a full boundary tone plus the preceding pitch accent, show that the pattern $H^*L - L\%$ heavily dominates across all studied emotions. These results by Stibbard (2000) may be explained by the fact that this is a usual intonational pattern for statements.

Results presented by Stibbard for natural emotional speech in English correspond to those of Mozziconacci (1998), based on acted affective speech in Dutch. Mozziconacci used the IPO grammar of intonation for Dutch, and identified a prominence-lending rise-fall pattern as the most widespread intonation contour across all affective states. This contour may also be transcribed as $H^*L - L\%$ in the Tran-

scription of Dutch Intonation (ToDI), the ToBI-like system developed for Dutch (Gussenhoven 2005). Thus, the same intonation pattern appears as most frequently used in affective and neutral speech based on the two presented studies, independent of their intonation model and language, even if certain associations between some intonation patterns and affective states are found. Mozziconacci suggests that intonation contours may primarily serve linguistic functions, such as realisations of different sentence types: *'Intonation pattern represents a phonologically significant pattern of pitch movements distributed over an utterance, it is an abstract representation of concrete phonetic realisations'* (Mozziconacci 2002). Thus, the $H^*L - L\%$ pattern found in all affective states just serves to represent statements. In contrast, pitch range is claimed by Mozziconacci (2002) as one of the primary parameters covering phonetic variability and serving paralinguistic functions. Mozziconacci notes that the high phonetic variability, realised through the pitch range, does not change the phonological meaning of the intonation pattern. These claims were supported experimentally by Chen (2005). She found cross-linguistic variability in the association of particular usage of pitch range and affective states.

Pitch range

Laver (1994) distinguishes four types of pitch range: organic pitch range, paralinguistic pitch range, linguistic pitch range and pitch span. Organic range represents the maximum pitch range of which the speaker's voice is physically capable. Paralinguistic range is the adjustment of the pitch range within the organic range for paralinguistic purposes of signalling particular affective information. Cultures vary in the way that they utilise pitch settings as part of paralinguistic communication. Linguistic range is the pitch range within which the phonologically relevant pitch of the speaker's voice habitually varies in paralinguistically unmarked conversation. Languages differ to some degree in the placement of this linguistic range within the organic range. Finally, the fourth type of pitch range is defined by Laver (1994) as a phonological pitch span which is the local range realised by the speaker within the
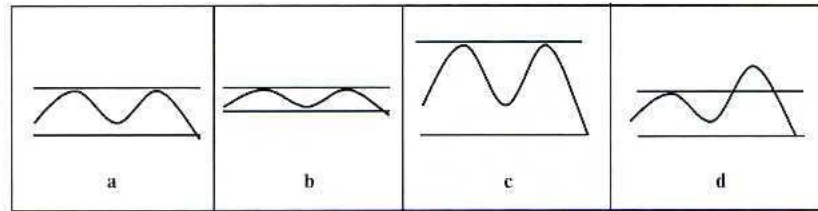
Figure 3.1: Table of pitch range variation, taken from Gussenhoven (2004). Variation in pitch range relative to panel a: higher register (panel b), greater span (panel c), and greater local-accent range.

whole or part of a particular utterance.

The pitch span notion is determined by two framing components (Laver 1994, Gussenhoven 2004, Ladd 1996). The first component, the baseline (also called bottomline, register or level) is the series of pitch values forming the floor of the current pitch span. Variation in the baseline amounts to the raising and lowering of the intonation pattern in the F_0 space. The second component, the topline, or plateau, is the series of pitch values forming the ceiling of the current pitch span. The interval between the baseline and the topline, the excursion size, defines the width of the pitch span. Speakers choose a pitch span specification for the entire utterance, but they may also independently differentiate local accentual contours by pitch span. See an example of pitch range variability in Figure 3.1. One of the characteristic pitch phenomena is the downward slope of both the baseline and the topline, with a progressive narrowing of the pitch span within the linguistic range, usually called declination. The width of the pitch span and the pitch value of the baseline can be manipulated by speakers for phonological and paralinguistic effects. Laver (1994) gives an example of phonological usage: in English, the start of a new topic is often signalled by a sudden widening of the pitch span and by a step-like raising of the baseline.

The majority of affective speech studies (Slaney & McRoberts 1998, Cohn & Katz 1998, Makarova & Petrushin 2003) use statistical approaches to pitch range estimation: mean F_0, standard deviation of F_0, the difference between maximum and minimum F_0. Even if these measures have been found to correlate with paralinguistic

information, they are only partially informative (Johnstone & Scherer 2000).

Patterson (2000) has applied pitch range measures, proposed by Ladd (1996), together with statistical measures in order to investigate which measures are the best in capturing pitch range conveying affective information. Patterson suggests that a linguistically motivated approach (acoustic measurement of linguistcally important pitch variation in the utterance, such as pitch peaks and valleys) gives better and more perceptually relevant results than a statistical approach (automatic extraction of pitch maximum and minimum values of the utterance). Sentence final low and post-accent valleys are estimated as competing linguistic measures for level, both of them were found to be successful in capturing speaker characteristics, though sentence final low is slightly more effective. The difference between non-initial accent peaks and post-accent valleys is shown as the best span measurement by Patterson (2000). See more details and examples in Chapter 4 'Methodology' and Appendix A.

Several studies of affective speech (Paeschke & Sendlmeier 2000, Bänziger & Scherer 2005) used pitch range measures developed and studied by Ladd (1996), Patterson & Ladd (1999), Patterson (2000). Results of Bänziger & Scherer (2005) show that pitch level, measured as suggested by Patterson & Ladd (1999), is mainly affected by emotional arousal. Low arousal emotions, i.e. sadness, calm joy and anxious fear, have a lower pitch level than high arousal emotions, i.e. despairing sadness, elated joy, panic fear and hot anger. The pitch level of cold anger (LA anger) is located between the pitch level of other low-aroused and of high- aroused expressions. The authors (Bänziger & Scherer 2005) have also observed a difference between the average location of the absolute maximum of F_0, measured as F_0 peak, for happy expressions (LA joy) and for elated expressions (HA joy). On average, F0 peak was measured at 46% of the total duration of the utterances for happy expressions, and at 72% of the utterances for elated expressions. The authors also state that the 'accentuation' of despairing sadness and elated happiness is, on average, more marked on the second part than on the first part of the utterances,

and the topline is consistently higher in those expressions until the final fall than in expressions of cold anger (LA anger) and panic fear (HA fear), which are also more 'accentuated' on the second part than on the first part of the utterances. In addition, the F_0 topline of those expressions is notably lowered before the final fall on the last syllable. The proper definition of this phenomenon, described by the authors as 'accentuation', is the local accent range (Gussenhoven 2004).

Paeschke & Sendlmeier (2000) analysed the corpus of German acted affective speech, only anger arousal is specified (high arousal). The intonation pattern representing sadness starts with the lowest F0 value and has the smallest differences between maxima and minima; it does not fall considerably at the end of the sentence like most other emotions. The baseline of neutral utterances is found around two semitones higher than for sadness, and their intonational pattern falls at their end nearly to the lowest border of the speaking range, it is significantly smaller than in the other emotions. Happiness and anger show nearly exactly the same results for all range measures, but their peak points lie significantly higher than for the other emotions. The final F_0 value of happiness and anger lies above the final pitch points of sad and neutral sentences and considerably lower than the final point in fearful utterances, which has the highest final F_0 value. The differences between maxima and minima of F_0 in fearful sentences are almost of the same size as in neutral sentences, but the baseline of fear is twice as high as in neutral sentences.

3.3 Biologically motivated theory of universal paralinguistic intonation meaning

Two biologically motivated theories have been developed to account for universality and language specificity in pitch variation (Ohala 1983, Gussenhoven 2004). Ohala (1983) identified the universal connection between F_0 and size of the organism, with smaller larynxes producing higher pitch than larger ones, and defined this relation as *Frequency Code*. Departing from Ohala's approach, Gussenhoven proposed an extended model of universal paralinguistic intonation meaning (Gussenhoven

2004). He identified two more biological codes associated with universal paralinguistic meanings in speech: the *Effort Code* and the *Production Code*. The Effort Code describes the association of effort variation with the excursion size of pitch movements, greater effort leading to wider excursions. The Production Code associates high pitch with beginnings and low pitch with endings of speech events which comes from the regulation of the breathing process. To summarise, biological codes are form-meaning relations which are based on the effects of physiological properties of the production process on the signal.

Gussenhoven also specifies that people do not need to have the actual physiological conditions required by the biological codes, they can communicate by means of these codes by creating the effects which are understood by others as associated with form-meaning relations. That means that the exploitation of the biological codes is a controlled use of pitch variation. Indeed, speakers cannot change the size of their larynx to manipulate pitch for the purposes of the Frequency Code, but they may use pitch height in peaks or peak delay, due to the mechanical connection between high peaks and late peaks. Gussenhoven (2004) notes that a higher pitch peak takes longer to reach than a lower one if the rate of change is the same, in other words, higher peaks tend to be later than lower peaks. Therefore, peak delay may be used as an enhancement of, or even as a substitute for, pitch raising. Another phonetic variable which may be used as a substitute of parameters used in the biological codes is pitch level, or pitch register: higher register may replace wider pitch span.

Gussenhoven (2004) states that the three biological codes explain the universality in the interpretation of pitch variation. A particular usage of the general form-meaning relation receives a number of specific interpretations: 'affective' signals attributes of the speaker, and 'informational' signals attributes of the message. The specific meaning derived from these codes may occasionally be in competition, for example, high pitch may signal opposite attributes depending on the code used for interpretation (higher pitch signals small size according to the Frequency Code and more effort according to the Effort Code). Moreover, similar meanings can be

achieved from different pitch characteristics, if these are interpreted in terms of different codes. Many languages adopt universal form-meaning relations derived from the biological codes as structural, i.e. phonological elements in their grammar. This phenomenon is defined by Gussenhoven (2004) as grammaticalisation of paralinguistic form-meaning relations. As these structural elements are apparently subject to the forces of language change, some languages possess form-meaning relations that go against the biological codes, for example, when a language possesses interrogative intonation contours ending in falling pitch.

Thus, the meanings of the biological codes, developed by Gussenhoven (2004), may be summarised as follows:

- The Frequency Code is widely used for the expression of affective meanings. The correlation between larynx size and vibration rates of the vocal cords is exploited for the expression of power relations: higher pitch may be used for creating the image of a 'small' person, lower pitch for a 'big' person. Low pitch is associated with masculinity, dominance/assertiveness, aggression, confidence, and protectiveness.

- The informational interpretation of the Effort Code is emphasis, based on the speaker's assumed intention to underscore the importance of the message. Affective interpretations are insistence, enthusiasm, and obligingness. The latter is due to the interpretation of effort as the speaker's desire to appear clear and unambiguous.

- The Production Code is used for the signalling of phrasing, it is due to the effect of energy dissipation in the course of the utterance. Its interpretations are informational only: high beginnings signal newness of topic, low beginnings the opposite; high endings signal continuation, low endings the opposite.

Gussenhoven underlines that the universal biological codes are not used in the same way in different languages, as grammars may favour or disfavour particular form-meaning relations. For example, cross-linguistic differences were found in the

degree to which F_0 peaks signal interrogativity (Hadding-Koch & Studdert-Kennedy 1964). Moreover, the exploitation of the biological codes is to some extent conventionalised within speech communities, for example, high pitch is used more extensively to signal feminine values in Japanese than in American English (Ohara 1992). It means that communities may choose a particular interpretation to a particular degree. In cases, when mutually incompatible interpretations are in competition, different communities may make different choices. For example, in the study of Chen et al. (2002, 1999) English listeners perceived higher registers as more friendly, but Dutch listeners did not; Dutch listeners perceived more emphasis as the register was higher, but English listeners less. Since friendliness and emphasis may be felt to be mutually exclusive meanings, the two speech communities have made different choices from universally available meanings: emphasis as expressed by pitch register used as a substitute for pitch span by the Dutch (Effort Code), and friendliness as expressed by high register by the English (Frequency Code).

Another competition concerns the availability of different strategies for the expression of emphasis. There are two ways in which the Effort Code can be recruited to express emphasis in F_0 peaks. One, as in Hamburg German (Peters 2002) as cited by Gussenhoven (2004), is enhancement of high pitch by late peaks. The other, as in Zagreb Croatian (Smiljanić & Hualde 2000) as cited by Gussenhoven (2004), is to locate the peak precisely inside the stressed syllable, rather than allowing it to drift rightwards. In other words, it takes more effort to have a peak inside its associated syllable, just as it takes more effort to have wide excursions. Thus, the implementation of the three biological codes shows language-related and culture-related differences.

3.4 Cross-linguistic studies of affective speech

Cross-linguistic research of vocal affective speech is primarily concerned with perception. Studies which have recently appeared in this field will be presented in this section. All of these studies used only adults as subjects for their experiments. While

some of them have only monolingual subjects, others also include second language speakers.

Scherer et al. (2001) reports the results from a study conducted in nine countries in Europe, the United States and Asia on vocal emotion portrayals of anger, sadness, fear, joy, and neutral voice as produced by professional German actors in language-free utterances. Data show an overall recognition accuracy of 66% across all emotions and countries. Although accuracy was substantially better than chance, there were sizable differences ranging from 74% in Germany to 52% in Indonesia. However, patterns of confusion were very similar across all countries. Anger was the best recognised in this study, followed by fear and sadness. Joy, which is one of the best recognized in studies of facial expression, reaches the lowest accuracy percentage in the present study. The authors suggest that one of the reasons may be that joy is strongly marked by smiling and this helps to give good results in facial recognition of emotions. No comparably iconic cue may exist for the voice. The most attention-getting vocal cues, such as loudness and fundamental frequency, may be quite similar in intense joy, fear, and anger (due to high arousal), possibly explaining the patterns of confusion. The authors suggest the existence of similar inference rules from vocal expression across cultures. Generally, accuracy decreased with increasing language dissimilarity from German in spite of the use of language-free speech samples. It is concluded that culture- and language-specific paralinguistic patterns may influence the decoding process.

Chen (2005) studied experimentally the relationships between Gussenhoven's Biological Codes and paralinguistic intonational meanings. On the one hand, her research supports Gussenhoven's claims on the universality of intonational meaning. On the other hand, results bring to light significant differences between speech communities in interpreting pitch variation. Language-specificity appears to occur both at the level of the association and at the level of the strength of the association. The association description specifies if a given parameter is perceived as carrying a meaning predicted by the biological codes. The strength of association defines how

much variation in meaning a given interval of pitch values is perceived to signal. At the level of association, two speech communities may occasionally interpret the relation between pitch variation and a given meaning in opposite ways; presumably only one of the two ways accords with the biological codes. For example, some languages possess interrogative intonation contours ending in falling pitch (Arvanti et al. 2006) which goes against the Frequency Code. Furthermore, one speech community may associate pitch variation with a certain meaning where the other does not.

Chen (2005) conducted several experiments to study the perception of paralinguistic intonational meanings. British English and Dutch subjects listened to stimuli in their native language and judged each stimulus on four semantic scales stemming from the four meanings: 'not self-confident' *versus* 'self-confident', 'not friendly' *versus* 'friendly' (derived from the Frequency Code - Experiment 1); 'not surprised' *versus* 'surprised', and 'not emphatic' *versus* 'emphatic' (derived from the Effort Code - Experiment 2). The stimuli were synthesised to have a range of differences in pitch contour, pitch register (or level) and pitch span in Experiment 1, and of pitch register, peak height, peak alignment and end pitch in Experiment 2. In the perception of 'confident' and 'surprised' as signalled by pitch register and 'emphatic' and 'surprised' as signalled by peak height, Dutch listeners perceived a significantly larger meaning difference for the given interval of pitch values (Type-1 difference). This Type-1 difference can be explained by the difference in standard pitch range: British English has a wide pitch range and Dutch has a narrow range. Thus, a given pitch interval will signal a larger meaning difference in Dutch than in English, even if British English and Dutch speakers express the same range of semantic meaning.

More of Chen's results show that, in the perception of 'emphatic' as signalled by pitch register, Dutch listeners associated a higher pitch register with a higher degree of emphasis, but British English listeners associated a higher register with a lower degree of emphasis (Type-2 difference). This difference comes from the fact that languages may make different choices whenever conflicting meanings are derived from the biological codes. Such conflicts are common, as the three biological

codes may use the same phonetic parameter, i.e. F_0. In particular, pitch register appears to signal friendliness in British English on the basis of the Frequency Code, but emphasis in Dutch on the basis of the Effort Code. Raised pitch register was perceived with a lower degree of emphasis and a higher degree of friendliness by British English listeners. Dutch listeners showed a reversed association. The association of pitch register with emphasis in Dutch makes the use of pitch register for the expression of friendliness less effective.

In the perception of 'surprised' and 'emphatic' as signalled by end pitch and 'surprised' as signalled by peak alignment, British English listeners associated variations in peak alignment and end pitch with different degrees of surprise and emphasis, whereas Dutch listeners perceived little difference of meaning (Type-3 difference). Later aligned peaks signalled higher level of 'surprised' only for British English listeners. Chen (2005) attributes this to the fact that variations in tonal alignments are used to signal 'routineness' in British English (i.e. the later the alignment, the less the routineness) (Gussenhoven 1984) but probably not in Dutch. In the perception of 'surprised' and 'emphatic' as signalled by end pitch, British English listeners appeared to have strong associations between L% (low tone) and 'emphasis', and between H% (high tone) and 'surprise', while Dutch listeners did not show any significant associations. Chen suggests that speech communities may attribute more or less importance to pitch variation at the phrase end.

Yanushevskaya et al. (2006) reports results from her cross-linguistic investigation on how voice quality (VQ) and fundamental frequency (F_0) combine in the signalling of affect for Japanese speakers. Three types of synthesised stimuli were used: (1) 'VQ only', involving variations in voice quality and a neutral F_0; (2) 'F_0 only', with different F_0 contours and modal voice; and (3) combined 'VQ + F_0' stimuli, where combinations of (1) and (2) were employed. Overall, stimuli involving voice quality variation (1 and 3) proved to be most consistently associated with affect. In series (2) only stimuli with very high F_0 yielded high affective ratings: indignation and fear. Some striking differences emerge in the ratings obtained for Japanese subjects

compared to those obtained for speakers of Irish English. First of all, there is relatively greater importance of the F_0 contour for Japanese subjects, at least when it involves very high levels. Secondly, Japanese speakers did not identify three affective states, i.e. 'formal', 'stressed' and 'relaxed', in the proposed synthesised stimuli, while Irish English listeners associated these affective states especially strongly with specific voice qualities. Thirdly, there are cases of different attribution of affect to the same stimulus by the two language groups: i.e. the combined stimulus 'tense + F_0 of indignation', which appears to be readily associated with 'indignation' for the Irish English subjects, is associated with 'intimacy' by the Japanese.

de Mareuil et al. (2002) report results on the generation of emotions in three languages; English, French and Spanish. Recordings of three male actors (one speaker per language, 25 utterances per speaker) were analysed, and F_0 curves were used for resynthesis. Though the authors speak about different cross-language strategies exhibited by their actors, it seems premature to make such generalisations on the basis of just one speaker per language, as one speaker per language does not allow distinction between individual and cross-linguistic variability. Speakers showed most variability in the expression of fear - the English actor simulated fear through whispering, while the French and Spanish actors shouted and cried respectively.

Braun & Katerbow (2005) conducted a cross-linguistic study on samples of dubbed affective speech. The American-English original of a popular TV series (*Ally McBeal*), as well as its German and Japanese dubbed versions, were used for selection of affective utterances (anger, joy, fear, sadness, neutral) in the productions of the main male and female characters. Forty-five utterances per language (about 5 utterances per emotion of one male and one female speaker) were selected for acoustic analysis and perception tests. Speakers of the same sex were compared across languages. Statistical measures of F_0 were reported: average F_0, F_0 standard deviation, and F_0 range. It is difficult to interpret the reported production results, as no speaker normalisation was performed on the data. Moreover, the number of speakers per language does not allow the abstraction of individual differences,

and analysis of language-specific characteristics. Perception test results show that German and American listeners could not identify joy in the Japanese male productions significantly, while anger and neutral were successfully identified. In contrast, Japanese female productions of joy were identified by German and American listeners even more successfully than by Japanese listeners. In general, German and American listeners had more similar associations in perception tests than either had with Japanese listeners, and the authors suggest that more differences in perception may exist between speakers of languages that are less related linguistically and culturally than between the more closely related ones.

Cross-linguistic research of vocal affective speech is primarily concerned with perception. Studies that have recently appeared in this field were presented in this section. All of these studies used only adults as subjects for their experiments. While some of them studied only monolinguals, others also include second language speakers. To the best of our knowledge, there are no cross-linguistic studies of affective speech using children. There appears to be a general consensus that both biological and cultural factors contribute to emotion expression and perception.

3.5 Conclusion

Cross-linguistic studies that have recently appeared in the field of affective speech research predominantly concentrate on perception. To summarise, they suggest that vocal expressions of emotions are based both on universal mechanisms and on social/cultural conventions. In the context of second language acquisition, perception studies show that both first language transfer and activation of universal paralinguistic intonational meanings are present in the interpretation of pitch variation in second language. A number of available means to express vocal emotions has been identified: pitch variation, voice quality, speech rate, intensity, and others. Perception studies show that their paralinguistic usage, level of importance and meaning vary in different languages. Different methods of analysis were developed and tested for these acoustic parameters, linguistic methods of speech analysis were identified

as giving better and more perceptually relevant results. Cross-linguistic production studies of affective speech are comparatively rare. Moreover, they present a number of important limitations. First, they use a very small number of speakers per language, usually between one and three speakers. This means that they disregard individual variability, which is characteristic of affective speech. Second, they mostly use statistical methods of speech analysis, which are easier to apply, but less effective than linguistic methods. In spite of these limitations, these studies try to generalise and speak about language differences in affective expressions. Indeed, the main difficulty in performing larger scale production studies is time-consuming acoustic analysis. Nevertheless, it is evident that to be able to make valid generalisations requires a higher number of speakers per language than one or two. The increase of the number of speakers in the production studies presents a challenge, which when overcome will bring better understanding of cross-language and within-language variability in vocal affective expressions.

Chapter 4

Bilingual language acquisition

4.1 Introduction

Concepts of bilingual language acquisition, particularly those related to phonetic learning and affective expression, are introduced and discussed in this chapter. Language acquisition research was historically concerned with monolingual children. Only quite recently, from the late 1980s, has research interest in bilingualism increased, partially due to the recognition of bilingualism as a normal and widespread way of language acquisition. To the best of our knowledge, there exist no studies of simultaneous bilingual paralinguistic phonetic acquisition, thus only research in monolingual and second language acquisition is reported.

4.2 Definition of bilingualism

The definition of bilingualism by Blanc (2000), adopted in the present study, is based on three dimensions; linguistic, communicative and cognitive. The linguistic dimension covers the important concept of *language competence*. Language competence comprises knowledge of multiple language components; lexicon, semantics, syntax, morphology, phonology and pragmatics. A distinction is made between the balanced bilingual, who has equivalent competence in both languages, and the dominant bilingual, for whom competence in one of the languages is superior to competence in the other. Note that dominance/balance is not equally distributed for all domains and functions of language; each individual has their own dominance configuration. Grosjean (1998) summarises two contrasting views on bilingual competence existing in research and education. According to the first approach, the bilingual is considered as 'two monolinguals in one person'. It assumes that bilinguals should be as proficient as monolinguals in their two languages. This 'monolingual approach' sees

the monolingual as the norm, and the bilingual as an exception. The second view argues that the bilingual is not the sum of two complete or incomplete monolinguals, but that bilinguals form a group of speakers with a unique linguistic profile. It also suggests that bilinguals should be measured and compared by reference primarily to other bilinguals, and only secondarily to monolinguals.

The other factors, proposed by Blanc (2000) for the definition of bilingualism, are sociocultural environment, age and context of acquisition. *The age and context of acquisition* play an important part in respect of various aspects of the bilingual's development, such as linguistic, neuropsychological, cognitive and sociocultural. There are two main types of bilingualism; simultaneous early or infant bilingualism, when the child develops two mother languages from the onset of language (the area under investigation in the present study), and consecutive bilingualism, when a person acquires a second language after the basic linguistic acquisition of the mother language has been achieved. The acquisitional context may vary, from family to institutional environment. *The sociocultural environment* describes the status of the two languages in the community. If the two languages are both socially valued in the child's environment, the benefit from the bilingual experience will be very important for the child and the stimulating environment will lead to greater cognitive flexibility. This type is called additive bilingualism. Subtractive bilingualism describes the opposite situation, when one of the child's languages is devalued, with the consequence that the child's cognitive development may be delayed.

Bilinguals communicate in two languages which also reflect two different cultures. The principal question which occupies many scientists in bilingual research is about the organisation of two languages in one mind: how two languages (and two cultures) are represented in the speaker's mind and how they are used by the same speaker. Do bilinguals have one or two linguistic systems, one or two affective and cultural representations, one or two identities? This dilemma of how one person manages to live with two languages and two cultures attracts much attention.

In the context of bilingual speech acquisition, one question is whether speech ac-

quisition of a simultaneous bilingual child (learning two languages from birth) starts from one linguistic system that eventually separates into two (the 'unitary-system hypothesis' (Redlinger & Park 1980, Volterra & Taeschner 1978, Vihman 1985)), or if children separate two linguistic systems from the beginning (the 'dual hypothesis' (Genesee 1989, Nicoladis & Genesee 1996)). Genesee (2003) argues that while bilingual children differentiate their developing languages and use them appropriately with different interlocutors, it often happens that one language becomes dominant over the other. Thus, bilingual children are expected to use their languages in proportion to their relative proficiency in each. The existing research (Peperkamp et al. 1999, Genesee 2003) suggests that out of two important contributors to the language dominance - the language of the primary caretaker and the language of the community - the latter is relatively more influential in bilingual language development.

In the context of speech production modelling, researchers design bilingual production models as composed either of one extended system or of two language specific subsystems (de Bot 1992, Cook 2002, Tomioka 2002). In the context of cultural psychology, scientists suggest two models of affective representation and expression for bilinguals; a cultural alternation model of bilingualism where bilinguals function in an alternating fashion between different cultural frames of mind, and a cultural fusion model where bilinguals adopt a fusion of the two cultures to form an integrated culture that frames affective behaviour of the bilingual person (Chang et al. 2002). In the psycholinguistic context, studies of emotion in autobiographical memory, linguistic repertoires, and conceptual representations of bilinguals investigate the complex question of bilingual identities (Pavlenko 2006). Do bilinguals have a single or a double identity, specific to their two languages and cultures?

One of the research questions of the present dissertation is to investigate how simultaneous bilingual children produce affective speech in their two languages. This study appears related to the bilingual research working on the question of whether bilinguals have one extended language system or two language-specific subsystems. The review of models of linguistic speech production will be covered in the following

section, these models will be tested for their ability to cover affective speech productions of bilingual children and will be discussed in the Conclusion chapter. The study of how bilingual children produce affective speech in their two languages may help to achieve better understanding of bilingual language organisation in general.

4.3   Models of bilingual speech production

Most speech production theories are exclusively based on monolingual data (Dell et al. 1997, Ferrand 2001, Caramazza & Miozzo 1997, Levelt 1989). Nevertheless, there have been some attempts to develop models of multilingual production. Some theories of multilingual production are limited only to specific areas, like the lexicon (Kroll & de Groot 1997, Green 1998). de Bot (1992) redesigned the monolingual model proposed by Levelt (1989) and described all the modules involved in language production: conceptualizer, lexicon, formulator, monitor system and articulator. Thus, this adaptation by de Bot presents a particular interest. De Bot postulates that the conceptualizer, responsible for the choice of language and for the formulation of the intended message, is common for both languages. Then, de Bot describes the lexicon and the formulator, responsible for lexical, morphosyntactic and phonological information, as having language specific subsystems, and finally one extended system for the articulator, responsible for the phonetic output of the message. He argues that bilinguals possess only one articulator, without a systematic division for the two languages.

Wilson (2006) studied the articulatory setting of monolingual and bilingual speakers of Canadian English and Québécois French through the window of interspeech posture (ISP) of the articulator (the position of the articulators when they are motionless during inter-utterance pauses) by ultrasound imaging. Significant differences were found both across monolingual groups and within individual bilingual speakers. The components of ISP that differ across these languages between monolingual groups are upper and lower lip protrusion, tongue tip height, and the degree to which the corners of the mouth are drawn towards the midsagittal plane

from a maximally-spread position. In Canadian English, the upper and lower lips are significantly more protruded, the tongue tip is higher, and the corners of the mouth are drawn farther toward the midsagittal plane. Amongst individual bilingual speakers who are perceived to be native speakers of both Canadian English and Québécois French, all speakers show the same upper and lower lip protrusion differences (i.e. English more protruded than French) as the monolingual groups, and half of the speakers show the same tongue tip differences (i.e. English higher than French) as the monolingual groups. These are the only relevant cross-linguistic differences between ISPs for bilinguals who are perceived as native in both languages. The other important finding reported by Wilson (2006) is that bilinguals who are perceived as native speakers of the two languages do not have a unique inter-speech posture for bilingual speech mode (i.e. when the bilingual is ready to speak in either language). The bilingual mode was stimulated by presenting French and English utterances for reading in a mixed order. The ISP for each of these speakers in bilingual speech mode is equivalent to the monolingual mode ISP of that speaker's dominantly-used language. Thus, in summary, this research shows that ISP (and hence articulatory setting) is language specific between monolingual subjects and within bilingual subjects perceived as native in both languages.

The findings reported by Wilson (2006) support those researchers (Cook 2002, Tomioka 2002) who criticise the application of the subsystem division only to the lexicon and the formulator in the de Bot model (de Bot 1992). Apparently, bilinguals perceived as native speakers in both languages have specific articulatory settings, and thus they have separate articulators for their two languages. Nevertheless, not all bilingual speakers are perceived as native speakers in the two languages, and consequently not all bilinguals separate their articulatory settings.

The speech learning model (SLM) (Flege 1995, Piske et al. 2002, Aoyama et al. 2004) reconstructs the acquisition of a new phonetic system by second language (L2) learners. The SLM suggests that when L2 learners encounter an L2 sound which does not exist in the phonological inventory of their first language (L1), at

the beginning L2 learners substitute this L2 sound by the nearest L1 sound. With more exposure to L2, learners gradually establish a new phonological category for this sound. The SLM posits that L2 learners develop two phonetic subsystems, specific to their L1 and L2, but these subsystems exist in a 'common phonological space' and have influence on each other. According to the SLM, L1 and L2 speech sounds interact through two distinct mechanisms; category assimilation and category dissimilation. The first mechanism, category assimilation, is thought to operate when a new category is not established for an L2 speech sound despite audible differences between it and the closest L1 speech sound. The SLM predicts that in such cases, a 'merged' category will develop over time that subsumes the phonetic properties of the perceptually linked L1 and L2 sounds. The second mechanism, category dissimilation, explains a different way of interaction between L1 and L2 phonetic subsystems. This mechanism is thought to operate when a new category has been established for an L2 speech sound. Category dissimilation causes a newly established L2 category and the nearest L1 speech category to shift away from one another in phonetic space. The SLM argues that this mechanism develops out of the necessity to maintain phonetic contrast between all of the elements in the combined phonetic space of the two languages.

The study of Sundara & Polka (2007) investigated the discrimination of English and French coronal stops by simultaneous bilingual, monolingual and advanced early L2 learners of English and French. This study tested the SLM and provided evidence for differential effects of simultaneous and sequential acquisition of two languages on the organisation of phonological categories. Interactions between the two languages were observed in L2 learners as well as simultaneous bilinguals, but the nature of this interaction was different in the two groups. Specifically, in L2 learners, interactions resulted in assimilated or merged categories, whereas in simultaneous bilinguals, interactions resulted in dissimilated or well-separated categories. Apparently, simultaneous exposure to two languages gives an advantage in the acquisition of some phonetic categories over L2 learners. Moreover, the mechanism of dissimilation to

maintain phonetic contrasts between all categories results in differences between learners exposed to dual languages and a single language, as simultaneous bilinguals separated the studied phonetic categories more than monolinguals.

Cook (2003) proposed a theory of the integrated continuum. It describes a model of the linguistic organisation in the bilingual mind which sees the language system of the multilingual speaker as a whole and covers all possible relations between different components of the two languages in one's mind, from the complete separation to the total integration. The continuum does not necessarily imply the direction of the movement (the organisation may be stable) or apply to the whole language system (some components may be integrated, others separated); nor does it necessarily affect all individuals in the same way. The point on the continuum may also vary according to the individual's perception of language mode, level of tiredness and other personal factors. The continuum may also be related to different stages of language development. This model was designed to cover all possible linguistic organisations in multilingual speakers. Thus, the model of integrated continuum gives much flexibility in describing particular research cases.

4.4 Bilingual acquisition of phonetic abilities

Several investigations (Bosch & Sebastián-Gallés 2003*b,a*, Burnham 2003, Sundara et al. 2006, Khattab 2002, Gordeeva 2006, Whitworth 2003, Kehoe 2002) have recently appeared on simultaneous bilingual acquisition, examining various perception and production phonetic abilities of children and adults. These studies show that it is possible to find both phonetic separation and interference between the two languages of bilinguals. Interference can be defined as 'those instances of deviation from the norms of either language which occur in the speech of bilinguals as a result of their familiarity with more than one language' (Weinreich 1953). Some studies also suggest that simultaneous exposure to two languages has an impact on child phonetic development - some phonetic abilities are acquired later by bilingual than by monolingual children. Furthermore, it appears that for some phonetic aspects,

not all simultaneous bilinguals attain monolingual phonetic proficiency in their two languages even in adulthood.

In the context of perception phonetic development, studies in (Bosch & Sebastián-Gallés 2003*b,a*, Burnham 2003, Peperkamp et al. 1999, Sundara et al. 2006) investigated phonetic contrasts at the segmental level and showed that dual language exposure may change the rate at which children's phonetic discrimination abilities become language specific. Simultaneous bilingual children show language-specific discrimination abilities for contrasts that occur only in one of the two languages at a later age than monolingual children. For example, Sundara et al. (2006) examined the discrimination of the voiced stop-voiced fricative contrast /d-ð/ by 4-year-olds learning either English or French, or both from birth, and by simultaneous bilingual adults. Their findings show that the ability to discriminate the native English contrast improved with age for bilingual and monolingual English children, but not for monolingual French children and adults. Furthermore, although simultaneous bilingual and monolingual English adults were comparable, children exposed to both English and French were poorer at discriminating this contrast when compared to monolingual English-learning 4-year-olds. The authors argue that even if language experience facilitates perception of the English /d-ð/ contrast, this facilitation occurs later in development when English and French are acquired simultaneously. The study of Peperkamp et al. (1999) examined phonetic discrimination of word stress by adult French-Spanish bilinguals who learnt their two languages from birth. In contrast to Spanish, French has stress on the last syllable. The perception test of the stress location with monolingual adults shows that French subjects are 'deaf' to stress contrasts. The replication of the test with Spanish-French simultaneous bilingual adults shows that the majority of these bilinguals show the ability to discriminate stress location like Spanish monolinguals. Nevertheless, some bilingual subjects exhibit certain 'stress deafness' such that their performance corresponds to the best results in French monolingual group.

In the context of production phonetic development (Khattab 2002, Gordeeva

2006, Whitworth 2003, Kehoe 2002), research supports the main findings of perception studies that simultaneous bilingual children produce some language-specific phonetic contrasts later than monolingual children, and not all bilinguals finally acquire monolingual phonetic proficiency in their two languages. Kehoe (2002) studied the acquisition of vowel systems by German-Spanish simultaneous bilingual children, aged 1-3 years old. Results show that while German monolingual children produce significant differences in short and long vowels at the age of 2;6[1], bilingual children do not significantly differentiate vowel length in their productions at this age. Gordeeva (2006) performed a longitudinal production study of two Scottish English-Russian simultaneous bilinguals. She examined the production of prominent syllable nuclear vowels in the two languages of bilingual children through the measurement of vowel quality, vowel duration and vocal effort. The subjects produced variable degrees of language differentiation and interaction depending on their age, language exposure, cross-linguistic structure and the studied acoustic parameter. The author reports that the two bilingual children showed a slower rate of Scottish vowel length acquisition than Scottish monolingual children at the beginning of observation (3;8 for child 1 and 3;4 for child 2), but by the end of observation (4;5), child 1 had apparently acquired and child 2 had significantly improved contrastive realisations of vowel length. Whitworth (2003) studied rhythm acquisition by six German-English bilingual children in comparison to their parents realisations, the ages of the children were 5;0, 7;6, 6;2, 8;10, 10;10, 13;2. Only the oldest child, aged 13;2, showed language specific rhythm realisation in his two languages. His younger brother, aged 10;10, showed the tendency to separate his two languages in rhythm parameters, but this differentiation was still not statistically significant.

Simultaneous exposure to two languages has an impact on child phonetic development - some phonetic abilities are acquired later by bilingual than by monolingual children. Paradis & Genesee (1996) argue that the rate differences of monolingual and bilingual development gives evidence for an interaction between the two lan-

---

[1]For ease of notation, an age of 2 years and 6 months is denoted as 2;6.

guages of bilinguals, but the interaction does not contradict the presence of differentiation between the two language. The nature of interaction between the two languages is influenced by the amount of exposure to each of the languages. Gordeeva (2006) reported that the amount of language exposure is reflected in the different extent of language differentiation. For a balanced bilingual child, the language interaction effects were bi-directional within the same sound structure variable. For a dominant bilingual child, the interaction effects were unidirectional, from the dominant language to the weaker one. In general, language interaction can be considered as a characteristic feature of bilingual phonetic acquisition, even if its presence is not necessarily obligatory. The slower rate of acquisition of some phonetic abilities by bilingual children must not be regarded as a delay in acquisition, as monolingual standards are not directly applicable to bilingual acquisition, which has its particular characteristics.

4.5 Monolingual and second language studies of affective phonetic acquisition

Research into affective phonetic acquisition has started very recently, and, to the best of our knowledge, it has not approached simultaneous bilinguals yet. This section presents existing research in monolingual and second language affective acquisition.

Chen (2005) studied the perception of paralinguistic intonational meaning in the listeners' second language. Advanced Dutch learners of British English (L2 English listeners) judged the English stimuli on the scales 'not surprised' *versus* 'surprised', and 'not emphatic' *versus* 'emphatic'. The results show that L2 English listeners resemble monolingual Dutch listeners (L1 Dutch listeners). In this study, stimuli were synthesised with different peak heights. L2 English listeners perceived a significantly larger meaning difference for the given interval of peak heights in the perception of both meanings than monolingual English listeners (Type-1 difference). In the perception of 'emphatic' as signalled by pitch register, L2 English listeners associated a higher pitch register with a higher degree of emphasis, whereas L1 English listeners

associated a higher register with a lower degree of emphasis (Type-2 difference). In the perception of 'surprised' as signalled by peak alignment, and of 'emphatic' as signalled by end pitch, L2 English listeners again follow L1 Dutch listeners (Type-3 difference). These findings provide evidence that first language (L1) transfer [1] plays an important role in the perception of second language (L2) paralinguistic intonational meaning. At the same time, L2 English listeners showed some knowledge of the difference between standard pitch ranges in British English and Dutch, and tried to adjust their perception of intonational meanings. In the perception of 'emphatic' as signalled by peak height and pitch register, and 'surprised' by peak height, the meaning difference perceived by L2 English listeners was somewhere between that perceived by L1 English listeners and that perceived by L1 Dutch listeners.

In the next experiment of Chen (2005), the hypothesis of L1 transfer was further tested in the perception of non-advanced British English learners of Dutch (L2 Dutch listeners), who judged the Dutch stimuli again on the scales 'not surprised' *versus* 'surprised', and 'not emphatic' *versus* 'emphatic'. The hypothesis of L1 transfer was again supported. L2 Dutch listeners distinguished a significantly smaller meaning difference for the given interval of peak heights in the perception of 'surprise' and 'emphatic', and for the given interval of pitch registers in perception of 'surprise' (Type-1 difference) they performed as L1 British English speakers. Nevertheless, in the perception of 'emphatic' and 'surprised' as signalled by end pitch, and 'surprised' as signalled by peak alignment, L2 Dutch listeners showed wide variability. They did not perform as L1 Dutch listeners, and they showed some deviations from L1 British English listeners. For example, in the perception of 'surprised', an increase in peak height and pitch register did not necessarily lead to an increase in the perceived degree of surprise in L2 Dutch listeners' ratings.

Chen (2005) suggests that L2 non-advanced Dutch listeners start gaining knowledge on the differences in the communicative usages of pitch variations between

---

[1]In second language acquisition research, there is a term closely related to interference - language transfer, which is applied to the influence resulting from similarities and differences between the target language and any other language that has been previously acquired.

British English and Dutch. As this knowledge is still incomplete, their adjustment led to 'fuzzy' associations of pitch variations with meaning differences. By contrast, L2 advanced English listeners showed better understanding of the difference in standard pitch range between the two languages and adjusted their perception more. In the perception of 'emphatic' as signalled by pitch register, L2 Dutch listeners performed like L1 Dutch listeners - they associated an increase in pitch register with an increase in the perceived emphasis. This successful performance of non-advanced Dutch speakers is explained by the hypothesis that L2 Dutch listeners used the universal form-meaning relation between pitch register and emphasis embedded in the Effort Code (defined and discussed in Chapter 3). However, the reason why the Effort Code was not activated in the perception of 'surprised' as signalled by peak height and pitch register, is not known. L2 learners may apparently use two strategies in interpreting pitch variation in their second language - the transfer from the first language and the activation of the intonational universals, as embodied in the biological codes. Moreover, the level of competence has an important impact - advanced L2 learners understand better differences between their first and second language and adjust their interpretation of L2 pitch variation more successfully than beginner and intermediate learners.

Mejvaldova (2001) also studied intended emotions and attitudes cross-linguistically in French and Czech in the context of second language learning. Her results show that even when second language learners understand the language used, there are some attitudes (i.e. boredom, admiration) which are misinterpreted in identification tests. For example, prosodic expression of admiration utterances in Czech appears to be similar to the expression of boredom in French. This proximity is symmetric - admiration utterances in French pronounced by Czech speakers and correctly identified by Czech perceivers, were identified as boredom by French listeners. Boredom utterances in Czech pronounced by French speakers and correctly identified by French perceivers, were identified as admiration by Czech listeners. Shochi et al. (2006) studied the perception of Japanese attitudes by three groups of French listen-

ers (naïve, beginner and intermediate in Japanese). The results show that French listeners who are naïve in Japanese can well recognise admiration, authority and irritation. At the same time, French listeners are unable to discriminate Japanese questions and declaratives until they have reached the intermediate level in Japanese. The extreme Japanese politeness is interpreted as impoliteness by French listeners, even at Japanese intermediate level.

In the study of Dromey et al. (2005), the recognition of affective prosody by speakers of English as a first and foreign language was compared. Adults who were fluent in English, and who grew up speaking English or one of 21 other languages listened to words spoken with angry or neutral intonation by an actor, an English mother tongue monoglot. The accuracy with which the listeners identified the intended emotion was measured and compared between three groups: English mother tongue (EMT) monoglots, English mother tongue polyglots, other mother tongue polyglots. In the definition of the authors, polyglots are considered as speakers, acquired their second language after the first language. The results show that English mother tongue polyglots scored higher than other mother tongue listeners, whereas EMT monoglots did not. Women were significantly more accurate than men across the three listener groups. There was a modest inverse correlation between accuracy and age. The authors hypothesise that the learning of a second language may have helped the EMT polyglots to develop additional perceptual skills in decoding speech emotion in their native language.

Pavlenko (2002) investigated discursive constructions of emotions in the speech of adults who were late Russian-English bilingual. She suggests that in cases where two speech communities differ in the conceptualisation of emotions, the process of second language socialisation may result in the conceptual restructuring of emotion categories of adult language learners, as evident in instances of second language influence on first language performance. She states that in the process of second language socialisation some adults may transform their verbal repertoires and conceptualisations of emotions, or at least internalise new emotion concepts and scripts.

It is possible that in the process of second language socialisation, not only internalisation of new concepts takes place, but also the process of attrition of concepts and scripts that would be marked and inappropriate in the new interpretive community. Her analysis of the corpus shows at first sight that the Russian-English bilinguals in the study use their lexical resources appropriately in each language, exhibiting an adjectival pattern in English and a verbal pattern in Russian. However, a closer look at the emotion scripts they draw on suggests that some of them have internalised and actively used American emotional concepts, and that rather than adhering to emotion concepts and scripts salient in monolingual Russians' narratives, they favour concepts and scripts shared by their two speech communities. The analysis of the bilinguals' narratives told in Russian demonstrates that some late Russian-English bilinguals attempt to frame emotions as states through uniquely English means, thus producing instances of L2-influenced semantic and morpho-syntactic transfer. Pavlenko suggests that this could result from the fact that the subjects in the study were late bilinguals, who learned their second language in adulthood, and made a transition as refugees and immigrants to a more powerful and prestigious linguistic and cultural community. The results of the study also suggest that these individuals may be at different stages with regard to their discursive assimilation. Some may have already approximated the emotion discourses of their new community, while others, as seen in the instances of L1 transfer, may either be in the process of doing so or may be adhering to Russian discourse patterns.

Koven (2006) reported his investigation on how one French-Portuguese bilingual female displays affect differently in her two languages. The bilingual comes from a Portuguese family who immigrated to France. The author argues that it is not structural differences between French and Portuguese that account for the subject's French *versus* Portuguese affective displays, so much as the range and kinds of discursively enactable affective performances and personas to which she has access in each language. Specifically, the author focused on the subject's affective displays in narratives of personal experience by comparing the different intensities and locally

recognised affective stances in stories she tells in her two languages, as well as what other bilinguals infer about her affect from those tellings. The results show that the subject speaks as an interlocutor and more as a neutral narrator in Portuguese than in French. Similarly, she performs the voices of quoted characters in more extreme, marked styles in French than in Portuguese. Both the subject and listeners perceive her as enacting a different kind of affect, from a different kind of social actor. She is perceived in French not just as angrier for a current moment, but as an angrier person. On the other hand, in Portuguese she comes across as someone who restrains herself, and is thus a more 'calm, reserved' person.

Cross-linguistic studies of affective speech production and perception by adult monolingual and bilingual speakers show universal similarities and cross-linguistic differences (Mejvaldova 2001, Chen 2005, Yanushevskaya et al. 2006). There are studies looking at monolingual children's affective speech production and perception (Rothman & Nowicki 2004, der Meulen et al. 1997, Clément 1999), but to the best of our knowledge, child affective speech has not been approached yet from the cross-linguistic angle or in the bilingual acquisitional context. Thus, studies of affective speech in the perception and production by monolingual children are presented.

Rothman & Nowicki (2004) present a test of the ability to decode emotion of varying intensity in the voices of children, the Diagnostic Analysis of Nonverbal Accuracy Child Paralanguage (DANVA2-CP). The test uses the voices of a 10 year-old European American boy and girl saying the same utterance with happiness, sadness, anger and fear at different levels of intensity. A review of various studies that have applied the test shows that errors in decoding emotion in speech decrease rapidly with age through childhood and adolescence, then more slowly into adult-hood. The results from studies quoted by Rothman & Nowicki (2004), including nearly 2000 participants aged between 6 and 12 provide substantial support for an association between accuracy in identifying emotion in child voices as measured by the DANVA2-CP and social competence assessed in a number of ways. Difficulties in reading negative emotions appears to be associated with lack of social compe-

tence. Moreover, different patterns of negative emotion errors may be associated with different types of psychological problems and disorders.

Gérard & Clément (1998) compared the performance of French children (5, 7 and 9 years old) and adults in production and perception tasks of prosodic representations. They investigated how subjects control two linguistic prosodic representations (question and declarative) and two affective prosodic representations (happiness and sadness). During speech production, control of the fundamental frequency of the voice emerged earlier in childhood (5 year-olds), than control of rhythmic parameters (7 year-olds). Children produced prosodic contrasts to differentiate sadness and happiness at the age of 5, even though happiness is realised in a less contrastive style by children than it is by adults. The second experiment examined subjects' identification of the same prosodic representations as above, plus irony, all produced by an expert speaker, using a word-by-word gating procedure. According to this procedure, utterances are cut into several samples, starting from one first word, and then each time adding a new word to it. The gating procedure serves to identify the point in the utterance when a particular meaning can be identified. During speech perception, children were able to identify different prosodic representations as well as adults from the age of 9, but only in their complete form. Only adults were able to anticipate prosodic representations. In other words, only adults can identify the correct prosodic representations before the end of the utterance. Taken together, the results suggest that children acquire production skills earlier than perception ones.

The study by der Meulen et al. (1997) investigated prosodic abilities of normally developing children and children with specific language impairment in Dutch. Three age groups were studied; 4 year-olds, 5 year-olds and 6 year-olds. Subjects performed the prosody imitation task and the emotion identification task. In the prosody imitation task, subjects were required to imitate sentences with different linguistic and affective intonation contours. Linguistic contours represented questions and statements. Four affective states were chosen; happiness, sadness, anger and fear.

Results show that linguistic intonation is more readily recognised and reproduced by children than emotional intonation. Child speech rate was increasing with age, but it was slower than that of the corresponding adult stimulus sentence for all the children. The emotion identification task required subjects to identify recordings of emotionally intoned sentences. Sadness had the highest recognition level among the studied emotions, and happiness had the lowest level. The 4 year-olds scored worse than chance expectation. The identification performance was improving with age. Children with specific language impairment performed significantly worse than normally developing children only in the imitation test.

4.6 Conclusion

The concepts of bilingual language acquisition, particularly those related to phonetic learning, have been introduced in this chapter. While there exist studies on simultaneous acquisition of linguistic aspects of speech, to the best of our knowledge, no studies cover paralinguistic acquisition in simultaneous bilingual children. Thus, research in monolingual and second language acquisition of vocal affective expressions is presented. The study of how bilingual children produce affective speech in their two languages will help to address two important questions: whether bilingual children have one extended language system or two language-specific subsystems in affective speech production, and whether interactions may be observed between the two languages of bilinguals in the affective speech. In general, this research into paralinguistic acquisition of simultaneous bilinguals will contribute to the better understanding of the human capacity to acquire and use more than one language, and to adapt to the standards of more than one culture and society.

Chapter 5

Methodology

5.1 Introduction

This chapter covers important methodological issues in affective speech recording, which were taken into account for designing the present corpus of child affective speech. The whole process of corpus development is described in detail; the development of visual materials and recording procedure, the subject recruitment, audio digitisation and processing, description of all analysed acoustic parameters, annotation process and data extraction.

5.2 Corpus design

5.2.1 Recording methods of affective speech

One of the main difficulties in affective speech research is how to obtain an appropriate speech sample. This is the case for adult speech research, but proves even more difficult for child affective speech research. There are three possible ways to record affective speech; spontaneous speech, acting and induction, each with its advantages and disadvantages. After providing a short summary of these three recording possibilities and their discussion in light of child affective speech recording, the chosen methodology for this study will be described.

Spontaneous speech

Spontaneous speech is considered to display the most authentic emotions (Scherer 2003), but its collection is very difficult, and raises issues of copyright (for example, the usage of television recordings) and ethical problems. The ultimate objective of the spontaneous speech recording is to get affective speech in the most natural con-

ditions with people who do not even suspect they are being recorded. Besides ethical problems, spontaneous speech gives rise to another difficulty - each affective state must be annotated in the recorded data and must have a statistically viable number of samples. It is not always straightforward to identify a particular affective state, as researchers use different definitions and the number of affective states varies across researchers. Several attempts have been undertaken to study affective spontaneous speech of adults, amongst others; affective speech extracted from TV programmes in the Reading-Leeds database (Stibbard 2001), interviews on problematic topics and extracts from TV programmes in the Belfast database (Douglas-Cowie et al. 2000), recording of airport passengers (Scherer & Ceschi 2000), and uninterrupted daily recording of subjects (Campbell 2000). In spite of problems concerning naturalness and ethics, the attempts to study spontaneous affective speech are very important for the progress of affective speech research, as they look at the complex realisation of affective states in real life speech. However, research using children must conform to ethical requirements more strictly, and ethical concessions necessary for spontaneous speech recording are not seen as possible for children, as a particularly vulnerable subject group.

Acted affective speech

Acting of emotions is the most widespread way of affective speech recording in the existing research. It does not have the ethical difficulties described above, and it makes it possible to obtain studio-quality, controllable recordings. There are several important issues related to acted affective speech:

- *Quality of acted speech.* Researchers use both professional actors (Scherer et al. 2001, Banse & Scherer 1996) and ordinary people (Amir et al. 2000) to record acted affective speech. While professional acting is often criticised for unnatural exaggeration, naïve acting is argued to be flawed by the lack of professional training to encode affective states voluntarily in a recognisable way (Campbell 2000). In case of using actors, Scherer (2003) argues for the

importance of professional actors to have mastered Stanislavski techniques (auto-induction methods), as it closely replicates the natural phenomenon of people acting under display rules, and to provide appropriate instructions for the portrayals (e.g. realistic scenarios) of affective states. Another important requirement is to pass the recorded affective displays through an identification test and to work only with those affective displays which are being perceived reliably as such. In this case, the recorded material can be considered as 'a shared communication pattern for emotional information in a culture' (Scherer 2003).

- **Identification of acted versus genuine emotions.** People have an ability to distinguish between acted and genuine emotions (Aubergé & Cathiard 2003). Thus, even if listeners may recognise the intended emotions in the perception tests, they may be conscious of the fact that these emotions were acted (Campbell 2000). This ability is vitally important for success in interpersonal communications. Emotion is expressed continuously during interaction, and each participant attempts to 'read' other participants' emotional states. Work by Porter & Samovar (1998) shows that in large measure, culture affects the display and recognition of emotion by specifying how, when, in what social context, and by whom emotion is being displayed and recognised. Moreover, display rules are described as cultural influences on what people learn about the need to manage the appearance of particular emotions in various situations: 'Culture imposes display rules as children are taught what they may and may not do, often through unconscious observation, imitation of peers and adults, and/or through parental reinforcement' (Porter & Samovar 1998).

- **Stereotypicity of acted emotions.** Acted emotions are often criticised for their stereotypicity (usage of most prototypical and easily interpretable emotive correlates) and exaggeration. The stereotypicity comes from the fact that not all characteristics can be voluntarily controlled, and those that can be controlled, are often exaggerated to reinforce the intended expression (Zetter-

holm 2003). Acting of emotions can be compared with the voice imitation techniques. The level of control and individual preferences for different voice characteristics can vary across speakers, nevertheless, the imitation can be well recognised, if some crucial voice characteristics are successfully imitated (Zetterholm 2003). The nature of acting procedures implies that the actor aims at one emotion at a time with a strong intention to be understood without the additional help of context and semantic meaning, but emotional realisations in real life are much more complicated. A study by Scherer (2003) suggests that the emotional stereotypes can be universally understood across different cultures, but at the same time the level of recognition depends on the culture - people from the same culture have less difficulty understanding each other than people from different cultures.

The usage of acted affective speech is a valuable research approach. Nevertheless, it is important to take into account its disadvantages and to be careful in generalisations.

Induced affective speech

A wide range of procedures has been developed in the literature to induce affective speech in a controlled way. They include the reading of positive/negative self-statements and stories (Campbell 2000), the use of music or films (Pavlenko 2002), the use of interactive tasks and games on computers, e.g. the 'Wizard of Oz' (WoZ) technique (Audibert et al. 2004), and putting a subject into a situation meant to evoke a specific emotion. However, evoking certain negative emotions (for example, to scare someone), particularly in child subjects, could be seen as unethical, just as it may be unethical to record someone who is already scared. As a result of this problem the induced emotions are often mild, as if there were an inverse relation between the strength of the induction and the unethical value. The induction of affective states has the positive feature that it gives control over the stimulus; on the other hand, ethical questions and individual differences in the reaction to the

stimulus contribute to its disadvantage.

5.2.2 Developed method of child affective speech recording

Recording of acted speech does not contradict the ethical requirements to the research on children. Moreover, acting of emotions can be presented as a game to children, so that they can be more naturally involved in the expression of emotions. Acting has already been effectively used in the study of child facial expressions of emotions (Field & Walden 1983). Thus, this study uses acting as the principal method of affective speech recording.

The most common procedure of affective speech recording through acting is to ask subjects to read or to say some words or utterances in a specified affective state. Sometimes, subjects are given information on the situation, in which the desired affective state may be expressed, but visual materials are usually not provided. Subjects are often asked not to overplay, nevertheless it is difficult to control the intensity and the way of expressing emotions in this way. It was decided to use visual materials in support of acting in the present study. It is easier for the subjects to play an emotions if they see its visual realisation (Field & Walden 1983). Acting affective states with a particular arousal and intensity simply on demand can be too challenging a task for children, so it was found more appropriate and convenient to propose visual materials to which children can relate, and which suggest the desired intensity and arousal level of affective states for acting.

Visual materials were developed in a realistic manner together with a professional illustrator and based on the research of facial expressions (Ekman 2003). They represent a child expressing four emotions; happiness, sadness, anger and fear. This particular choice of emotions was motivated by the fact that they are used in the existing vocal perception studies of affective states by children (Rothman & Nowicki 2004, Gérard & Clément 1998, der Meulen et al. 1997) which can serve as a basis for comparison with the present work. The gender and age of the depicted child are adapted to that of the recorded subjects. The arousal types of expressed

emotions are described in Section 5.4.

The choice of realistic pencil drawing was intentional. While photos represent a concrete person, pencil drawing represents a more abstract image which makes it easier to associate oneself with a particular affective state. At the same time, caricature faces, often used in perception studies with children, are too abstract, they are more associated with cartoon characters than with real children or adults. A production study using caricature faces (Grichkovtsova 2002) as visual materials showed that some speakers changed their normal style of speaking in order to imitate typical cartoon voices.

The visual materials are presented to the children as a pile of randomized cards, including 16 repetitions for each emotion. The subject is instructed to say the token utterance (see below) in the same way as the child on the picture; thus the child is playing emotions through association with the drawing. Visual materials played a very important part in the recordings, they served as the reference of the expressed emotions, and as an affect-inducing material, as the subjects associate themselves with the picture. Detailed instructions and illustrations are given in Appendix C.

The usage of acted affective speech raises several important methodological issues, as has already been mentioned. The most important question concerns the quality of acted speech, because children are not professional actors. The developed methodology used the visual materials to help children to express a required affective state, but it was still essential to test the quality of affective speech acted by children through the identification test. This approach for the validation of acted affective speech has been suggested by Scherer (2003). Moreover, the decision was taken to make the corpus validation even stronger by recording a professional actress according to the same methodology, and comparing her level of identification with that of children. It is also important to comment that acted speech recording is a valuable research technique for the study of affective speech and for the progress in this complex research area, but it does not permit straightforward generalisations to the natural, real-life usage of affective speech.

5.2.3   Speech material

One token utterance was selected with a similar sound and prosodic structure for English and French: 'I see a banana there./Je vois une banane, moi'. The utterance presents a simple structure. The words used were relatively neutral and were thought not to have the potential to elicit or invite affective bias in English or French. French sentences were close translations of the English stimuli. There was still a potential problem of differences in the melody of the sentences, caused by different placement of stress in the two languages, but we used words with the same vowel in the utterance's main accented syllable to increase the cross-linguistic comparability of pitch and intensity. The sentence structure was designed to have an unstressed word after the final tone in the utterance, so that there would be enough voiced segmental material available for the final pitch accent to be fully realised. The utterance was a declarative, this type of sentence having more or less similar intonation patterns in the two languages.

5.2.4   Recording method of neutral utterances

Neutral was not represented as a facial expression card, neutral utterances were recorded in a separate test at the very beginning of the recording session. The idea was to record neutral utterances in a neutral context, and not as an intended state, played by the subject. In this test, the child was instructed to say the token utterance ('I see * there.'/'Je vois *, moi.') with a word from a pile of cards (picture naming) in which 16 repetitions of the original word (banana/banane) were mixed among the others. Half of the cards had a banana, the other half had different objects, like an apple, a strawberry, a coin, etc. Later all the utterances with the original token wording were selected for analysis. Cards with other than banana objects were added to make this recording less repetitive and more attractive to the children. Neutral utterances serve as a reference of affectively uncoloured, or linguistic usage of prosody.

5.2.5 Participants

Children

Bilingual children were recruited according to the following primary requirements:

1. to be raised as bilinguals in English and French simultaneously from birth;

2. to be born and living in Edinburgh (dialect control);

3. to have Scottish English as dominant language;

4. to use French on a daily basis and to receive some school instruction in it;

5. to be in the age range of 7 to 10 years old.

Six bilingual children have been initially recruited for the study. One bilingual child has been used for the pilot project to test recording and analysis methodology, and this subject was not included in the main analysis. The pilot project was an important step in the validation of the recording methodology. It allowed testing of the experimental set up and to see if the child was able to act emotions on the same neutral utterance with the help of visual materials. It also gave the possibility to establish the procedure for the annotation and acoustical analysis. After the pilot study, the recording of the cross-linguistic corpus of monolingual and bilingual children was started.

Five Scottish English/French bilingual children (B1, B2, B3, B4, B5) were recorded for the corpus. All of them correspond to the requirements for bilingual children. They were all born in Edinburgh, Scotland. They attend mainstream schools in Edinburgh, and attend the same French afternoon school twice a week. English is the dominant language for all the children, as it is the language of their school, their friends, and the country they live in. Child B1 is 10, child B2 is 7, child B3 is 9, child B4 is 8, and child B5 is 8, thus, the average age of the bilingual children is 8;4 years old with standard deviation of 1.1.

There are some differences in the families' situations and parental strategies regarding language usage; as a consequence the amount of input the children are

Table 5.1: Background linguistic profiles of recorded children

Child	Age	Place of birth/residence	Mother	Father	Home language
B1	10	Edinburgh, Scotland	French	Scottish	French/English
B2	7	Edinburgh, Scotland	French	Scottish	French
B3	9	Edinburgh, Scotland	French	French	French
B4	8	Edinburgh, Scotland	French	Scottish	French/English
B5	8	Edinburgh, Scotland	Scottish	French	French/English
MF	8	Normandy, France	French	French	French
MSE	8	Edinburgh, Scotland	Scottish	Scottish	English

receiving in French varies. Children B1, B2 and B4 have a French mother and a Scottish father, child B3 has both French parents, and child B5 has a Scottish mother and a French father. The parents of children B1, B4 and B5 use one-parent one-language strategy, but they speak English between themselves, so their home dominant language is English. Parents of child B2 choose to speak the less dominant language, French, at home. The father of child B2 is Scottish, but he speaks French with his wife and daughter. Child B3 has both French parents, and they speak French at home. Children B2 and B3 have French as their dominant home language. Child B5 has the least French input amongst all the children, as her mother, the principal caregiver, is Scottish, and these parents claimed the lowest frequency of visits to France. All the other families go to France at least once a year.

The variability in the individual family context of bilingual children is a well known issue (Gordeeva 2006). Circumstances, such as parents' linguistic strategies and preferences, number and length of visits to the country of the other language, etc., make it impossible to find bilinguals with identical backgrounds. This variability in the linguistic and family background of bilinguals explains why groups of bilingual speakers are always less homogeneous than groups of monolingual speakers. Therefore, care needs to be taken when interpreting group results in studies of bilingual speakers.

French monolingual children (children labeled MF1 to MF6) were recorded in Normandy, France. They are all 8 years old and have both French parents. Scottish monolingual children (children MSE1 to MSE6) were recorded in Edinburgh. They are 8 years old and have both Scottish parents. It was decided to have monolingual

children of the same age in the two languages for the higher level of control and reference, the two monolingual groups are thus directly comparable. French and Scottish monolingual children were matched to the average age of the bilingual children. The socio-economic status of children was also matched, all the children are middle class and their parents have university degrees. As all the bilingual children happened to be girls, it was decided to record only female monolingual children as well. Thus, the corpus has only female participants.

The described methodology allowed recording of a cross-linguistically comparable corpus of five bilingual Scottish-French children and twelve monolingual peers (six for each language) with an average age of 8 years. The information regarding linguistic and social background of recorded children was collected through a questionnaire filled in by the children's parents (it is presented in Appendix E).

Adults

One Scottish professional actress has been recorded, according to the same methodology as the children, with visual materials adapted to her age. She makes an additional control and reference for the recorded children. The recording of a professional actress, who has mastered the Stanislavsky acting method, allows testing of the recording methodology and comparison of the actress's production and the level of identification with that of children.

Parents of the recorded bilingual children could have also been an interesting object of the research. All of them are also bilinguals in English and French. But while the children are simultaneous bilinguals (they learn their two languages from birth), their parents are adult consecutive bilinguals, as they acquired one of their languages as a foreign language later at school and university. The study of affective realisations by adult bilinguals allows to investigate the affective organisation of consecutive bilinguals which may be different from that of both monolinguals and simultaneous bilinguals. Moreover, the comparison of parents' and children's affective realisations may give important information on the parents' influence on

the development of individual and language specific characteristics in the affective speech of bilingual children. Mothers of bilingual children were recorded in their first language according to the developed methodology. Nevertheless, after thorough consideration of the time necessary to complete and analyse this additional corpus, it was decided to postpone this work for further research.

5.3 Recording

All speakers were recorded with a TASCAM DA-P1 portable DAT recorder and a cardioid condenser Lavalier microphone (Audio-Technica AT831) either in the recording studio or in a quiet school room. The experiment was also recorded with a video camera. The bilingual children were recorded in two sessions (one for each language) by different experimenters (an English speaker and a French speaker) in order to control for language mode and code-switching (Grosjean 2001), with an interval of about two weeks between the sessions. This approach allowed creation of a monolingual mode of recording, the children spoke only the language of the session and did not use code-switching. The analysis of affective realisations in the monolingual mode is used to investigate the level of separation between the two languages by bilingual children, which is one of the research questions of the present study.

The concept of language mode was developed by Grosjean, who defined it as 'the state of activation of the bilingual's languages and language processing mechanisms, at a given point in time' (Grosjean 2001). It is a notion of a situational continuum ranging from a monolingual to a bilingual speech mode. In the monolingual speech mode, the bilingual deactivates (but never totally) one of the languages; this happens in the presence of and during conversation with monolingual speakers. In the bilingual mode, the bilingual speaker chooses a base language, activates the other language and calls on it from time to time in the form of code-switches and borrowings; this happens in the presence of and during conversation with other bilinguals. Code-switching is 'the alternative use by bilinguals of two or more languages in the

same conversation' (Milroy & Muysken 1995). Grosjean points out the necessity to account for the language mode in bilingual research. It helps to better understand how bilinguals process their two languages, separately or together, and it can partly account for problematic findings relating to such topics as language representation and processing, interference, code-switching, language mixing in various bilingual populations. He stresses that language mode is invariably present in bilingualism research as an independent, control or confounding variable and hence it needs to be heeded at all times.

The usage of two different persons for the recording of bilinguals (an English speaker for the English session, and a French speaker for the French session) created a monolingual mode of recording - the children spoke only the language of the session and did not use code-switching. The analysis of affective realisations in the monolingual mode is used to investigate the level of separation between the two languages by bilingual children, which is one of the research question of the present study.

The recording time of each session was about 30 minutes. The speech was digitised and stored at 44100 Hz. The analysis was performed on the speech down-sampled to 22050 Hz, as processing and storing capacities of informatic resources easily allowed it, and as people may perceive frequencies of up to 20 000 Hz. In total, 16 utterances for each of the studied affective states (anger, fear, sadness and happiness) and for the neutral realisation were recorded per individual monolingual speaker (80 utterances per speaker). Bilingual speakers had the recording in their two languages, thus they had the corresponding number of utterances in each of their languages. About 10% of the recorded corpus contained utterances with noise, hesitation or incorrect wording. It was decided to take 8 utterances for each emotion and the neutral state in the middle of the recording, where it was assumed that children were performing at their best. At the very beginning of the recording, children hesitated most and more often made errors in wording, as they were adapting to the experiment. They were starting to get tired by the end of the experiment.

The decision to take only half of the recorded corpus for the acoustic analysis was also motivated by the large number of recorded speakers and the feasibility of the project.

To summarize, twelve monolingual children, five bilingual children in two languages and one professional Scottish actress were recorded for the corpus, with a total of eighteen speakers. Eight repetitions of four affective states (anger, fear, sadness, happiness) and of a neutral state were selected from the recorded corpus for the acoustic analysis. Forty utterances were taken per each monolingual speaker, and 80 utterances per bilingual speaker. Thus, a total of 920 utterances was selected for acoustic analysis from the recorded corpus.

5.4 Arousal classification of the recorded corpus

According to the theory of emotional arousal (Scherer 2003), emotional categories have two levels of emotional arousal or activation; low arousal (LA) and high arousal (HA). Thus, anger can be realised as cold (or controlled) anger (LA anger) and hot anger (or rage) (HA anger), fear as anxious fear (LA fear) and panic fear (HA fear), sadness as depressed sadness (LA sadness) and despaired sadness (HA sadness), happiness as calm happiness (LA happiness) and elated happiness (HA happiness). Emotional arousal has an important impact on vocal expressions.

The children's acting of emotions was supported by visual material depicting HA anger, HA fear, LA sadness and LA happiness (copies of these visual materials are given in Appendix C). These types of arousal were chosen because they are better described in affective speech research and are easily depictable as visual expressions. It is important to say that many studies do not mention the arousal type of studied emotions. It was not seen as possible to give direct instructions to children concerning the type of arousal. The majority of realisations do appear to follow the intended arousal level. Nevertheless, some speakers appear to use an alternative arousal type. Children who used an alternative arousal for the expression of a particular affective state, did so consistently during the whole recording, thus

the level of arousal relates not to a particular utterance, but to all the utterances of an affective state expressed by the speaker.

The level of arousal was first judged perceptually on the basis of vocal characteristics of emotions summarised by Johnstone & Scherer (2000), the description of both high and low arousal forms for anger, fear, sadness and happiness is given in section 3.2 of Chapter 3. A number of speakers (BSE2, B5, MSE1, MSE4, MF6) appeared to target crying sadness, which involved higher arousal. MF2 and MSE6 appeared to realise only LA emotions. MF4 appeared to realise LA anger. B3 appeared to realise LA fear in both languages. The cases with alternative arousal were included in the identification test and in the analysis of individual production results. The results of these analyses will be discussed in relation to the level of arousal in Chapter 7 'Production results'.

5.5   Measurements

Using *Wavesurfer* software, acoustic measurements of pitch range variation, rhythm and speech rate were taken. Pitch range variation and speech rate have been identified as important acoustic correlates of affective speech in previous studies Johnstone & Scherer (2000). Rhythm has also been suggested as an acoustic correlate of affective speech Mozziconacci (1998), but it is less studied in affective research. All the utterances were manually labelled at the following pitch points: IF_0 (initial value of F_0), H_1 (the peak of the first accented word), V (valley, the lowest point between two peaks), H_2 (the peak of the main accented word, the main accent), FL (final low, the lowest value after the peak). In cases where there was no well-defined peak, an absolute F_0 maximum value was taken near the accented syllable. Segment duration measurements were taken manually on the whole utterance. An example of pitch and segmental annotation is shown in Fig. 5.1. The fundamental frequency and time values were then extracted automatically to data files with the help of ESPS algorithm, and again checked manually for any pitch perturbation or voice quality errors. The analysis settings for the detection of possible frequencies was set
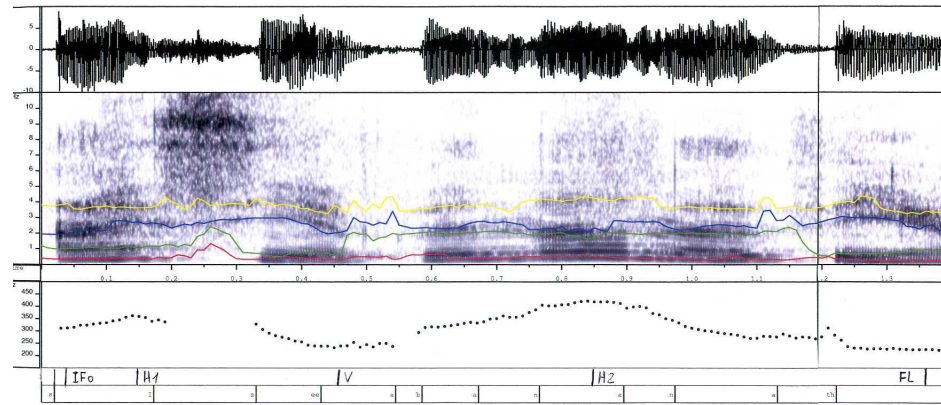
Figure 5.1: Example of pitch and segmental annotation for child B3 in English. The intended affective state was happiness.

at 50-1200 Hz. The choice of this range is justified by the fact that the recorded data represents child paralinguistic usage of pitch; some children produced pitch as high as 1100 Hz.

5.5.1 Pitch range

Following Ladd (1996), pitch range is described by using two partially independent measures of variation: overall level and span. Overall level refers to the "height" of a speaker's voice, span refers to the width of pitch frequencies covered by a speaker (how big the excursions are). These measures have been shown (Patterson 2000) to give better and perceptually relevant results, than the more commonly used measures, such as statistical moments (mean, standard deviation, difference between maximum and minimum F_0, etc.).

According to Patterson (2000), sentence final low and post-accent valleys are considered as competing linguistic measures for level. Both of them were found to be successful in capturing speaker characteristics, though sentence final low is reported as slightly more effective. The difference between non-initial accent peaks and post-accent valleys is shown as the best span measurement by Patterson. While Patterson's corpus consisted of read passages of six or seven long utterances, the corpus of the present study is based on one short utterance, pronounced orally with different intended affective states. The utterance can potentially have two

peaks and one valley. The main accent peak and final low were the most reliable intonational points across different speakers, languages and emotions. Thus, the difference between the main accented peak and the final low ($H_2 - FL$) was calculated as a global pitch span measurement for the utterance, coded as *span*. The final low, or FL, is the lowest value after the peak of the main accent, which is equivalent to Patterson's 'sentence final low' (Patterson 2000), it is coded as *level*.

Research ('t Hart et al. 1990, Hermes & Rump 1994) suggests that a linear representation of the number of periods per second does not correspond to the auditory impression of pitch changes. For example, the difference between 400 Hz and 450 Hz is perceived as a smaller pitch change than the difference between 100 Hz and 150 Hz. Several measurement scales have been claimed to give better auditory representation, of which the most commonly used are the semitone scale (st) and the equivalent rectangular bandwidth, ERB. The semitone scale is logarithmic, as it represents differences between two tones, so it is most applicable to span which represents the difference between some topline and some bottomline pitch points. It was also suggested as the best representation of span by Patterson (2000). The derivation of the equivalent rectangular bandwidth is designed to describe the frequency selectivity of the auditory system, based on the perception of a signal through noise, ERB is used as the unit of frequency. Hermes & Rump (1994) claim that the most appropriate scale for measuring F_0 is the ERB scale. Though Patterson (2000) has not found any significant differences of level measurements between Hz or ERB, it was decided to opt for ERB, especially since the corpus consists of both child and adult data. Thus, apart from span measurements, which were calculated in semitones, level measurement and all pitch points measures in general were converted to ERB.

The conversion of frequencies in Hertz into semitones for measurement *span* is performed according to the following formula:

$$st = \frac{12}{\log_{10}(2)} \times \log_{10}\left(\frac{H_2}{FL}\right),$$

where H_2 is the fundamental frequency of the main accented peak in Hz, FL is the fundamental frequency of the final lowest point in Hz, and $\log_{10}$ is the logarithm function to base 10.

The conversion of frequencies in Hertz into equivalent rectangular bandwidth (ERB) is performed according to:

$$ERB = 16,6 \times \log_{10}(1 + \frac{F_0}{165,4}),$$

where F_0 is the fundamental frequency in Hz, and $\log_{10}$ is the logarithm function to base 10.

5.5.2 Pitch turning points

Five pitch points which were thought to best capture the pitch changes in the utterance were measured: IF_0 (initial value of F_0), H_1 (the peak of the first accented word), V (valley, the lowest point between two peaks), H_2 (the peak of the main accented word - main accent), FL (final low - the lowest value after the peak). All the pitch points were transformed following two normalisation methods, Z-scores and percentage transformations. The normalisation methods have been developed and used in the linguistic studies of pitch (Rose 1987) to circumvent unnecessary complications due to individual variability in pitch range usage. It was decided to test the two methods of normalisation in the present study of paralinguistic usage of pitch.

Pitch range is a complex acoustic parameter to study, as it includes several levels of representation. Let us recall that as explained in Chapter 3, Laver (1994) specifies the following types of pitch range: organic pitch range, paralinguistic pitch range, linguistic pitch range and pitch span. The placement of both the linguistic range and paralinguistic range within the organic range is language-specific. Individual speakers of a given language thus adapt their placement of linguistic and paralinguistic ranges within the organic range accordingly; nevertheless this pitch range placement shows variability across speakers. This variability between individual speakers ex-

plains the difficulty to study the pitch range, and the necessity to normalise pitch range across individual speakers in order to compare between speakers and between groups of speakers. Two methods of normalisation (Rose 1987) have been widely used in the linguistic studies of pitch range: Z-score transformation and percentage transformation.

Z-scores normalisation

Z-scores are a special application of transformation rules. The z-score for an item, indicates how far and in what direction, that item deviates from its distribution's mean, expressed in units of its distribution's standard deviation. The mathematics of the z-score transformation are such that if every item in a distribution is converted to its Z-score, the transformed scores will necessarily have a mean of zero and a standard deviation of one; in other words the distribution of items thus transformed is centered on zero and scaled to the standard deviation. Z-scores are sometimes called "standard scores". The Z-score transformation is very useful when seeking to compare the relative standings of items from distributions with different means and/or different standard deviations. Z-scores are also specially informative when the distribution to which they refer, is normal (Gaussian). In every normal distribution, the distance between the mean and a given Z-score cuts off a fixed proportion for each possible score.

For *each* pitch point measurement (IF_0, H_1, V, H_2 and FL), the corresponding values of the fundamental frequency F_0 were transformed into Z-scores according to the following formula:

$$Z = \frac{X - \mu_{\text{neutral}}}{\sigma_{\text{neutral}}},$$

where X is a measured F_0 value for a given affective state and a given speaker in ERB, and μ_{neutral} and σ_{neutral} are the mean and the standard deviation of F_0 values for neutral utterances produced by *all* child speakers. Z-score analysis gives the distance between the measured value X and the mean μ_{neutral} as a number Z times the

standard deviation σ_{neutral}. Therefore, this approach allows appropriate comparison not only between individual child speakers, but also between child groups, since for a given pitch point, the average, μ_{neutral}, provides a unique baseline against which particular F_0 values, including individual neutral values, can be compared. Moreover, comparison between mothers and children, and actress and children are also possible this way. Note that because of their lower pitch, adults were not included in the calculation of μ_{neutral} and μ_{neutral} since this would introduce a bias that would not only prevent comparison between children, and between groups of children, but also between adults and children.

Percentage normalisation

For each measurement, F_0 values were transformed to a percentage of the speaker's pitch range according to the following formula:

$$ERB_{\text{norm}} = \frac{ERB - ERB_{\text{min}}}{ERB_{\text{max}} - ERB_{\text{min}}} \times 100,$$

where ERB_{norm} is a measured F_0 value for a given affective state and a given speaker in ERB, and ERB_{min} and ERB_{max} are the minimum and maximum F_0 value used by the speaker in the entire corpus. For bilingual children, these are the minimum and maximum F_0 values used by the speaker across the two languages. This type of normalisation allows to abstract from the specific pitch range of individual speakers.

Both types of normalisation were computed on the five measured pitch points: IF_0, H_1, V, H_2 and FL. In the course of the corpus analysis, it was found out that several children and adults used falsetto to express some of their emotions. Falsetto is a phonation setting which results in a particularly high fundamental frequency: the bottom end of the falsetto range overlaps with the top end of the voiced phonation range. Thus, the percentage normalisation for those speakers who used falsetto in their affective expressions may show biased results: those pitch points which are realised in the normal voiced phonation are transfered to smaller values (maximum F_0 value, used for percentage calculation, comes from the falsetto range), than pitch

points of those speakers who use only normal voiced phonation. That is why it was decided to report and analyse pitch values transformed only to Z-scores.

5.5.3 Rhythm

Speech rhythm is the *'complex perceptual pattern produced by the interaction in time of the relative prominence of stressed and unstressed syllables'* (Laver 1994). Two methods are frequently used to calculate the acoustic correlate of speech rhythm; the first, by Ramus et al. (1999), the second, by Low et al. (2001).

Ramus et al. (1999) suggests that the proportion of vocalic intervals $\%V$ and the standard deviation of the intervocalic interval duration σIV in a sentence may be used as direct correlates of syllable structure and thus of rhythm class. A low vowel percentage ($\%V$) and high standard deviation (σIV) in intervocalic interval length is associated with stress-timing. The main criticism (Grabe & Low 2002) of this method of measuring speech rhythm is that it does not take account of changes in speaking rate.

Low et al. (2001), Grabe & Low (2002) proposed a different acoustic correlate for speech rhythm. The Pairwise Variability Index (PVI) measures rhythmic variability by calculating the mean difference in the duration of successive intervals. The PVI is calculated separately for successive vocalic and intervocalic intervals. Intervocalic intervals stretch from the offset of one vocalic interval to the onset of the next. Grabe and Low suggest the application of a normalising procedure for vocalic intervals, since their duration is very much affected by changes in speech rate (Gay 1981). A high PVI corresponds to a great amount of rhythmic variability which is typical for stress-timed languages. The variability of vocalic intervals is related to the vowel reduction which seems to be a characteristic of stress-timed languages, such as English and German (Dauer 1983). High intervocalic variability depends on the complexity of syllable structures which permits a greater variation in the number of consonants around the syllable nucleus. Complex syllable structure is another characteristic of languages that are traditionally classified as stress-timed

(Abercrombie 1967, Dauer 1983). A low PVI corresponds to a small amount of rhythmic variability, and low intervocalic variability depends on a simple syllable structure, which are both characteristic of syllable-timed languages. French is described as syllable-timed language (Abercrombie 1967, Dauer 1983).

The method devised by Low and Grabe was chosen for the present study, particularly because it has already been successfully applied in bilingual rhythm acquisition (Whitworth 2003). The final syllable of each of the utterances was not included in the calculation of the PVI, since final lengthening presents a particular phenomenon (Grabe & Low 2002). Thus, final syllable was measured and analysed separately.

Normalised Pairwise Variability Index for vocalic intervals (PVI)

$$PVI = \frac{1}{m-1} \sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| \times 100,$$

where m is the number of intervals and d_k the duration of the kth interval.

Raw Pairwise Variability Index for intervocalic intervals ($rINT$)

$$rINT = \frac{1}{m-1} \sum_{k=1}^{m-1} |d_k - d_{k+1}|,$$

5.5.4 *Speech rate*

Speech rate was calculated as the number of syllables per second (syll/s).

5.5.5 *Voice quality*

A full auditory and acoustic analysis of the voice quality was not possible to perform in the scope of the present study. Nevertheless, the rich usage of voice quality in the production of the recorded speakers makes it important to give some general description. Using definitions proposed by Laver (1994) and presented in Chapter 3, the most prominent phonation settings were taken on each utterance; normal voicing, creakiness, breathiness, harshness, falsetto and whisperness. Different phonation types are illustrated by the examples from the corpus, see Figures 5.2-5.6. The presence of smiling (lip spreading) has also appeared as an important voice quality
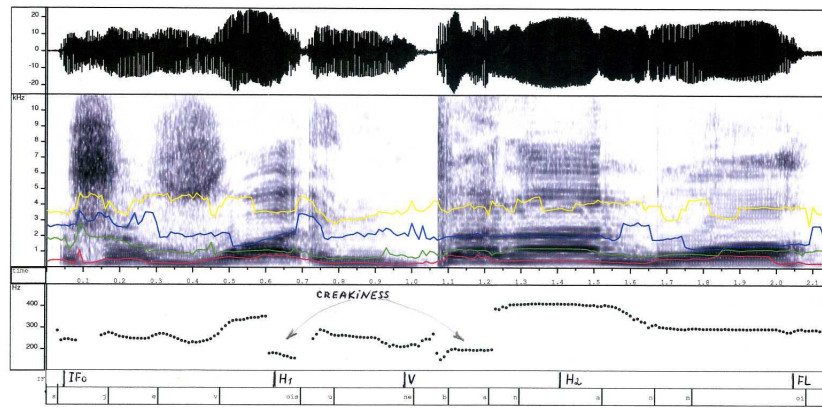
Figure 5.2: Example of creakiness, used for the expression of anger. Creakiness has a characteristic low fundamental frequency with aperiodic glottal pulses. The parts of the utterance with creakiness are marked.

parameter in the vocal expressions of happiness in the analysed corpus, hence it was decided to note smiling as well.

For each utterance of the analysed corpus, an auditory analysis of the voice quality was carried out: the most prominent phonation setting (normal voicing, creakiness, breathiness, harshnesss, falsetto and whisperness) was evaluated. To facilitate this evaluation work, a visual examination of the acoustic signal was also performed (as examplified on Figs. 5.2-5.6). Each of the five affective states expressed by an individual speaker was represented by eight utterances. To deal with the variability of the phonation setting across the eight utterances produced for each affective state, it was decided to proceed as follows: the identified phonation setting was assumed to be consistently used by the speaker for a particular affective state and reported in production results (see Appendix), only if the majority of the utterances (above 50%) contained this phonation setting. Thus, if, for example, a speaker used creakiness in two utterances and normal voicing in six utterances for anger, this affective state was reported as realised with normal voicing. Note that in the few cases where there were 50% of normal voicing and 50% of an other phonation setting, the latter was considered the most prominent. The description of phonation settings mainly serves to present the voice quality in the corpus and to draw attention to its prominence for paralinguistic information. The full analysis of voice quality will be the domain of the future exploration of the recorded corpus.
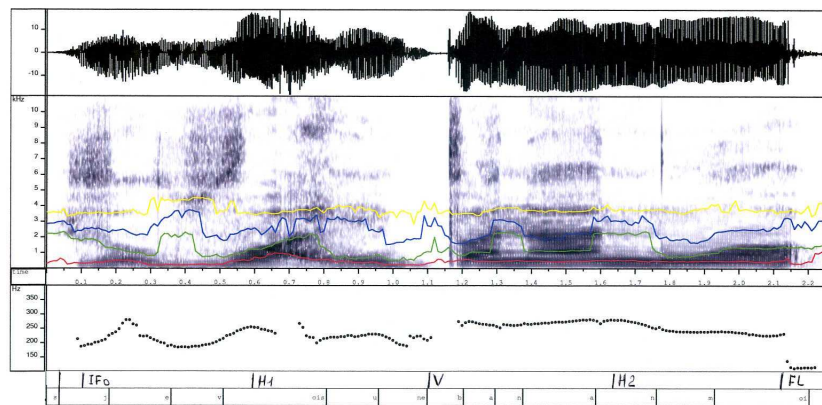
Figure 5.3: Example of harshness, used for the expression of anger. Harsh voice has an aperiodic noise, which characterises both harsh and creaky voices, but a normal fundamental frequency distinguishes it from a creaky voice quality.
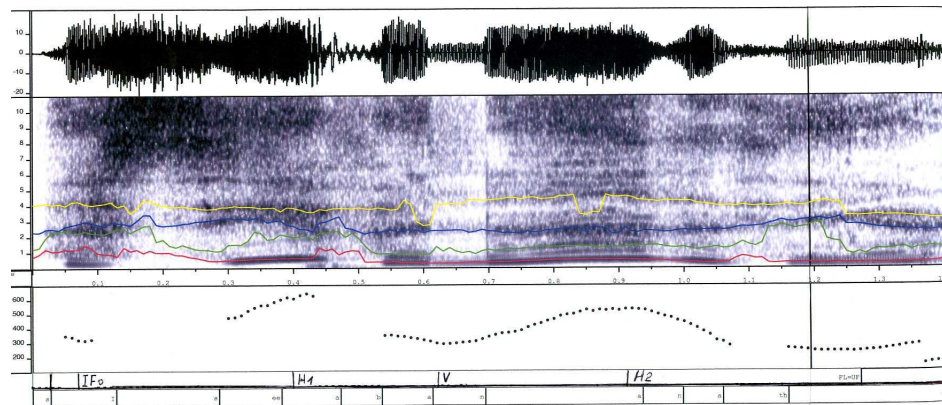
Figure 5.4: Example of breathiness, used for the expression of fear. The breathy voice is characterised by a high rate of air flow, accompanied by friction. It is possible to see an overall decreased acoustic intensity and noise in spectrum.
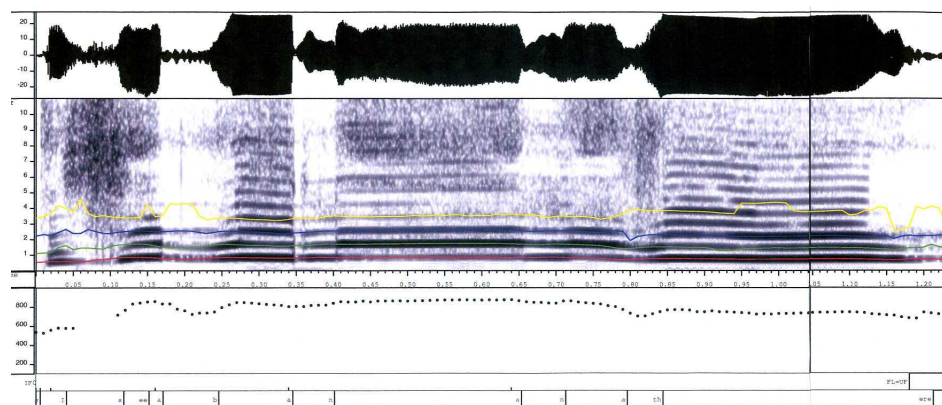
Figure 5.5: Example of falsetto, used for the expression of fear. The whole utterance was pronounced with an extremely high fundamental frequency, about 800 Hz.
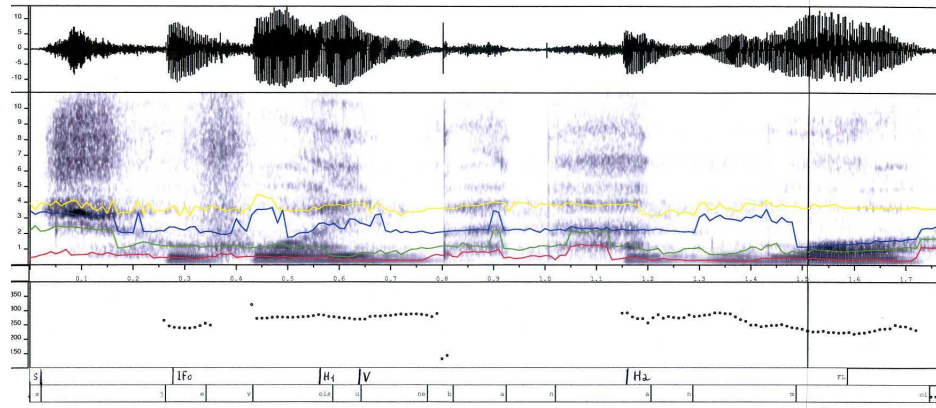
Figure 5.6: Example of whisper, used for the expression of fear. It is possible to see that the middle part of the utterance has no fundamental frequency, which is characteristic of whisper.

5.6 Conclusion

The size of the developed corpus (18 speakers, 920 utterances, more than 9 hours of recording) proved very demanding in terms of work and time required for its analysis, especially taking into account the choice to follow a *linguistic approach* in acoustic analysis, and the number of measured acoustic parameters. For comparison, other studies of affective speech, using linguistic methods of pitch analysis, have a smaller number of analysed utterances; Mozziconacci & Hermes (1999) had 315 utterances, Paeschke & Sendlmeier (2000) had 213 utterances, Bänziger & Scherer (2005) had 144 utterances. Studies using automatic statistical pitch analysis are much more numerous and they usually have a larger number of analysed utterances: Slaney & McRoberts (1998) had 509 utterances, Cohn & Katz (1998) had 621 utterances, Makarova & Petrushin (2003) had 600 utterances.

In the light of this, the present study demonstrates a significant difference. The successful completion of detailed acoustic analysis of such a large corpus (18 speakers, 920 utterances) was made possible thanks to effective usage of numeric means for data extraction and manipulation; the annotation was performed manually, which required a considerable amount of time and personal effort and manipulation of data thus generated, and required specific writing of scripts so that repetitive tasks could be automated as much as possible. Even if generalisation of the results should be

still performed carefully, the number of analysed speakers allows better observation of individual and general tendencies in the data of the present study, and it makes a contribution to the area of affective speech research.

Chapter 6

Perception experiment: Method and results

6.1 Perception experiment

According to the developed methodology described in Chapter 5, child affective speech has been recorded through acting with the help of visual materials. One of the important methodological requirements of research based on acted affective speech is the perceptual validation of the recorded corpus. The recorded expressions may be considered as reliable representations of affective states if they are successfully identified. Moreover, taking into account the fact that children are not professional actors, the perceptual validation is especially important, as it tests whether children are able to enact the studied affective states (anger, sadness, fear and happiness) according to the developed methodology. The recorded corpus also contains one professional actress who is included to strengthen the perceptual validation of the child-acted affective speech, this makes it possible to test whether the level of identification for children's acted affective expressions is comparable to that for a professional actress.

Apart from the corpus validation, the perception test addresses directly one of the research questions: 'How do monolingual French and Scottish adults identify affective speech, encoded by bilingual and monolingual children in the two languages?' The perception test with monolingual French and Scottish adults makes it possible to investigate if there exist cross-linguistic differences in the perception of child affective speech in Scottish English and French. Moreover, the test allows comparison of the level of identification for the affective speech encoded by monolingual and bilingual children in the same language. It is of interest to see if the affective states of bilinguals are easier/more difficult for identification than those of monolinguals, or are simply comparable. The perception experiment also serves for the evaluation

of affective speech encoding strategies used by individual speakers, as some strategies may be more successful than others. To answer these questions, perception test results are analysed both at the group level (monolingual *versus* bilingual children) and at the individual level for each child.

6.1.1 Stimuli description

The corpus of affective speech selected for the acoustic analysis contains 920 utterances. Each studied affective state (anger, sadness, happiness, fear + neutral) is represented by eight utterances. Eighteen speakers are included in the corpus; five Scottish English-French bilingual children (recorded in the two languages), six French monolingual children, six Scottish monolingual children and one Scottish actress. A representative number of utterances (five out of the total of eight utterances) was randomly selected for each of the studied emotions, i.e. anger, fear, sadness and happiness, produced by individual speakers in a given language. Neutral utterances were not included in the perception test. Thus, the total number of stimuli for the perception test was 460 utterances.

6.1.2 Participants

Twelve French monolingual subjects (three females and nine males) were recruited at the Université de Caen, Normandy, France. Twelve Scottish English monolingual subjects (six males and six females) were recruited in Queen Margaret University College, Edinburgh, Scotland. All the French subjects were originally from Normandy. All the Scottish subjects were originally from Edinburgh. The recruited French and Scottish subjects were students and professionals working in the universities. Their average age was 28 years old, with standard deviation of 9.5. None of the subjects reported having any hearing difficulty.

6.1.3 Procedure

Special software for psycholinguistic perception tests *Perceval* was used for the experiment design. Perceval was developed in the Laboratoire Parole et Langage, Aix-en-Provence, France, and is free for academic purposes Guion (2003). The subjects were instructed to listen to children's affective expressions, and to identify the intended affective state from four possibilities; anger, fear, sadness and happiness. The experiment was run on a computer with high quality headphones in a quiet language laboratory room. The instruction was displayed on the screen: 'You will hear emotions played by children. Choose which emotion the child is playing and press the key: 1 - Anger, 2 - Fear, 3 - Sadness, 4 - Happiness'. The names of emotions and their key numbers were displayed on the screen during the experiment. All the stimuli were presented in a randomised order. French and English utterances pronounced by monolingual and bilingual children were mixed. The whole experiment was run by the Perceval software, responses and response times were automatically recorded in the data file. On average, the experiment took about 40 minutes.

6.2 Perception test results for child speakers' groups

Perception test results are presented through a confusion matrix (Table 6.1). Encoded affective states are plotted against the decoded affective states. This enables study of the percentages of correct responses and the pattern of confusions. Affective states were recognised with a higher accuracy rate than chance: Cohen's kappa coefficient [1] (κ) for French listeners is found to be $\kappa = 0.418$, $p < 0.0001$, and for Scottish listeners $\kappa = 0.381$, $p < 0.001$. The correlation between French and Scottish listeners was significant (Cohen's kappa $= 0.09$, $p < 0.001$), but suggests also some differences, as it is close to zero. Chi-Square tests were applied to each subgroup of speakers and listeners to investigate the observed and expected frequencies of

---

[1]Cohen's kappa coefficient is a statistical measure of inter-rater reliability (Cohen 1960). Since the calculation of κ takes into account the agreement occurring by chance, it is generally viewed as more robust than percentage agreement calculations. If the raters are in complete agreement, $\kappa = 1$; and if there is no agreement other than what would be expected by chance, $\kappa \leq 0$.

correct responses, and their statistical significance. The table 6.1 shows the percentages of emotion recognition and confusion for two groups of listeners: Scottish English and French. Groups of speakers are presented separately: BSE (bilinguals in Scottish English), BF (bilinguals in French), MSE (Scottish English monolinguals), MF (French monolinguals), ASE (Scottish English actress). Statistically significant results are displayed in bold.

The perception results for anger (Table 6.1) show that the Scottish actress has the highest level of identification: 80% for Scottish listeners and 73.3% for French listeners. The expressions of anger by bilingual children are identified in the two languages, but with better rate in French (71.9% for Scottish listeners, 74.1% for French listeners), than in English (62.2% for Scottish listeners, 52.9% for French listeners). The same observations may be derived from the results of monolingual children: French monolingual children have higher identification level (55.2% for Scottish listeners, 69.6% for French listeners) than Scottish monolingual children (28.8% for Scottish listeners, 51.3% for French listeners). Note that Scottish monolingual children are identified better by French listeners than by Scottish listeners, who significantly confused anger with sadness and happiness.

Fear results (Table 6.1) show that bilingual children have been successfully identified both in English and in French. In contrast, fear encoded by French monolingual children has not been successfully identified by Scottish listeners (18.2%). French listeners have identified fear just above chance (28.5%), but it was significantly confused with other emotions. Another interesting observation is that the Scottish actress's fear is identified by Scottish listeners at 98.2%, while French listeners confused her fear with sadness (53.3% *versus* 46.7%).

The identification of sadness (Table 6.1) was significant across French and Scottish listeners for all groups of children. The main difference between French and Scottish listeners is that Scottish listeners also show significant confusion of sadness with fear for all the groups of children, except French monolingual children. Even sadness, encoded by the Scottish actress received a significant confusion with fear

(sadness 47.3% *versus* fear 52.7%). French listeners showed higher identification level for sadness across all speakers. Two significant confusions with fear occurred only for bilingual children in English and for Scottish monolingual children. Interestingly, the Scottish actress has been successfully identified by French listeners at 84.4%.

Overall, happiness was successfully identified in the present study. Results displayed in Table 6.1 also show that bilinguals' emotions not only were identified with higher accuracy in Scottish English than in French by Scottish listeners (70.9% *versus* 62.9%), but also by French listeners (56.0% *versus* 44.2%). Note that Scottish listeners showed higher levels of identification of bilinguals, both in English and French, than did the French. Happiness expressed by both French and Scottish monolingual children was confused with sadness by Scottish listeners, but not by French listeners (see last column of the 'Happiness block' at the bottom of Table 6.1. Note also that French listeners identified better happiness, expressed by the Scottish actress (60.0%), than Scottish listeners who confused it with fear (40% *versus* 36.4%).

The perception test showed that monolingual children could encode the four affective states, i.e. anger, fear, sadness and happiness, recorded according to the developed methodology. The analysis of group results suggests that there are some cross-linguistic differences in the perception of child affective states by French and Scottish adults. It was found that each studied affective state had a specific pattern of identification, depending both on the group of listeners and on the group of speakers. Several patterns of identification by Scottish and French listeners were observed. Both French and Scottish listeners could identify a group of speakers in a similar way (for example, sadness encoded by French monolingual children was identified at 65% both by Scottish and French listeners). One group of listeners could be more successful in the identification of a particular group of speakers. Listeners could show better identification in their native language (for example, Scottish listeners identified fear of the Scottish actress at 98%, while French listeners

only at 54%). Interestingly, several cases were observed when listeners performed in the unknown language better that listeners for whom it was the native language (for example, French listeners identified sadness encoded by the Scottish actress at 84%, while Scottish listeners only at 47%). This phenomenon was especially noticeable in the perception of affective states encoded by bilingual children: for example, happiness encoded by bilingual children was identified at a higher rate by Scottish listeners than by French listeners in both English and French.

The answer to the question whether bilinguals' affective states are easier or more difficult to identify than those of monolinguals does not appear straightforward. The results of the perception test show that the level of identification of affective states encoded by bilinguals is high. At the same time, two interesting phenomena were observed. The identification of bilinguals' affective states may be better in one of the two languages for both groups of listeners. As, for example, anger of bilingual children is identified with a higher rate in French than in English by both Scottish and French listeners. Moreover, one group of listeners may be more successful than the other in the both languages (in the native, but also in the unknown language).

It would be of interest to understand the reasons for the observed phenomena in the perception of monolingual and bilingual children by Scottish and French listeners. In order to do this, the analysis of the perception test on the level of individual speakers and the analysis of the production results is necessary.

6.3 Perception test results for individual child speakers

Series of statistical tests were performed to study the level of identification and patterns of confusion for each individual speaker. Chi-Square tests were applied to each child speaker to investigate the observed and expected frequencies of correct responses, and their statistical significance. These tests were calculated on the total of all the responses, Scottish and French listeners were not separated. These results are presented in Table 6.2. They allow us to study the level of identification for each affective state, encoded by individual speakers. Moreover, complementary

Chi-Square tests were performed on the perception of stimuli encoded by bilingual children, in this analysis the identification by Scottish and French listeners was analysed separately. This complementary analysis allows a more profound insight in the identification of bilingual speech. The results are presented in Table 6.3 in percentage figures. Results which showed significance in Chi-Square tests are displayed in bold.

The results of the perception test for individual speakers, displayed in Table 6.2, show the level of identification for the studied affective states across all child speakers and the Scottish actress. For bilingual children, there was only one case when an affective state encoded by the child was not significantly identified (anger for bilingual child B1 in English). In all the other cases, the intended emotions by bilingual children were identified above chance. The results for the Scottish actress show that her affective expressions were successfully identified. The analysis of the perception data for monolingual children show that there are cases when one or even several affective states were not identified for some individual speakers.

The results for 'anger' show that bilingual children B2 and B3 have successfully encoded anger, as it was recognised by listeners at the maximum level (above 97%) in their two languages, which is even higher than for the professional actress whose anger was identified at 77%. Anger expressed by bilingual children B4 and B5, was identified above chance in the two languages, but with a higher rate in French than in English: 61% *versus* 37% for B4, and 63% *versus* 46% for B5. For child B1, anger was identified only in French (42.9%). Anger of three Scottish children (MSE2, MSE5, MSE6) was not identified. Angry expressions of all French monolingual children were successfully identified.

Fear results for individual bilingual children (Table 6.2) show that this affective state was decoded above chance. Fear expressed by child B2 was identified with a similar high rate in the two languages (73% in English and 70% in French). For child B5, fear was identified with a higher rate in English than in French (79% *versus* 49%). The level of identification for the fear of these two children was comparable

to the professional actress who had 78%. The level of identification for fear of the other bilingual children was significant, but lower. Fear of three Scottish children (MSE2, MSE5, MSE6) was not identified. These are the same children whose anger was not successfully identified. For French monolingual children, only fear encoded by children MF1 and MF3 was identified above chance.

Sadness was identified with a higher rate in French than in English for two bilingual children (B1 and B2). Sadness expressed by child B4 was identified with a high rate in the two languages (78% and 77%). Children B3 and B5 have the same level of identification for their sadness in French, at 47%. Only sadness of one Scottish child (MSE5) and one French child (MF5) was not identified. Sadness of three French children (MF2, MF3, MF4) was even identified with a higher level than that of the professional actress (64%). Sadness of four Scottish children (MSE1, MSE2, MSE3, MSE4) was significantly confused with fear.

Happiness was successfully identified for bilingual children, some children had an especially high level of identification, for example, child B2 received 94% in English, and child B3 received 87% in English. Happiness of all Scottish and French monolingual children was identified above chance. Happiness encoded by the Scottish actress received 49% of identification.

Results of the perception test for individual speakers showed that affective states of some monolingual children were not successfully identified. Across Scottish monolingual children, only two children (MSE3 and MSE4) received high levels of identification for the four studied affective states. MSE1 was also statistically recognised, though significant confusions are present for all her affective states. Only sadness and happiness were recognised for MSE2. Interestingly, MSE5 was identified significantly as happy in all the affective states. Only sadness was identified for MSE6. Affective states of monolingual French children showed higher identification than those of Scottish children. Anger, sadness and happiness were successfully identified for all French monolingual children (except child MF6 whose sadness was not identified). Fear received the lowest level of recognition for French children, only

two children MF1 and MF3 received above-chance identification. An interesting observation was made about child MF2 - all her affective states received a significant confusion with sadness.

6.3.1 Perception of bilingual children by Scottish and French adults

Complimentary Chi-Square tests were performed on the perception of stimuli encoded by bilingual children. The level of identification was calculated separately for French and Scottish listeners in order to investigate better the variability observed in the perception of bilingual affective speech. The results are presented in Table 6.3 in percentage. These results indicate whether a particular affective state of a bilingual child is more easily identified by Scottish or by French listeners. Interestingly, the same phenomena which were observed on the group level (see section 6.2) were noticed for individual bilingual children. Some affective states were identified with the same level by both groups of listeners: for example, anger encoded by child B3 in French was identified by French and Scottish listeners at 100%. Some affective states were identified more successfully in one of the languages by the both groups of listeners. For example, French anger, encoded by child B4, was identified with a higher level than English anger both by Scottish and French listeners. Some affective states were identified more successfully by listeners in their native language: for example, happiness of child B1 in English was identified at 69% by Scottish listeners and at 40% by French listeners. Some affective states were identified better by listeners in the unknown language: for example, happiness expressed by child B3 in French was identified by Scottish listeners at 92.7% and by French listeners at only 43.2%. These important results suggest the existence of cross-linguistic and cross-cultural differences in the affective speech of English and French, it is also possible to assume that interactions may be present in the affective speech of bilingual children.

6.4 Conclusion

To summarise, affective expressions played by bilingual children are identified by adult listeners. In their majority, affective expressions played by monolingual children are also recognised by adults. Nevertheless, there were some individual speakers whose displays of specific affective states were not identified. The proportion of not-identified affective states represents 16% of the whole child corpus. The perception test proved its importance for the present study, as it allowed us to validate the recorded corpus of child affective speech. Those affective states which did not receive a significant level of identification were analysed acoustically and presented in the production results for individual speakers, but they were not included in the group analysis of production data. The variability observed in the perception tests was found to be related to the group of listeners, to the groups of child speakers and to the individual speakers. The discussion of the reported results from the perception test will be presented in Chapter 8, together with the discussion of the production results, where the research questions will be addressed as well.

Table 6.1: Results of the perception test for Scottish and French listeners, presented in a confusion matrix. Groups of speakers are presented separately: BSE, BF, MSE, MF, ASE. The list of encoded emotions by children is on the first column of the table. Perceived emotions are listed in the row at the top of the table. The identification level is displayed in percentages, statistically significant values are in bold.

Encoded emotions	Listeners	Speakers	Anger	Fear	Sadness	Happiness
Anger	Scottish	BSE	**62.2**	8.5	17.8	11.5
		BF	**71.9**	5.8	16.8	5.5
		MSE	**28.8**	6.4	**35.5**	**29.4**
		MF	**55.2**	3.9	27.3	13.6
		ASE	**80**	10.9	9.1	-
	French	BSE	**52.9**	16.9	20.9	9.3
		BF	**74.1**	5.8	12.9	7.1
		MSE	**51.3**	13.4	10.4	24.9
		MF	**69.6**	3.7	13.7	13.0
		ASE	**73.3**	11.1	8.9	6.7
Fear	Scottish	BSE	5.0	**54.6**	18.5	21.9
		BF	6.2	**54.4**	18.6	20.8
		MSE	3.4	**44.5**	10.3	**41.7**
		MF	8.2	18.2	**37.1**	**36.5**
		ASE	-	**98.2**	1.8	-
	French	BSE	7.9	**61.6**	19.0	11.6
		BF	15.1	**48.9**	22.2	13.8
		MSE	13.0	**37.2**	16.5	**33.3**
		MF	16.7	**28.5**	**27**	**27.8**
		ASE	-	**53.3**	**46.7**	-
Sadness	Scottish	BSE	0.3	**37.8**	**59.8**	2.1
		BF	0.4	**49.2**	**49.2**	1.1
		MSE	0.3	**41.5**	**44.5**	13.6
		MF	4.7	21.3	**64.9**	9.1
		ASE	-	**52.7**	**47.3**	-
	French	BE	4.7	**26.1**	**61.5**	7.7
		BF	0.5	24.5	**73.1**	1.9
		MSE	4.5	**25.4**	**48.5**	21.6
		MF	2.7	20.8	**65.0**	11.5
		ASE	-	15.6	**84.4**	-
Happiness	Scottish	BSE	6.9	9.1	13.1	**70.9**
		BF	6.5	10.5	20	**62.9**
		MSE	3.9	3.9	**26.7**	**65.5**
		MF	8.0	5.4	**36.9**	**49.7**
		ASE	-	**36.4**	23.6	**40.0**
	French	BE	13.3	14.2	16.4	**56.0**
		BF	**29.5**	12.5	13.8	**44.2**
		MSE	17.2	10.1	10.1	**62.7**
		MF	11.5	8.7	17.4	**62.5**
		ASE	6.7	20.0	13.3	**60.0**

Table 6.2: Results of the perception test for individual speakers in a confusion matrix. The level of identification is calculated taking the responses of French and Scottish listeners together. The list of encoded emotions by children is in the first column of the table. Perceived emotions are listed in the row at the top of the table (PA - Perceived Anger, PF - Perceived Fear, PS - Perceived Sadness, PH - Perceived Happiness). The identification level is displayed in percentages, statistically significant values are in bold. MF1-MF6 - monolingual French children, MSE1-MSE6 - monolingual Scottish children, B1(SE or F)-B5(SE or F) - bilingual children (speaking in Scottish English or in French), ASE - Scottish actress.

Speaker	Encoded Anger				Encoded Fear				Encoded Sadness				Encoded Happiness			
	PA	PF	PS	PH	PA	PF	PS	PH	PA	PF	PS	PH	PA	PF	PS	PH
B1SE	8.4	14.7	**46.3**	**30.5**	6.3	**51.0**	12.5	**30.2**	-	**35.0**	**58.0**	7	1.0	23.0	20.0	**56.0**
B1F	**42.9**	2.0	**45.9**	9.2	10.0	**62.0**	27.0	1.0	1.0	25.0	**72.0**	2.0	10.0	10.0	31.0	**49.0**
B2SE	**98.0**	1.0	-	1.0	10.0	**73.8**	2.5	13.8	2.0	**59.0**	**39.0**	-	6.0	-	-	**94.0**
B2F	**97.0**	1.0	1.0	1.0	3.0	**70.0**	1.0	26.0	-	**37.0**	**62.0**	1.0	16.0	24.0	21.0	**39.0**
B3SE	**98.0**	1.0	-	1.0	4.0	**43.0**	**26.0**	**27.0**	8.0	24.0	**54.0**	14.0	11.0	1.0	1.0	**87.0**
B3F	**100.0**	-	-	-	3.0	**43.0**	**51.0**	3.0	-	**51.0**	**47.0**	2.0	24.2	3.0	2.0	**70.7**
B4SE	**37.0**	9.0	**40.0**	14.0	5.0	**45.0**	**43.0**	7.0	2.0	18.0	**78.0**	2.0	1.0	15.0	**29.0**	**55.0**
B4F	**61.0**	8.0	24.0	7.0	9.1	**35.4**	10.1	**45.5**	1.0	21.0	**77.0**	1.0	3.0	5.0	27.0	**65.0**
B5SE	**46.0**	**36.0**	11.0	7.0	7.0	**79.0**	6.0	8.0	-	30.0	**69.0**	1.0	**30.0**	18.0	23.0	**29.0**
B5F	**63.0**	18.0	5.0	14.0	**26.0**	**49.0**	12.0	13.0	-	**52.0**	**47.0**	1.0	**31.0**	15.0	5.0	**49.0**
MSE1	**32.0**	10.0	**48.0**	10.0	16.0	**35.0**	22.0	27.0	-	**40.0**	**58.0**	2.0	13.1	10.1	**44.4**	**32.3**
MSE2	23.0	17.0	**32.0**	28.0	2.5	13.8	8.8	**75.0**	-	**32.7**	**44.9**	22.4	3.8	1.3	6.3	**88.8**
MSE3	**48.0**	5.0	14.0	**33.0**	-	**88.0**	12.0	-	-	**51.0**	**48.0**	1.0	11.0	5.0	12.0	**72.0**
MSE4	**98.0**	1.0	-	1.0	15.0	**68.0**	3.0	14.0	-	**54.0**	**46.0**	-	11.7	-	-	**88.3**
MSE5	22.0	8.0	2.0	**68.0**	7.0	14.0	4.0	**75.0**	2.0	13.0	24.0	**61.0**	8.1	2.0	14.1	**75.8**
MSE6	11.0	16.0	**49.0**	24.0	5.0	23.0	28.0	**44.0**	11.0	15.0	**57.0**	17.0	10.0	22.0	**40.0**	28.0
MF1	**39.0**	5.0	28.0	28.0	10.0	**37.0**	12.0	**41.0**	5.0	26.0	**62.0**	7.0	11.8	12.6	**35.3**	**40.3**
MF2	**26.0**	3.0	**49.0**	22.0	11.0	7.0	**61.0**	21.0	3.0	12.0	**81.0**	4.0	8.1	3.0	**33.3**	**55.6**
MF3	**85.0**	1.0	9.0	5.0	**30.0**	**31.0**	10.0	29.0	3.8	18.8	**76.3**	1.3	15.0	10.0	11.0	**64.0**
MF4	**78.0**	3.0	18.0	1.0	7.0	27.0	**40**	26.0	3.0	8.0	**88.0**	1.0	4.0	10.0	**47.0**	**39.0**
MF5	**62**	4.0	23.0	11.0	10.1	13.1	**59.6**	17.2	5.0	13.8	**66.3**	15.3	13.3	3.3	**35.8**	**47.5**
MF6	**80.0**	7.0	-	13.0	4.0	22.0	13.0	**61.0**	3.4	**42.0**	26.1	28.6	4.0	2.0	3.0	**91.0**
ASE	**77.0**	11.0	9.0	3.0	-	**78.0**	22.0	-	-	36.0	**64.0**	-	3.0	**29.0**	19.0	**49.0**

Table 6.3: Results of the perception test for individual speakers in a confusion matrix. The level of identification is calculated taking the responses of French and Scottish listeners together. The list of encoded emotions by children is in the first column of the table. Perceived emotions are listed in the row at the top of the table (PA - Perceived Anger, PF - Perceived Fear, PS - Perceived Sadness, PH - Perceived Happiness). The identification level is displayed in percentages, statistically significant values are in bold. MF1-MF6 - monolingual French children, MSE1-MSE6 - monolingual Scottish children, B1(SE or F)-B5(SE or F) - bilingual children (speaking in Scottish English or in French), ASE - Scottish actress.

Listeners	Speaker	Encoded Anger				Encoded Fear				Encoded Sadness				Encoded Happiness			
		PA	PF	PS	PH	PA	PF	PS	PH	PA	PF	PS	PH	PA	PF	PS	PH
Scottish	B1SE	12	4	**48**	**36**	-	**51**	5.9	**43.1**	-	**36.4**	**58.2**	5.5	-	12.7	18.2	**69.1**
French	B1SE	4.4	26.7	**44.4**	**24.4**	13.3	**51.1**	20.0	**15.6**	-	**33.3**	**57.8**	8.9	2.2	35.6	22.2	**40**
Scottish	B1F	**35.2**	1.9	**59.3**	3.7	-	**87.3**	10.9	1.8	-	**30.9**	**65.5**	3.6	7.3	5.5	36.4	**50.9**
French	B1F	**52.3**	2.3	**29.5**	15.9	22.2	**31.1**	**46.7**	-	2.2	17.8	**80**	-	13.3	15.6	24.4	**46.7**
Scottish	B2SE	**98.2**	-	-	1.8	4.5	**75**	-	25.5	1.8	**74.5**	23.6	-	1.8	-	-	**98.2**
French	B2SE	**97.8**	2.2	-	-	16.7	**72.2**	5.6	5.6	2.2	**40**	**57.8**	-	11.1	-	-	**88.9**
Scottish	B2F	**94.5**	1.8	1.8	1.8	1.8	**69.1**	1.8	27.3	-	**59.9**	**47.3**	1.8	3.6	29.1	18.2	**49.1**
French	B2F	**100**	-	-	-	4.4	**71.1**	-	24.4	-	20	**80.0**	-	**31.1**	17.8	24.4	26.7
Scottish	B3SE	**96.4**	1.8	-	1.8	1.8	**30.9**	**38.2**	**29.1**	-	20	**74.5**	5.5	5.5	-	1.8	**92.7**
French	B3SE	**100**	-	-	-	6.7	**57.8**	11.1	24.4	17.8	**28.9**	**28.9**	24.4	17.8	2.2	-	**80**
Scottish	B3F	**100.0**	-	-	-	5.5	**32.7**	**61.8**	-	-	**60**	**40**	-	1.8	3.6	1.8	**92.7**
French	B3F	**100.0**	-	-	-	-	**55.6**	**37.8**	6.7	-	**40**	**55.6**	4.4	**52.3**	2.3	2.3	**43.2**
Scottish	B4SE	**43.7**	3.6	**34.5**	14.5	7.3	**47.3**	**40**	5.5	-	18.2	**81.8**	-	1.8	9.1	25.5	**63.6**
French	B4SE	24.4	15.6	**46.7**	13.3	2.2	**42.2**	**46.7**	8.9	4.4	17.8	**73.3**	4.4	-	22.2	**33.3**	**44.4**
Scottish	B4F	**67.3**	5.5	21.8	5.5	7.4	**27.8**	9.3	**55.6**	1.8	25.5	**72.7**	-	1.8	1.8	**34.5**	**61.8**
French	B4F	**53.3**	11.1	26.7	8.9	11.1	**44.4**	11.1	**33.3**	-	15.6	**82.2**	2.2	4.4	8.9	17.8	**68.9**
Scottish	B5SE	**52.7**	**32.7**	9.1	10.9	7.0	**72.7**	3.6	12.7	-	**41.8**	**58.2**	-	25.5	23.6	20	**30.9**
French	B5SE	**37.8**	**40**	13.3	8.9	2.2	**86.7**	8.9	2.2	-	15.6	**82.2**	2.2	**35.6**	11.1	**26.7**	**26.7**
Scottish	B5F	**61.8**	20	1.8	16.4	16.4	**54.5**	9.1	20	-	**74.5**	**25.5**	-	18.2	12.7	9.1	**60**
French	B5F	**64.4**	15.6	8.9	11.1	**37.8**	**42.2**	15.6	4.4	-	24.4	**73.3**	2.2	**46.7**	17.8	-	**35.6**

Chapter 7

Production results

This chapter presents the analysis of production data for monolingual and bilingual children in Scottish English and French. The corpus of affective speech validated by the perception test was analysed according to the methodology described in Chapter 5. This corpus contains eighteen speakers: five Scottish English-French bilingual children (recorded in the two languages), six French monolingual children, six Scottish monolingual children and one Scottish actress. Each studied affective state (anger, sadness, happiness, fear + neutral) of an individual speaker in a given language is represented by eight utterances. Thus, the total of 920 utterances was acoustically analysed. Acoustic correlates of pitch range variation, rhythm and speech rate were measured. Auditory analysis of voice quality was also performed. The results of the acoustic analysis for pitch range, rhythm and speech rate are reported in tables at Appendix B as mean and standard deviations, calculated over 8 utterances for each affective state of individual speakers.

This chapter also reports the statistical tests performed on the production data and the joint analysis of production and perception results in order to address the research questions of the dissertation. The first research question is whether acoustic correlates of affective speech can be identified in child production. Pitch range, rhythm and speech rate, associated with emotions in adult speech by previous research studies, were measured in the corpus of child affective speech. Statistical tests were performed to see if these speech parameters were used by children to differentiate between the studied affective states (anger, sadness, happiness, fear + neutral). The second research question is whether there are cross-linguistic differences in the usage of acoustic correlates of affective speech by French and Scottish monolingual children. To address this question, production results must be compared between

the two groups of children: French and Scottish monolingual children. The third research question is how bilingual children produce affective speech in their two languages and how their affective speech compares with that of monolingual children. It is therefore necessary to compare acoustic correlates used by bilingual children in their two languages, and to compare affective productions between bilingual and monolingual children.

7.1   Analysis of production data for groups of children

In order to address the three research questions, a doubly multivariate repeated measures model was chosen for the comparative analysis of acoustic correlates. This statistical model was applied to four subgroups of speakers: Scottish monolinguals (MSE), French monolinguals (MF), bilinguals in French (BF) and bilinguals in Scottish English (BSE). This model allows testing to see if analysed acoustic correlates are associated with affective states and if there are differences between groups of children. As each analysed acoustic parameter was measured by 8 utterances for each of the five speakers' affective states (anger, fear, sadness, happiness + neutral), its corresponding statistical variable had 8 levels for a repeated measures factor (also called a within-subjects factor). Two between-groups factors were chosen for the analysis of the data: emotion (five levels - anger, fear, sadness, happiness, neutral) and speakers' subgroup (four levels - MSE, MF, BF and BSE). As the repeated measures factor had 8 levels, the nature of the design was multivariate, so multivariate test statistics were applied. The statistical test was performed with SPSS statistical software.

The statistical test was performed separately for two types of acoustic correlates: first, for pitch measures, second, for rhythm and temporal measures. Pitch measures may be considered directly comparable in the two studied languages, thus the four subgroups of children (MSE, MF, BF and BSE) may be compared with each other. In contrast, rhythm and temporal measures cannot be compared in French and Scottish English in the recorded corpus, as it consists of affective expressions played on

Table 7.1: Multivariate tests (Pillai's Trace) for between subject effects on the model applied to four groups: monolingual French children, monolingual Scottish English children, bilingual children in French and bilingual children in Scottish English. Dependant variables are *span* and *level*. Value is significant at $p < 0.05$.

Effect	Value	F	Hypothesis df	Error df	Sig.
Group	0.870	3.372	24.0	198.0	**0.0001**
Emotion	1.377	4.399	32.0	268.0	**0.0001**
Interaction	0.948	0.795	96.0	568.0	0.918

only one token utterance, which has close but not identical phonetic structure in the two languages. In the context of one short utterance, a specific segmental environment may influence rhythm and speech rates measures, and thus the comparison of rhythm and speech rate between Scottish English and French would not be reliable. The results for the rhythm and speech rate may therefore only be compared between speakers of the same language, for example, between French monolingual children and French bilingual children.

7.1.1 Pitch measures

A doubly multivariate repeated measures model was used to compare two pitch parameters: pitch span and pitch level, used by four groups of speakers (6 monolingual French children (MF), 6 monolingual Scottish children (MSE), 5 bilingual children in French (BF) and 5 bilingual children in Scottish English (BSE)). Two dependent variables were analysed: *span* and *level*. Between-subject factors were GROUP and EMOTION. The factor GROUP has four levels: MF, MSE, BF and BSE. The factor EMOTION has five levels: anger, fear, sadness, happiness and neutral. The within-subject factor UTTERANCE has 8 levels, i.e. 8 repetitions of the utterance for each affective state. The analysis was run on type IV sums of squares.

Table 7.1 displays tests of significance for between-subject effects on the model. As Pillai's trace test is considered more robust than other statistics to violations of model assumptions (Olson, 1974), its results are shown. Pillai's trace is a positive-valued statistic. Increasing values of the statistic indicate effects that contribute more to the model. Both GROUP and EMOTION contribute significantly to the

model, but not the interaction between them. Table A.1 of Appendix A shows tests of between subject effects for the two measures individually. Only the measure of *span* does not show significance for the factor GROUP. The two measures, *span* and *level*, show statistical significance for factor EMOTION with $p < 0.05$. Thus, the statistical test showed that pitch range measures, both pitch level and pitch span, are used by children to differentiate their emotions; moreover, there are group differences in the usage of *level*. Bonferroni post hoc tests were run to study pitch range differences between particular emotions in the four groups.

Pitch span

The results for pitch span are shown in Figure 7.1. Bilingual children (both in French and English) and Scottish monolingual children have a wider span than French monolingual children. Only happiness is differentiated by bilingual children: it is realised with a narrower span in French than in English. The differences between the four subgroups of children do not attain the level of significance. Post hoc Bonferroni tests for span show that all children realise happiness and anger with a wider span; Scottish monolingual children and bilingual children in French also realise fear with a wide span. The other affective states have a lower span.

Pitch span results for groups of children show that French monolingual children have a narrower span than both Scottish monolingual children and bilingual children in English and French. Bilinguals differentiate only happiness in their languages: French happiness has a narrower span, similar to that of monolingual children. These results are presented schematically in Figure 7.2.

The differentiation of emotions by span follows the same pattern in Scottish English and French, see Figure 7.3. Happiness and anger are realised with a wide span. Sadness and neutral are realised with a narrow span. It is not possible to characterise fear by a particular span usage, as apparently it may be realised both with a narrow span and a wide span. Bilinguals, as a group, tend to realise fear with a narrow span in English and a wide span in French. In contrast, monolinguals
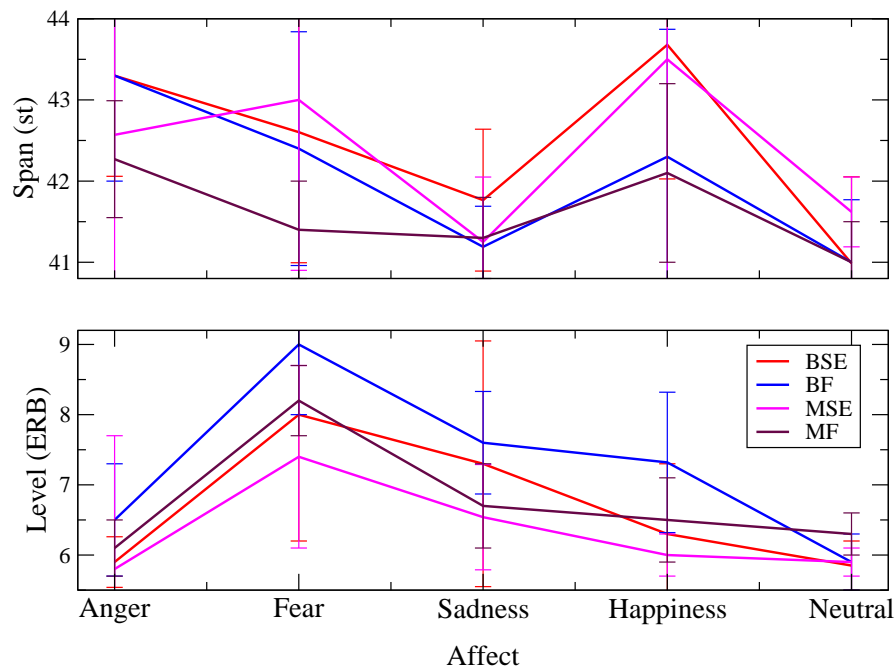
Figure 7.1: Pitch range results (span and level) for four groups of children: bilingual children in English, bilingual children in French, Scottish monolinguals and French monolinguals.
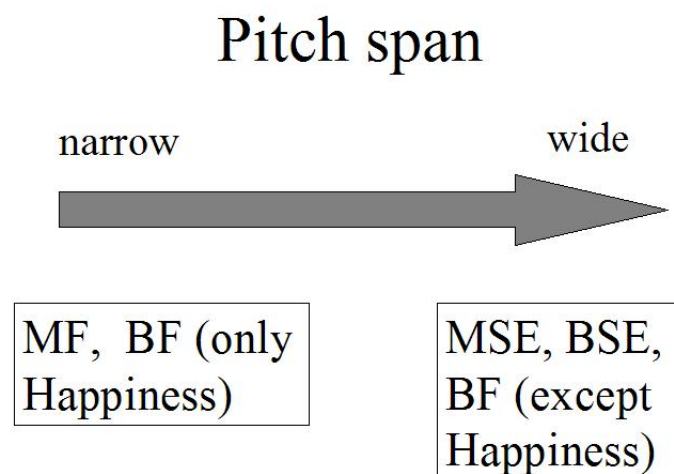
Figure 7.2: Pitch span results are presented schematically for four groups of children: bilingual children in Scottish English (BSE), bilingual children in French (BF), Scottish monolinguals (ME) and French monolinguals (MF).

Differentiation
of emotions by span

narrow   wide

Neutral
Sadness
Fear (MF, BSE)

Happiness
Anger
Fear (MSE, BF)
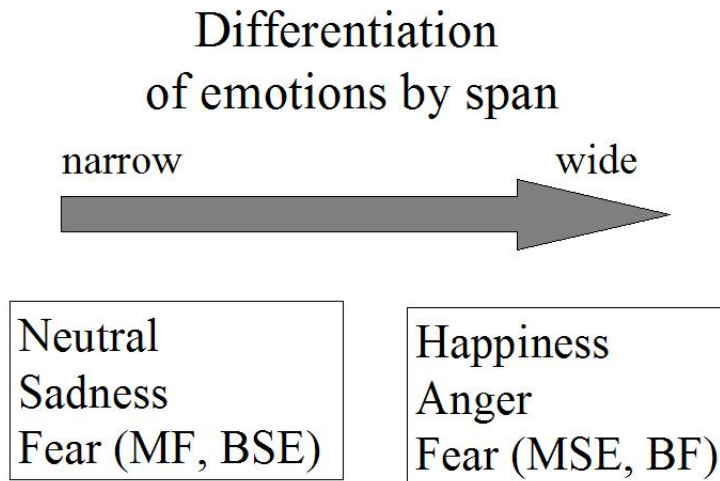
Figure 7.3: Differentiation of emotions by pitch level is presented schematically for four groups of children: bilingual children in Scottish English (BSE), bilingual children in French (BF), Scottish monolinguals (ME) and French monolinguals (MF).

tend to realise fear with a narrow span in French and a wide span in English. It is important to see how much variability exists in the realisation of fear by pitch span in the results of individual speakers.

Pitch level

The results for pitch level are shown in Figure 7.1. Level results show that bilingual children in French have the highest level, Scottish monolingual children have the lowest level, and the difference between these two groups is significant: ($p <0.001$). The other two groups are in between and overlap.

Bilingual children in English and Scottish monolingual children have fear with the highest level, then sadness and, finally, anger, happiness and neutral have a similar low level. Bilingual children in French and French monolingual children show a slightly different realisation pattern: they realise happiness with a higher level, similar to that of sadness.

Pitch level results show that French monolingual children have a higher level than Scottish monolingual children. Bilingual children follow the same pattern as monolingual children, as they differentiate their languages by using a higher level in

Pitch level

low   high

MSE < (BSE < MF) < BF
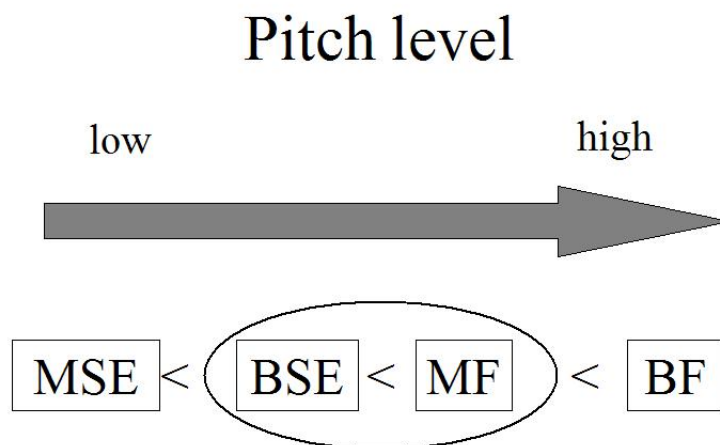
Figure 7.4: Pitch level results are presented schematically for four groups of children: bilingual children in Scottish English (BSE), bilingual children in French (BF), Scottish monolinguals (ME) and French monolinguals (MF).

Differentiation of emotions by level

low   high

Anger Neutral Happiness -MSE,BSE	Sadness Happiness -MF,BF	Fear
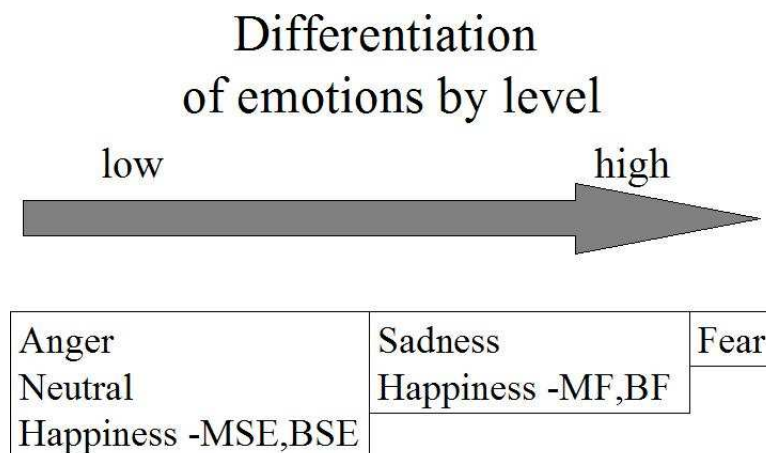
Figure 7.5: Differentiation of emotions by pitch level is presented schematically for four groups of children: bilingual children in Scottish English (BSE), bilingual children in French (BF), Scottish monolinguals (ME) and French monolinguals (MF).

French than in English. Realisations of bilingual children are higher, however, than those of monolingual children in the corresponding languages. Thus, the English pitch level of bilinguals is similar to monolingual French, and their French pitch level is significantly higher than that of French monolinguals. These results are presented schematically in Figure 7.4.

The pattern of differentiation of emotions by level is similar for English and French with the one exception of happiness, see Figure 7.5. Fear has the highest level. It is followed by sadness and, finally, by anger and neutral. The difference between languages is that French happiness is differentiated from anger and neutral by a higher level, similar to that of sadness. In contrast, English happiness has a low level, and it is not statistically different from anger and neutral. This pattern of happiness realisation found in the monolingual children is followed by bilingual children in English and French.

7.1.2 Rhythm and temporal measures

A doubly multivariate repeated measures model was used to compare the following parameters: speech rate, rhythm and final lengthening. As it was explained earlier (see also Chapter 5: Methodology, section 'Rhythm'), these parameters cannot be directly compared across languages in the studied corpus. Thus, monolinguals and bilinguals are compared only in the same language. The first test studies the stated parameters in the production of monolingual French children and bilingual children in French, the second those in the production of monolingual Scottish children and bilingual children in English. Four dependent variables were analysed: $rate$, PVI, $rINT$, $syllable$. Between-subject factors were GROUP and EMOTION. The factor GROUP has in this case two levels: monolinguals and bilinguals. The factor EMOTION has five levels: anger, fear, sadness, happiness and neutral, and the within-subject factor UTTERANCE has 8 levels. The analysis was run on type IV sums of squares. The tables 7.2 and 7.3 display tests of significance for between-subject effects on the model. Only the factor GROUP contributes significantly to

Table 7.2: Multivariate tests (Pillai's Trace) for between subject effects on the model applied to the two groups: French monolinguals and bilingual children in French. Value is significant at $p < 0.05$.

Effect	Value	F	Hypothesis df	Error df	Sig.	Observed Power
Group	0.546	11.746	4.000	39.000	**0.0001**	1.0
Emotion	0.503	1.512	16.0	168.0	0.101	0.871
Interaction	0.263	0.740	16.0	168.0	0.750	0.492

Table 7.3: Multivariate tests (Pillai's Trace) for between subject effects on the model applied to MSE and BSE groups. Value is significant at $p < 0.05$.

Effect	Value	F	Hypothesis df	Error df	Sig.	Observed Power
Group	0.240	3.085	4.0	39.0	**0.027**	0.759
Emotion	0.597	1.843	16.0	168.0	**0.029**	0.940
Interaction	0.197	0.544	16.0	168.0	0.920	0.355

the model for MF and BF groups. Both GROUP and EMOTION are significant for the model with MSE and BSE groups. Interaction is not significant for the two tests.

Table A.2 of Appendix A shows tests of between-subject effects for each individual measure calculated on the model for bilingual children in French and French monolingual children. Measures of $rINT$ and *syllable* show significance for the factor GROUP. Only the measure $rINT$ shows statistical significance for the factor EMOTION. No significant effect of interaction is observed. Bonferroni post hoc tests were run to study differences between particular emotions in the two groups. The realisation of the final syllable, Figure 7.6, shows significant GROUP difference between MF and BF, as bilingual children have a longer final syllable *syllable* in their French than French monolinguals. Bilinguals also show significantly higher consonantal variability than French monolinguals, Figure 7.8. Moreover, bilinguals differentiate their emotions by $rINT$.

Table A.3 of Appendix A shows tests of between-subject effects for each in-

Table 7.4: Results for temporal parameters: monolingual vs. bilingual children.

Measure	English	French
rINT	MSE - lower value than BSE	MF - lower value than BF
Rate	MSE - higher rate than BSE	MF = BF
syllable	MSE = BSE	MF - lower value than BF
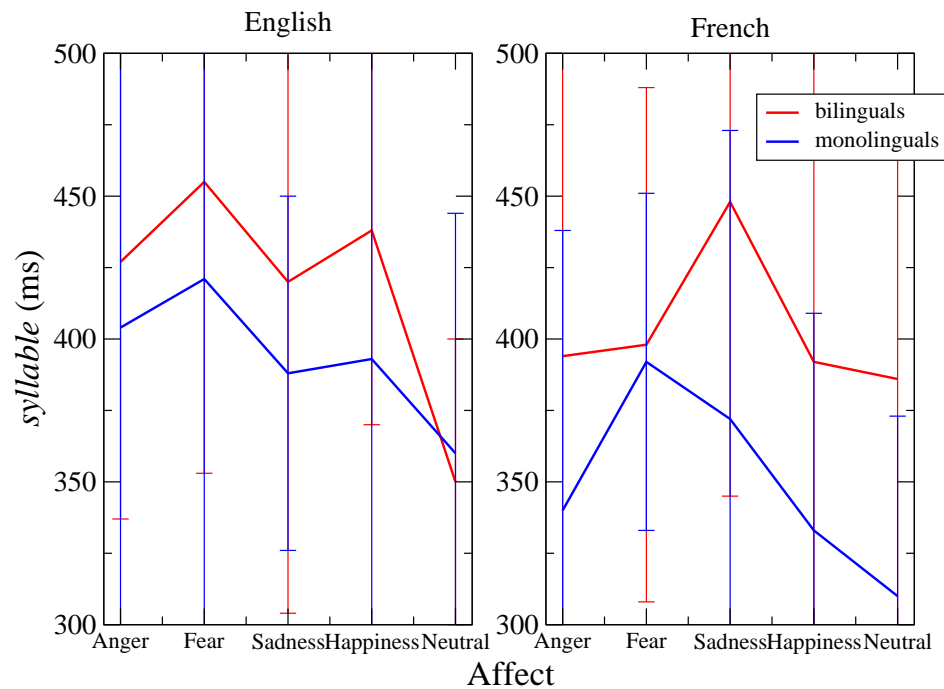
Figure 7.6: Results for the duration of the final syllable are presented for the two languages: French and Scottish English.

dividual measure calculated on the model for BSE and MSE children. Measures of $rINT$ and *rate* show significance for the factor GROUP. No statistical significance is observed for the factor EMOTION and for the interaction between the two factors. Bonferroni post hoc tests were run to study differences between particular emotions in the two groups. The realisation of speech rate (Figure 7.7) shows significant GROUP difference between MSE and BSE, as bilingual children have a slower speech rate in their English than Scottish monolinguals. Happiness and neutral have the highest speech rate of affective states, realised by bilingual children in English. Bilinguals also show significantly higher consonantal variability than Scottish monolinguals, Figure 7.8. Bilingual children realise sadness and anger with the highest $rINT$. Scottish monolingual children differentiate only anger from the other affective states, also with the highest consonantal variability.

7.1.3 Summary of group results

Group results based on the doubly multivariate repeated measures model show that pitch range measures (span and level) are used by children to differentiate emo-
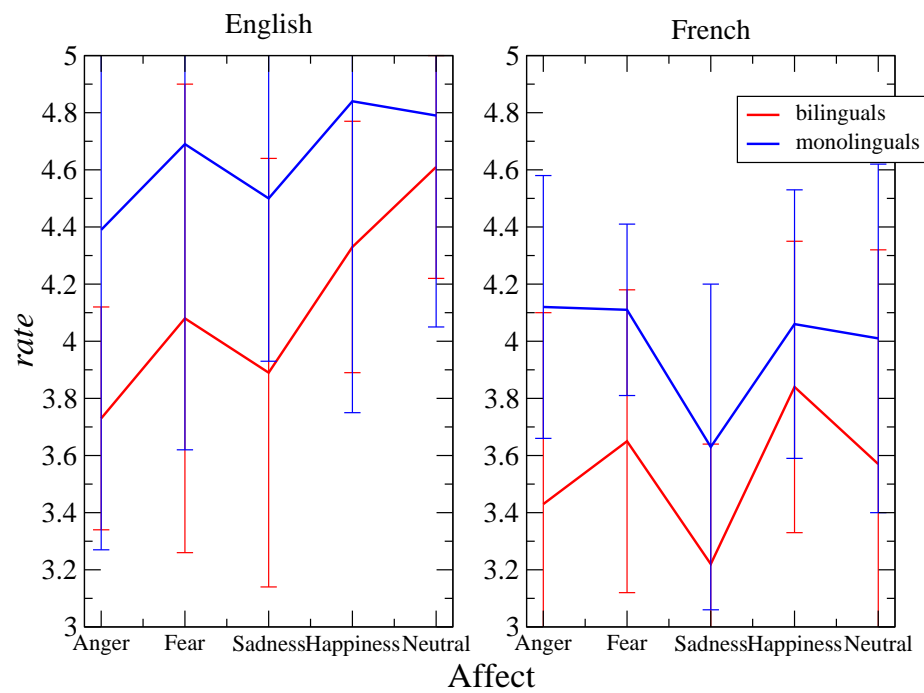
Figure 7.7: Results for speech rate are presented for the two languages: French and Scottish English.
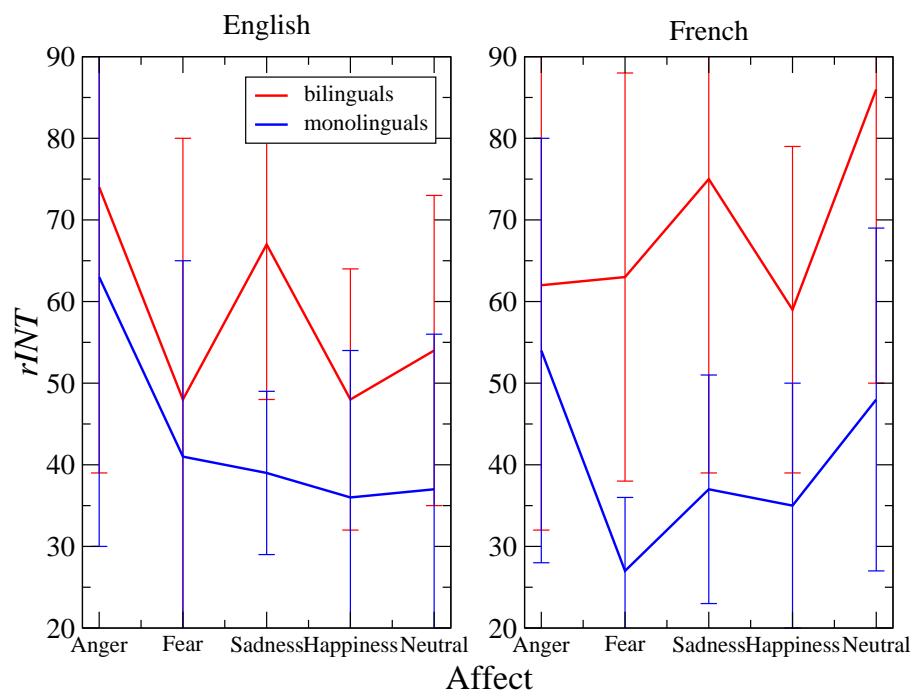
Figure 7.8: Results for $rINT$ are presented for the two languages: French and Scottish English.

tions, and that there are differences in the usage of pitch range measures by the four subgroups of children. The results showed that pitch span was narrower in French than in Scottish English, and that pitch level was higher in French than in Scottish English. At the same time, differences in the differentiation of emotions by level were significant between Scottish and French monolingual children, while the differentiation of emotions by span was realised in the same way.

The differentiation of emotions by span has the same pattern in Scottish English and French: happiness and anger are realised with a wide span; sadness and neutral are realised with a narrow span, fear may be realised with a narrow or a wide span. The differentiation of emotions by level is similar in the two languages for anger, fear, sadness and neutral: fear has the highest level, it is followed by sadness, and anger and neutral have the lowest level. The realisation of happiness is different in Scottish English and French. In French, happiness is realised with a higher pitch level, similar to that of sadness. In English, happiness has a lower level: it is not statistically different from anger and neutral.

Four acoustic correlates of rhythm and speech rate were statistically analysed: speech rate - $rate$, vocalic variability - PVI, consonantal variability - $rINT$, final syllable lengthening - $syllable$. These measures were analysed on their association with affective states: they were also compared between monolingual and bilingual children of the same language, but not across the languages, as was explained in Section 7.1.3. The results show that only one measure - consonantal variability, or $rINT$ - was consistently associated with affective states expressed by children in Scottish English and French. Group differences were found for $rINT$ and $syllable$ measures between monolingual and bilingual children in French: bilinguals had longer final lengthening and higher consonantal variability than French monolinguals. In Scottish English, group differences were found for $rINT$ and $rate$ measures between monolinguals and bilinguals: bilinguals had slower speech rate and higher consonantal variability than Scottish monolinguals.

The differentiation of emotions by consonantal variability follows the same pat-

tern across the two languages: anger is realised with a higher $rINT$ than the other affective states. Moreover, bilingual children in Scottish English also differentiate their affective states by speech rate: happiness and neutral have higher speech rate than the other emotions.

The presented production results for child affective speech in Scottish English and French allows us to address the first research question: can acoustic correlates of affective speech be identified in child production? Pitch range, rhythm and speech rate were measured in the corpus of child affective speech and statistically analysed. It was shown that the two measures of pitch range (span and level), one measure of rhythm (consonantal variability) and to some extent speech rate were used by children for the differentiation of vocal emotions. Thus, acoustic correlates of affective speech were identified in child productions. Note that out of three measured acoustic correlates of speech rhythm (PVI, $rINT$, $syllable$), only $rINT$ was associated with affective states. The perception test, reported in Chapter 6, demonstrated that adults were able to identify affective states encoded by children. The analysis of production results showed that the studied acoustic measures of pitch range, rhythm and speech rate are used by child speakers in the vocal encoding of affective states.

The second research question stands as to whether there are cross-linguistic differences in the usage of acoustic correlates of affective speech by French and Scottish monolingual children. Through the doubly multivariate repeated measures model, production results were compared between the studied groups of children. It was found that French and Scottish monolingual children differentiate their emotions by span in the same way: happiness and anger have a wide span; sadness and neutral have a low span; fear may have a wide or narrow span. For pitch span, it was found that anger, fear, sadness and neutral are differentiated in French and Scottish English according to the pattern: fear has the highest level, it is followed by sadness and finally anger and neutral have the lowest level. In contrast, a cross-linguistic difference was found in the realisation of happiness by level. French monolingual children realise happiness with a higher pitch level, similar to sadness. Scottish

monolingual children realise happiness with a lower level, similar to anger and neutral. Rhythm and speech rate measures could not be compared cross-linguistically in the two languages because of the specificity of the recorded corpus. Nevertheless, the pattern of differentiation of affective states by consonantal variability was the same in the two languages: anger had the highest $rINT$ both in Scottish English and French.

The third question asks: how do bilingual children produce affective speech in their two languages and how does their affective speech correspond with that of monolingual children? Thus, production results of the four subgroups of children (French monolinguals, Scottish monolinguals, bilinguals in French and bilinguals in Scottish English) were statistically compared. The statistical analysis showed that bilingual children differentiate their affective states by the measured acoustic correlates, and their pattern of differentiation of affective states corresponds with that of monolingual children. Nevertheless, some differences in the affective speech production were observed between monolingual and bilingual children.

- Production results showed that Scottish monolinguals had wider pitch span than French monolinguals. Bilingual children realised their pitch span like Scottish monolinguals in their two languages: they had a wide pitch span in both English and French. The only exception concerns happiness: this affective state was realised with a wider pitch span in Scottish English and a narrow pitch span in French, like monolingual speakers in the corresponding languages. The differentiation of emotions by span was performed according to the realisation pattern used by monolingual children.

- Scottish monolinguals had a lower pitch level than French monolinguals. Bilingual children also used lower pitch level in their English than in French. The difference between monolingual and bilingual children was that bilinguals used higher pitch levels than monolinguals in the corresponding languages. The differentiation of emotions by level was performed according to the realisation pattern used by monolingual children.

- Bilinguals have higher $rINT$ values than monolinguals in the corresponding languages. Bilinguals have a slower speech rate than Scottish monolinguals and a longer final syllable than French monolinguals.

The statistical analysis of production results for groups of children gave the first response to the research questions of the dissertation. Particularly, it allowed us to answer the first research question as to whether it is possible to identify acoustic correlates in child affective speech. A number of measured acoustic correlates were associated with emotions enacted by children. The results also showed some cross-linguistic differences in the affective speech of monolingual and bilingual children. In order to address more profoundly the research questions concerning cross-linguistic differences in child affective speech and the realisation of affective speech by bilinguals, the analysis of individual results is required, as well as the analysis of the production results in correlation with the perception data.

7.2 Production data analysis for individual speakers

Production data analysis for individual speakers is motivated by the fact that affective speech production is characterised by high individual variability - see the discussion of this issue in Chapters 3 and 5. Moreover, bilingual children make a specific group of speakers: a bilingual acquisitional context presents much more variability than a monolingual one, and the level of homogeneity is always lower between bilinguals than between monolinguals (see Chapter 5 for more details). This explains the necessity to analyse the production results for individual speakers, and particularly for bilingual children. The statistical analysis of production data for individual bilingual children may give insight into how each individual bilingual child can separate the affective speech in her two languages and whether the interaction between the two languages exists in the affective speech. Thus, individual analysis of the production data may help to address the third research question of the dissertation: how bilingual children produce the affective speech in their two languages and how their affective speech corresponds to that of monolinguals.

Separate multivariate analysis of variance (MANOVA) was carried out for each individual speaker, for a total of eighteen speakers. Their results are displayed in tables A.4-A.6 of Appendix A. One MANOVA test was run for each monolingual speaker from the corpus. The test was based on one factor - EMOTION (five levels: anger, fear, sadness, happiness, and neutral). Six dependent variables were included: *span*, *level*, *syllable*, *rate*, *PVI*, *rINT*. Bonferroni post hoc tests were run to study the differences between particular emotions.

Three MANOVAs were performed for each bilingual child. The first MANOVA test was designed for pitch measures. It was based on two factors: EMOTION (five levels: anger, fear, sadness, happiness, and neutral) and LANGUAGE (two levels: English and French). Two dependent variables were considered: *span* and *level*. The next two MANOVAs were run for rhythm measures: one MANOVA for English, and one MANOVA for French of the bilingual child. Only one factor was considered in these tests: EMOTION (five levels: anger, fear, sadness, happiness, and neutral). Four dependent variables were analysed: *syllable*, *rate*, *PVI*, *rINT*. Bonferroni post hoc tests were run to study the differences between particular emotions. MANOVA results for individual speakers are presented in sections of corresponding parameters.

7.2.1 Pitch measures

Statistical results of individual MANOVAs are presented in Table A.4 of Appendix A. The factor EMOTION in the individual MANOVA tests is significant for pitch span and pitch level of all children, except Scottish monolingual child MSE1 who does not differentiate her emotions by *span*. Bilingual children were also tested for the factor LANGUAGE. For the dependent variable *span*, only child B5 did not show significant results for factors LANGUAGE and INTERACTION; thus she does not differentiate her languages by *span*. Factor LANGUAGE is not significant for the variable *level* of children B1 and B2, but factor INTERACTION is significant for the measure *level* of all bilingual children. Individual results are presented graphically

in Figures 7.9 - 7.16, and numerically in Appendix B. Post hoc Bonferroni tests were run to study the differences between particular emotions.

Bilingual child B1 has a significantly wider span for fear and anger in French than in English. Though a higher level is realised in French than in English by B1, it is not significant. Child B2 has a wider span in English than in French, except for anger, which is realised with a wider span in French. The realisation of level overlaps in the two languages but only anger tends to have a higher level in French. Child B3 realises all the affective states with a wider span in English than in French and it is significant for happiness. The level is higher in French than in English, and it is significant for anger, sadness and happiness. B4 tends to use a wider span in English, but it is significant only for happiness. The level is significantly higher in French for all the affective states. Child B5 realises a wider span in English, but it is not statistically significant. The French level is significantly higher for all the affective states except anger.

Results of Bonferroni tests (see Appendix B) show how speakers differentiate their affective states.It is possible to see that only one bilingual child, B5, has the same pattern of differentiation of affective states by span in the two languages: sadness and neutral have a similar narrow span, and they are significantly different from the other affective states. Other bilingual children show different patterns of realisation for their two languages. The pattern used by child B5 is also found in the realisations of children B1F and MF3. Children B2SE, MF4, MF5 separate only neutral from the affective states with the narrowest span. Two bilingual children B2F and B3F and the monolingual child MF1 separate only anger from the other affective states by the widest span. Three bilingual children B1SE, B3SE, B4SE and the monolingual child MSE4 separate happiness from the other emotions with the widest span. Thus, based on the individual results, a cross-linguistic difference was observed in the differentiation of affective states by span. Happiness tends to be differentiated significantly by the widest span by speakers in English. In contrast, anger tends to be differentiated significantly by the widest span by speakers in

French. Neutral and sadness are realised with a narrow span both in English and French across individual speakers. Fear realisation shows most variability as it is realised with both a wide span and a narrow span by individual speakers.

The differentiation of affective states by pitch level shows less individual variability than that of pitch span. Exactly the same differentiation patterns are observed in some French and Scottish speakers. Neutral and anger are realised with the lowest level. Happiness is generally realised with a low level similar to anger and neutral, but some speakers (B2SE, B2F, B3F, MF1) use a significantly higher level for happiness; except with child B2SE it occurs only in French. Sadness is realised with a high level, but lower than that of fear. Fear is realised with the highest level both in English and in French. It important to draw attention to the fact that a number of speakers (B5SE, B5F, MF6) used a high arousal version of sadness. These speakers were not included in the group analysis, but their individual results show that sadness (and not fear, as for other speakers) is realised with the highest level (see differentiation patterns of children B5SE, B5F, MF6).

7.2.2 Rhythm and temporal measures

Rhythm and temporal measures, *rate*, *PVI*, *rINT*, *syllable*, are reported for each individual speaker in tables A.5 and A.6 of Appendix A. Post hoc Bonferroni tests were run to study differences between emotions. These results are reported in the corresponding subsections: speech rate, vocal variability, consonantal variability, final syllable lengthening.

Speech rate results

Individual MANOVAs were significant for the majority of speakers, but monolingual children MF5 and MSE6 do not show significance for the factor EMOTION. The variable of speech rate is significant for the factor EMOTION for all bilingual children in their two languages. The detailed statistical results are shown in Table A.5 of Appendix A. The results for all individual speakers are displayed in Figure
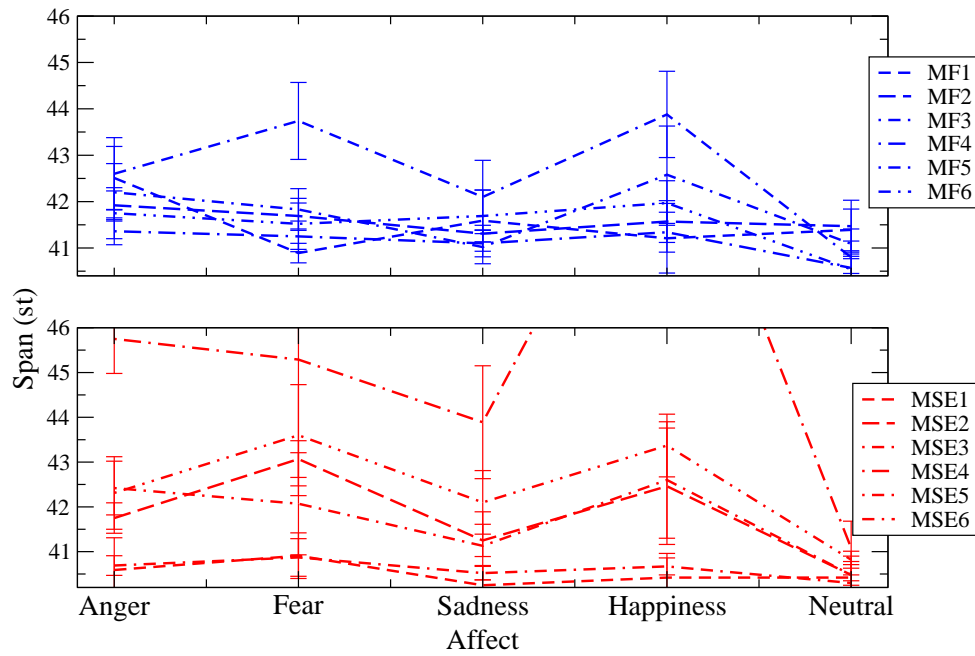
Figure 7.9: Individual results of *span* for Scottish English and French monolingual children.
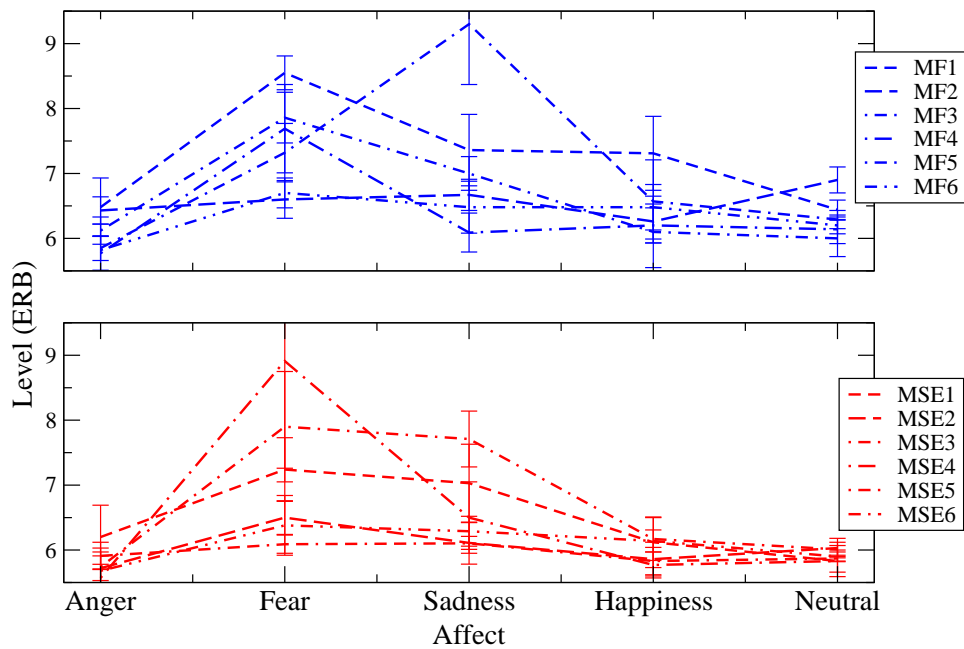
Figure 7.10: Individual results of *level* for Scottish English and French monolingual children.
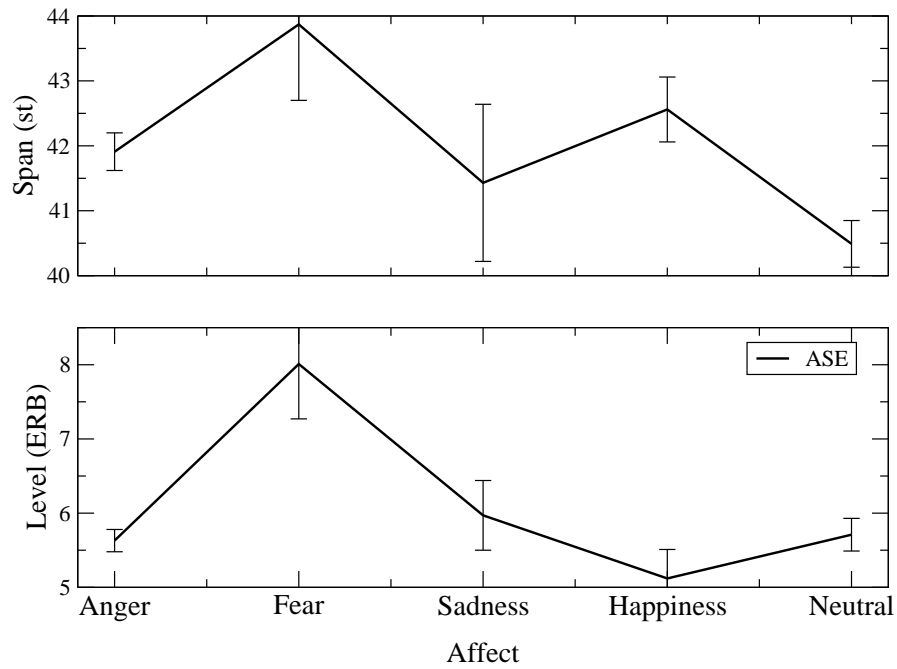
Figure 7.11: Pitch range for Scottish actress (ASE).
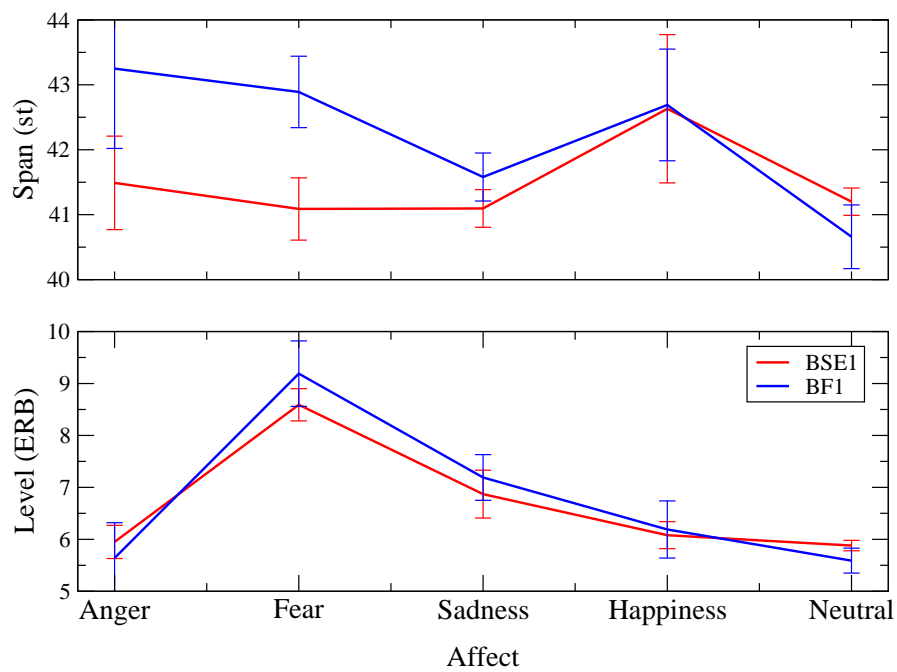
Figure 7.12: Pitch range for bilingual child B1 in Scottish English and French with standard deviations.
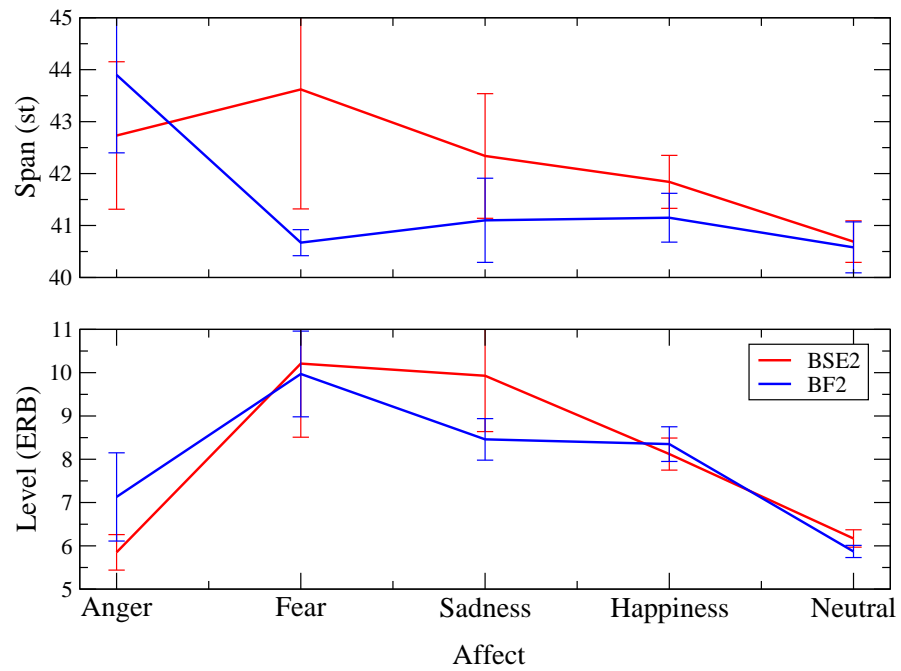
Figure 7.13: Pitch range for bilingual child B2 in Scottish English and French with standard deviations.
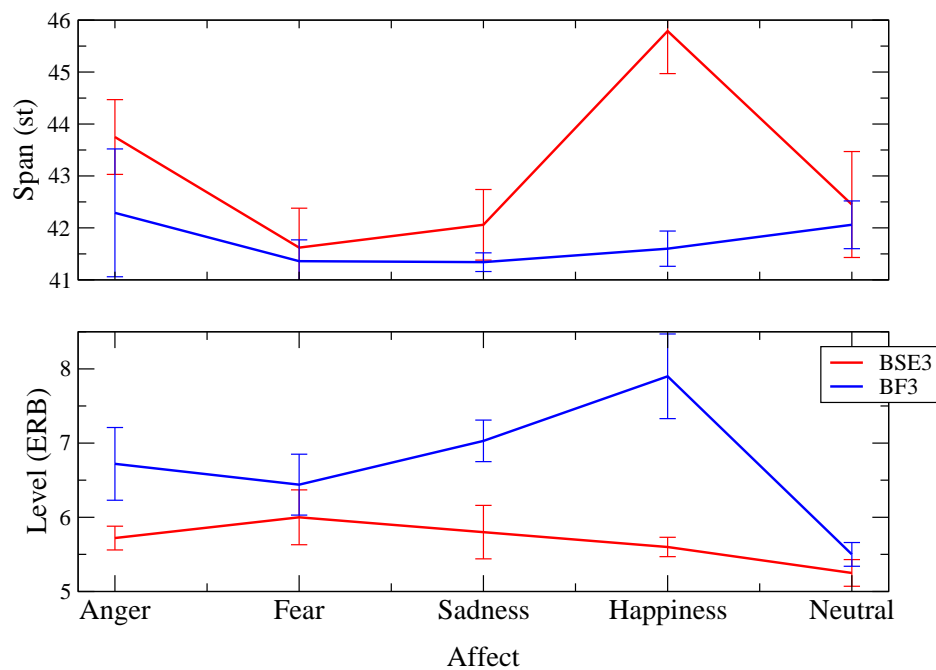
Figure 7.14: Pitch range for bilingual child B3 in Scottish English and French with standard deviations.
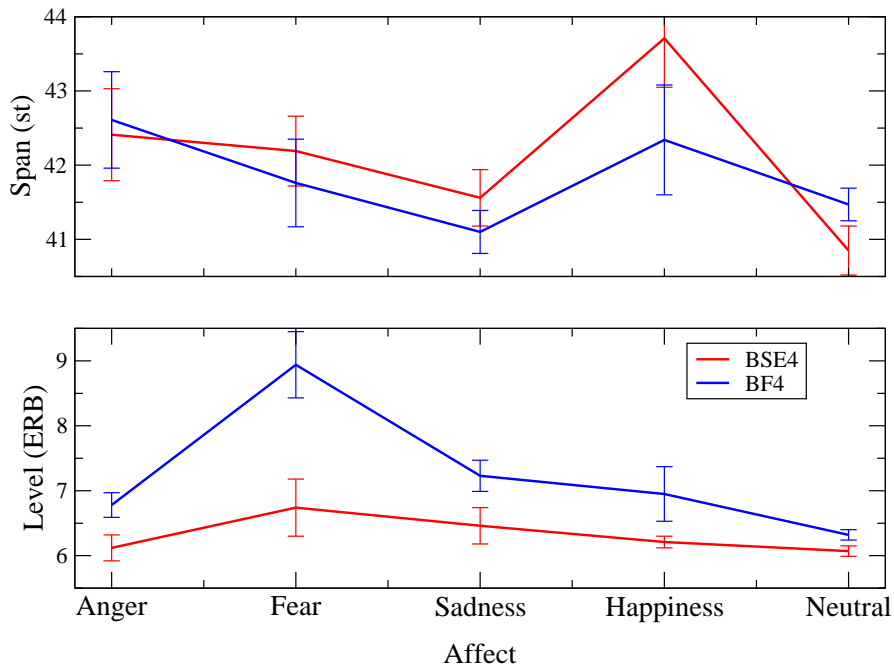
Figure 7.15: Pitch range for bilingual child B4 in Scottish English and French with standard deviations.
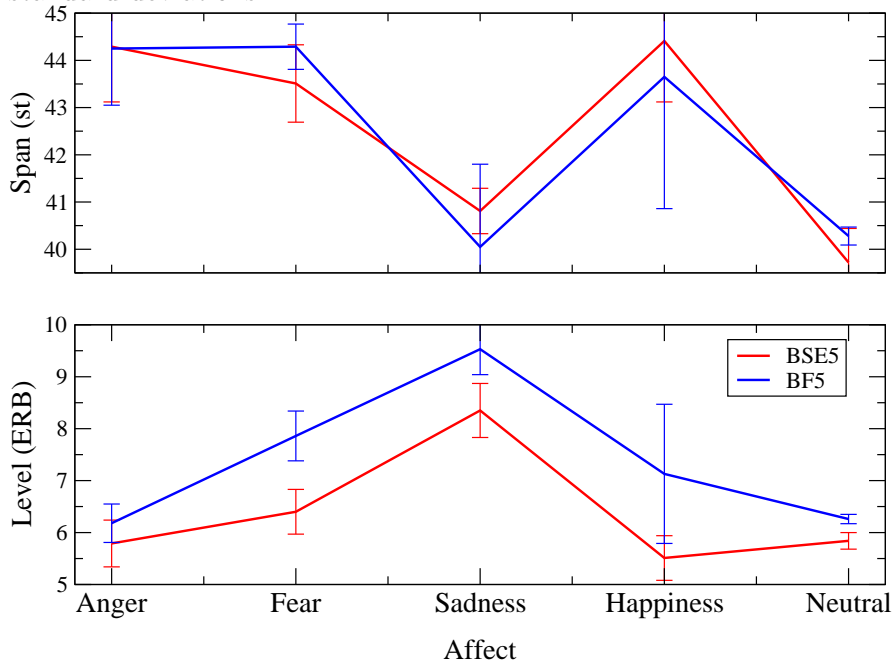
Figure 7.16: Pitch range for bilingual child B5 in Scottish English and French with standard deviations.
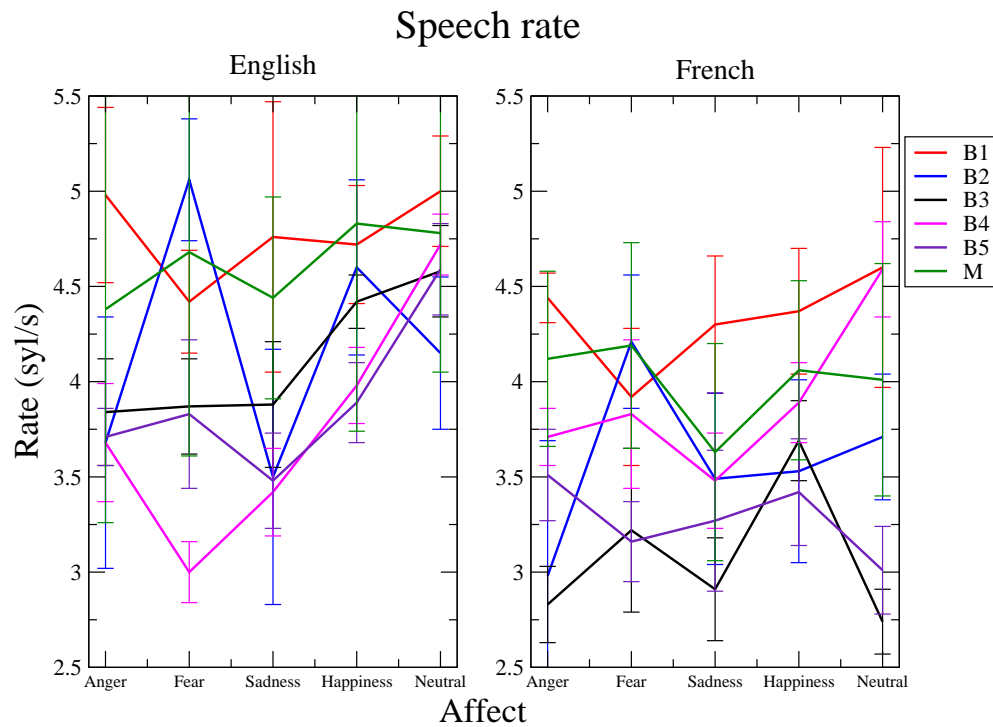
Figure 7.17: Results for speech rate of individual speakers in Scottish English and French.

7.17.

B1 does not differentiate her emotions by speech rate in English. In French, fear has the slowest rate and is significantly different only from neutral. In English, B2 has fear and happiness with the highest rate, anger and sadness with the lowest rate; these two groups are significantly different from each other. In French, fear has the highest rate, and sadness has the lowest; they are significantly different from each other, but not from the other emotions. B3 realises the same pattern in her two languages: happiness has the highest rate and it is significantly different from the other emotions with a lower rate. B4 also realises happiness with the highest rate in her two languages: sadness and fear have a lower rate. B5 also realises happiness with the highest rate, and fear and sadness with a lower rate in her two languages, but the difference between happiness and sadness is significant only in English. The Scottish actress realises anger with the lowest speech rate, and neutral with the highest.

The differentiation of affective states by speech rate looks similar in English and French for monolingual and bilingual children. Individual speakers realise happiness
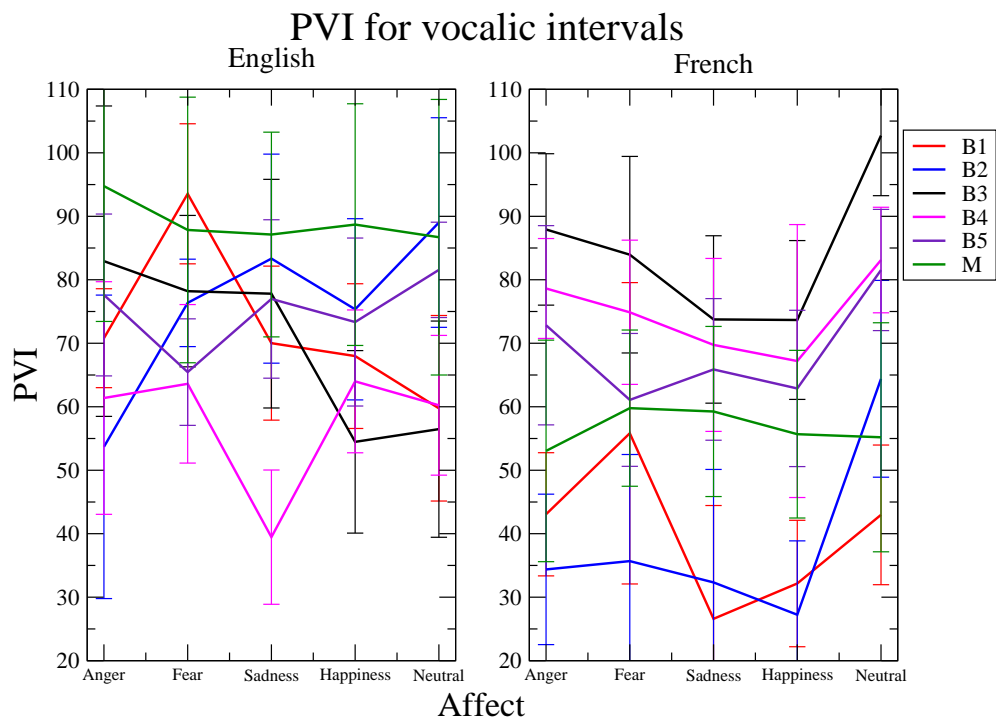
Figure 7.18: Results for PVI of individual speakers in Scottish English and French.

with a high speech rate, anger and sadness with a low speech rate. Fear realisation varies, as it may be realised both with a high and a low speech rate.

Vocalic PVI

Individual MANOVAs were only partially significant. Among French monolingual children, only one French monolingual (MF1) and three Scottish monolinguals (MSE3, MSE5, MSE6) show significance for the factor EMOTION. The Scottish actress (ASE) does not use it significantly to differentiate her emotions. The dependent variable PVI is significant for the factor EMOTION of bilingual children, except for children B3E, B4F and B5E. The detailed statistical results are shown in Table A.6, and a graphical representation is in Figure 7.18.

French monolingual child MF1 realises anger with the lowest vocalic variability, which overlaps only with sadness. MSE3 separates only neutral with the lowest variability, and emotions themselves are not differentiated. MSE5 has sadness with the highest PVI, significantly different only from fear. B5SE, B5F, B3F, B2F separate only neutral from the other emotions with the highest PVI. B3SE and
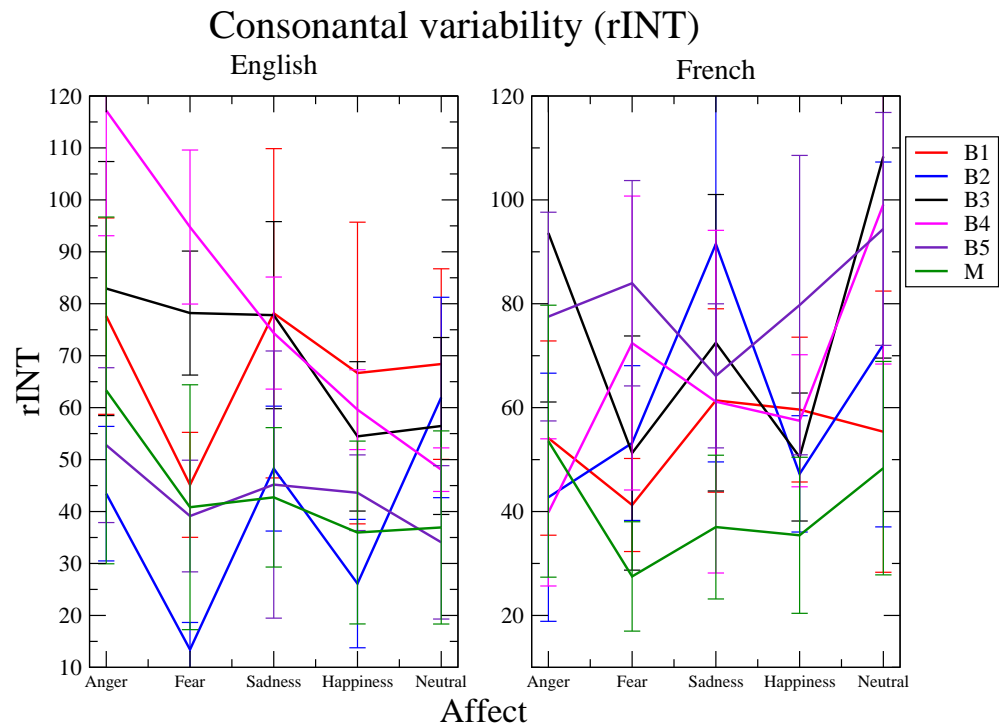
Figure 7.19: Results for $rINT$ of individual speakers in Scottish English and French.

B4F do not differentiate their emotions by vocalic PVI. B1 separates fear from the other emotions with the highest PVI: it does not overlap with the other emotions in English, it overlaps only with anger in French. B2SE realises anger with the lowest PVI, and it is significantly different only from sadness and neutral. B4SE realises sadness with the lowest PVI, and it is significant. ASE does not differentiate her emotions by PVI.

Consonantal $rINT$

Individual MANOVAs were significant for the majority of subjects. Among monolingual children, only MF2, MF6 and MSE6 do not show significance for the factor EMOTION. Bilinguals B1F and B5E do not use it significantly. The Scottish actress uses this acoustic correlate to differentiate her emotions. The detailed statistical results are shown in Table A.6. See also Figure 7.19 for the graphical display of individual data.

French monolingual MF1 separates anger with the highest $rINT$; MF2 does not differentiate her emotions by $rINT$, and MF3 has neutral and anger with the highest

$rINT$. MF4 has anger with the highest $rINT$, which differs significantly from fear and happiness. Scottish monolingual child MSE1 realises anger and sadness with the highest $rINT$, which overlap only with fear. MSE2 differentiates anger with the highest $rINT$ from the other emotions. MSE3 realises anger, sadness and neutral with a similar high $rINT$. MSE4 realises anger with the highest $rINT$, which overlaps only with sadness. MSE5 realises sadness and neutral with the highest $rINT$, which differ from happiness.

B1SE realises anger with the highest $rINT$, which is separated only from fear. B2 realises neutral, sadness and anger with a higher $rINT$, and happiness and fear have a lower $rINT$. In English, B3 has anger with the highest $rINT$, which overlaps only with fear and sadness: in French it is neutral, anger and sadness showing the highest $rINT$. B4 separates anger with highest $rINT$, followed by fear, which is also significantly different from the other emotions. In French, neutral has the highest $rINT$, and overlaps only with fear. BF5 realises neutral with the highest $rINT$, and overlaps with both anger and fear. ASE differentiates anger with the highest $rINT$ from the other emotions.

High consonantal variability is consistently used for anger by different individual speakers in English and French. Some speakers also use high $rINT$ for sadness and fear.

Final syllable

Individual MANOVAs were significant for the majority of subjects -see Table A.5 in Appendix A. Monolingual children MF2, MF5 and MSE6 do not show significance for the factor EMOTION. In bilingual children, B2 and B3F do not use final syllables significantly to differentiate their emotions.

In English, B1 realises fear and happiness with the longest final syllable: they are significantly different from the other emotions, but not from each other. In French, B1 realises fear and neutral with the longest final syllable, which overlaps with happiness, but not with the other emotions. B2 does not differentiate significantly

her emotions. B3 separates neutral from the emotions with the shortest syllable in English. In English, B4 differentiates fear and sadness with the longest syllable while happiness, anger and neutral have the shortest syllable: happiness overlaps also with sadness. In French, fear has the longest syllable, and it is significantly different only from happiness with the shortest final syllable. In English, B5 separates only neutral from the emotions with the shortest syllable. In French, sadness also has the longest syllable, which is significantly different only from anger, which has the shortest final syllable. The Scottish actress ASE has fear with the shortest final syllable: anger and sadness have the longest final syllable.

7.2.3   Summary of individual results

Group results for monolingual and bilingual children showed that span is wider in English than in French, and that level is higher in French than in English. Individual results for bilingual children follow group tendencies for pitch level and span. They also show variability: some children separate their languages for all the affective states, some do it only for some of the affective states and others do not separate their languages for the studied acoustic correlates.

Bilingual children B3 and B4 realise wider span for English than for French but it is significant only for happiness. Children B1 and B2 also use wider span in English, except for anger, which is used with a wider span in French. Child B5 uses wider span in English, but it is not significant. Pitch span is not wider for all the affective states of bilinguals in English. Bilingual children B1 and B2 use higher level in French than in English, but it is not significant. Children B3 and B5 use higher level in French, which is significant for some of the affective states. Child B4 uses higher level for all the affective states in French.

Individual results for pitch span support the group results, as happiness and anger tend to be realised with a wide span, while neutral and sadness have a narrow span. Fear realisation is variable: both wide and narrow span are observed. The specific pattern of differentiation between the studied affective states varies across

speakers, and this is true for monolinguals and bilinguals. Only the bilingual child B5 used the same realisation pattern in her two languages: sadness and neutral have a similar narrow span and they are significantly different from the other affective states. Interestingly, this pattern is also used by two other children in French (B1F and MF3).

The analysis of individual results allowed us to observe an interesting cross-linguistic difference in the differentiation of affective states by span. Though anger and happiness have a wide span in both English and French, as also shown by group result, there is a difference in their realisation. Happiness tends to be differentiated significantly by the widest span by speakers in English. In contrast, anger tends to be differentiated significantly by the widest span by speakers in French. Neutral and sadness are realised with a narrow span both in English and French across individual speakers. Fear realisation shows most variability, as it is realised with a wide span and a narrow span by individual speakers.

The differentiation of affective states by pitch level shows less individual variability than that of pitch span. Exactly the same differentiation patterns are observed in some French and Scottish speakers. Neutral and anger are realised with the lowest level. Happiness is generally realised with a low level similar to anger and neutral, but some speakers (B2SE, B2F, B3F, MF1) use a significantly higher level for happiness. Low arousal sadness is realised with a high level, but lower than that of fear. Fear is realised with the highest level both in English and in French. High arousal sadness, realised by some speakers (B5SE, B5F, MF6), differs from low arousal sadness in level, as it has a higher level than fear.

The differentiation of affective states by speech rate follows the same pattern in English and French for monolingual and bilingual children. Individual speakers realise happiness with a high speech rate, anger and sadness with a low speech rate. Fear realisation varies, as it may be realised both with a high and a low speech rate.

High consonantal variability is consistently used for anger by different individual speakers in English and French. Some speakers also use high $rINT$ for sadness and

fear. Vocal variability and final syllable are used by some speakers to differentiate their affective states, but their usage is less consistent across speakers.

The summarised results of individual monolingual and bilingual speakers support the findings from group analysis, and give a scope of possible variability both for monolingual and bilingual children. Different realisation patterns were observed for bilingual children, depending on the affective states and analysed measures. The following section serves to reunite individual and group production data with perception results from Chapter 6, as the effectiveness of individual strategies in encoding affective states may be judged only by their level of successful identification.

7.3 Joint analysis of production and perception data

7.3.1 Anger

The summary of production and perception results for anger is presented in Table 7.5. Perception results show that anger expressed by French monolingual children is identified with a higher level than anger expressed by Scottish monolingual children. The same observation is true for bilingual children: their French anger is identified with a higher level than their English anger. Individual results for bilinguals show that two children, B2 and B3, encoded anger extremely successfully: almost 100% of identification in their two languages. Their anger identification is even higher than that of the professional actress. Anger, expressed by children B4 and B5, was identified with a higher level in French than in English. For child B1, anger was successfully identified only in French. Anger expressions were identified with very high recognition level for four French monolingual children but only for two Scottish monolingual children.

Production results show that all bilingual and monolingual children use significantly wider pitch span than for neutral in French. Two bilingual children (B1SE and B5SE) use the same pitch span as for neutral in English: these are also children who have the lowest level of identification among bilingual children in English. Three monolingual Scottish children do not use wider span for their anger, and their

level of identification is also low. In French, pitch level is used by all French monolingual children and one bilingual child, B4F, significantly lower than in neutral. Two Scottish monolingual children and one bilingual child B2SE use lower level for their anger.

Many bilingual and monolingual children use significantly higher consonantal variability for their anger than in neutral. One bilingual child, B2, in both languages, and B5SE and ASE2 use lower speech rate for anger. The majority of speakers use also creakiness or harshness in their encoding of anger. Interestingly, the presence of creakiness does not guarantee high identification level of anger: in contrast, all the speakers using harshness are successfully identified.

Speakers who were successfully identified used wide pitch span, low pitch level, high consonantal variability, low speech rate, creakiness or harshness. This does not imply that each successful encoding uses all the listed acoustic correlates. Several strategies for use of these acoustic correlates were observed: the usage of all the mentioned correlates (e.g. B2SE - 98%), the usage of pitch range (span and level) without voice quality (MF4 - 78%), the usage of wide pitch span, high consonantal variability and creakiness (B3SE - 98%), the usage of wide pitch span, high consonantal variability and harshness (e.g. ASE2 - 77%). It is important to note that if the identified acoustic correlates of anger are used, they follow the same direction. For example, speakers used similar or significantly lower than neutral pitch level, but never significantly higher than neutral pitch level.

7.3.2 Fear

Production and perception results for fear are summarised in Table 7.5. Bilingual children have been successfully identified both in English and in French. Fear encoded by French monolingual children received very low level of identification. Two monolingual Scottish children (MSE3 and MSE4) were identified with a high level. Another interesting observation was made in the perception test: the Scottish actress's fear was identified by Scottish listeners at 98.2%, while French listeners

confused her fear with sadness (53.3% vs. 46.7%). Individual results for bilingual children show that some children have the same identification level in the two languages (B2, B3), while others were better identified in one of their languages. Bilingual child B5 was identified better in French, child B1 was identified better in English. It is not possible to say that all bilingual children expressed fear more in the same language, as was done for anger.

Production results show that almost all speakers use higher than neutral pitch level. Many, but not all, speakers use significantly wider than neutral pitch span. Consonantal variability is widely used in the two languages, but both high and low $rINT$ is possible. Speech rate and vocal variability are used to lesser extent, and no single tendency is observed. Final syllable is realised longer than in neutral by all the children who used it (B1, B4, B5, MSE4, MSE6, MF1, MF3), but the Scottish actress realised it shorter than in neutral. Two bilingual children (B4 and B5) used whisper, child B1 used creakiness, and several monolingual children (MSE2, MSE3, MSE4, MF3) used breathiness.

Different successful strategies were observed across individual speakers. For example, the Scottish actress (ASE2 - 78%) used wide span, high level, low consonantal variability, high speech rate and short final syllable. Some bilingual children (B1SE - 51%, B2F - 70%) did not use wide span, but they were still successfully identified. Apparently this variability in the usage of identified acoustic correlates is acceptable in the realisation of fear. The presence of the observed variability in the production of fear makes it more difficult to judge in the cross-linguistic perspective, and a higher number of speakers is required.

7.3.3 Sadness

Sadness recognition rates were significant across French and Scottish listeners for all groups of children. The main difference between French and Scottish listeners is that Scottish listeners also show significant confusion of sadness with fear for all the groups of children, except French monolingual children. Even sadness, encoded

by the Scottish actress, received a significant confusion with fear (sadness 47.3% vs. fear 52.7%). French listeners showed higher identification level for sadness across all groups of encoders. Two significant confusions with fear occurred only for bilingual children in English and for Scottish monolingual children. Interestingly, the Scottish actress has been successfully identified by French listeners at 84.4%.

Perception results for sadness show that bilingual children B1 and B2 were identified better in their French, and children B3 and B5 in their English. Sadness, expressed by child B4, was identified with a similar high rate in the two languages. Only one French monolingual child (MF6) was not successfully identified: other French children were decoded with the same or even higher level of identification than the professional actress. Scottish monolingual children received lower levels of identification.

Different strategies in the usage of acoustic correlates may be successfully decoded. All bilingual children use the same pitch span as in neutral and higher pitch level in their two languages. The usage of higher pitch span is also possible, but in this case with the same pitch level as in neutral (MF4 - 88% and MF5 - 66%). The Scottish actress (ASE - 64%) uses the same pitch range (both span and level) as in neutral, but in combination with slow speech rate, long final syllable and creakiness. Similar realisation strategy was found in French monolingual child MF2, who was successfully identified at 81%.

7.3.4 Happiness

Perception results show that bilinguals' happiness not only was recognised with higher accuracy in Scottish English than in French by Scottish listeners (70.9% vs. 62.9%), but also by French listeners (56.0% vs. 44.2%). Another interesting observation is that Scottish listeners showed higher levels for the identification of bilinguals, both in English and French, than the French listeners.

The realisation of happiness shows some differences in English and French. Almost all speakers in English used wider pitch span to encode happiness. Only one

Scottish child, MSE2, did not differentiate her happiness by the wide span and her identification was low - 32%. Bilingual child B2 is the only child in English who used significantly higher pitch level, and she also used lower consonantal variability, lower vocalic variability, higher rate and smiling; her identification was 94%. Other bilingual and monolingual children did not use pitch level in their realisation strategies in English. Child MSE2 used an opposite realisation strategy to B2SE: she realised happiness with a wide span and smiling, and it was identified at 89%. Smiling usage is spread across different speakers in English and French, but it is not obligatory for the successful recognition of happiness, as the strategies used by children B3 (78%), MSE3 (72%), MSE4 (88%), suggest. In French, three bilingual children (B2F - 39%, B3F - 47%, B5F - 49%) and two monolingual children (MF1 - 43 % and MF5 - 47%) realise higher pitch level than for neutral. Two French monolingual children, MF2 and MF3, realise their happiness with span and level as in neutral, but in combination with smiling their encoding was identified at 55% and 64%.

7.4 Summary

In this chapter, results based on the acoustic analysis of the recorded corpus were presented. Group comparisons were performed, together with the detailed analysis of individual speakers. Each speaker was analysed individually and compared with other speakers of the same language and with speakers of the other language. This analysis of individual speakers allowed us to observe the scope of existing variability in the production of affective speech among monolingual and bilingual children. Affective productions of each bilingual child were explored in the two languages and compared with other bilingual and monolingual children in order to see individual realisation patterns in the bilingual affective speech production. Finally, a joint analysis of individual production data and perception test data was performed in order to identify and compare successful strategies of affective speech production for monolingual and bilingual children in Scottish English and French. To sum-

Table 7.5: Summary of production and perception results for each individual speaker. If the acoustic correlate is significantly differentiated from the neutral value, arrows are used. Upper arrow means that the value is significantly higher than that of neutral, and down arrow means that the value is significantly lower than that of neutral.

	Anger in English									Anger in French							
Speaker	span	level	rINT	PVI	rate	syl	VQ	Id	Speaker	span	level	rINT	PVI	rate	syl	VQ	Id
B1SE	-	-	↑	-	-	-	-	8.4%	B1F	↑	-	-	↑	-	↓	creaky	43%
B2SE	↑	↓	-	↓	↓	-	creaky	**98%**	B2F	↑	-	-	-	↓	-	creaky	**97%**
B3SE	↑	-	↑	-	-	-	creaky	**98%**	B3F	↑	-	↑	-	-	-	harsh	**100%**
B4SE	-	-	↑	-	-	-	creaky	37%	B4F	↑	↓	-	-	-	-	-	**61%**
B5SE	↑	-	-	-	↓	-	creaky	46%	B5F	↑	-	-	-	-	↓	creaky	**63%**
MSE1	-	-	↑	-	-	-	-	33%	MF1	↑	↓	↑	↓	-	-	-	39%
MSE2	-	↓	↑	-	-	-	-	23%	MF2	-	↓	-	-	-	-	-	26%
MSE3	↑	-	↑	-	-	-	creaky	48%	MF3	↑	↓	↑	-	-	-	creaky	**85%**
MSE4	↑	-	↑	-	-	-	creaky	**98%**	MF4	↑	↓	↑	-	-	-	-	**78%**
MSE5	↑	↓	-	-	-	-	creaky	22%	MF5	↑	↓	-	-	-	-	creaky	**62%**
MSE6	-	-	-	-	-	-	creaky	11%	MF6	↑	↓	↑	-	-	-	harsh	**80%**
ASE	-	-	↑	-	↓	↑	harsh	**77%**									

	Fear in English									Fear in French							
Speaker	span	level	rINT	PVI	rate	syl	VQ	Id	Speaker	span	level	rINT	PVI	rate	syl	VQ	Id
B1SE	-	↑	↓	↑	-	↑	creaky	**51%**	B1F	↑	↑	↓	↑	↑	↑	creaky	**62%**
B2SE	↑	↑	↓	-	↑	-	-	**74%**	B2F	-	↑	↓	-	↑	-	-	**70%**
B3SE	-	↑	↑	-	-	-	whisper	43%	B3F	-	↑	↓	↓	-	-	whisper	43%
B4SE	↑	↑	↑	-	↓	↑	-	45%	B4F	-	↑	↑	-	-	↑	-	35%
B5SE	↑	-	-	-	-	↑	whisper	**79%**	B5F	↑	↑	↑	-	-	↑	-	49%
MSE1	-	↑	↑	-	-	-	-	35%	MF1	-	↑	↓	-	-	↑	-	37%
MSE2	↑	↑	-	-	↑	-	breathy	14%	MF2	-	-	-	-	-	-	-	7%
MSE3	↑	↑	↓	-	-	-	-	**88%**	MF3	-	↑	↓	-	-	↑	breathy	31%
MSE4	↑	↑	↓	-	-	↑	-	**68%**	MF4	↑	↑	↓	-	-	-	-	27%
MSE5	↑	↑	↓	↓	-	↑	breathy	44%	MF5	↑	↑	↓	-	-	-	-	13%
MSE6	-	-	-	-	-	-	breathy	23%	MF6	↑	↑	-	-	-	-	-	22%
ASE	↑	↑	↓	-	↑	↓	-	**78%**									

Table 7.6: Summary of production and perception results for each individual speaker. If the acoustic correlate is significantly differentiated from the neutral value, arrows are used. Upper arrow means that the value is significantly higher than that of neutral, and down arrow means that the value is significantly lower than that of neutral.

	Sadness in English									Sadness in French							
Speaker	span	level	rINT	PVI	rate	syl	VQ	Id	Speaker	span	level	rINT	PVI	rate	syl	VQ	Id
B1SE	-	↑	-	-	-	-	-	**58%**	B1F	-	↑	-	-	-	-	creaky	**72%**
B2SE	-	↑	-	-	↓	-	-	39%	B2F	-	↑	-	↓	-	-	-	**62%**
B3SE	-	↑	↑	-	-	↑	creaky	**54%**	B3F	-	↑	-	↓	-	-	harsh	47%
B4SE	-	↑	-	↓	-	↑	-	**78%**	B4F	-	↑	↓	-	-	-	creaky	**77%**
B5SE	-	↑	-	↓	↓	↑	creaky	**69%**	B5F	-	↑	↓	↓	-	↑	-	47%
MSE1	-	↑	↑	-	↓	↑	creaky	**58%**	MF1	-	↑	-	↑	-	-	-	**62%**
MSE2	-	-	-	-	-	-	-	45%	MF2	-	-	-	-	-	-	creaky	**81%**
MSE3	-	↑	↑	↑	-	-	-	48%	MF3	-	↑	↓	-	↓	↑	creaky	**76%**
MSE4	↑	-	↑	-	-	-	creaky	46%	MF4	↑	-	-	-	↓	↑	-	**88%**
MSE5	↑	-	-	↑	↓	-	creaky	24%	MF5	↑	-	-	-	-	-	-	**66%**
MSE6	-	-	↑	-	-	-	breathy	**57%**	MF6	↑	↑	-	-	-	-	-	26%
ASE	-	-	-	-	↓	↑	creaky	**64%**									

	Happiness in English									Happiness in French							
Speaker	span	level	rINT	PVI	rate	syl	VQ	Id	Speaker	span	level	rINT	PVI	rate	syl	VQ	Id
B1SE	↑	-	-	-	-	↑	-	**56%**	B1F	↑	-	-	-	-	-	smiling	49%
B2SE	-	↑	↓	↓	↑	-	smiling	**94%**	B2F	-	↑	-	↓	-	-	-	39%
B3SE	↑	-	-	-	↑	↑	-	**87%**	B3F	-	↑	↓	↓	↑	-	-	**70%**
B4SE	↑	-	-	-	↓	-	breathy	**55%**	B4F	↑	-	↓	↓	↑	↓	smiling	**65%**
B5SE	↑	-	-	↓	↓	↑	-	29%	B5F	↑	↑	↓	↓	-	-	smiling	49%
MSE1	-	-	-	-	-	-	-	32%	MF1	-	↑	-	-	↑	-	-	43%
MSE2	↑	-	-	-	-	-	smiling	**89%**	MF2	-	↓	-	-	↑	-	smiling	**55%**
MSE3	↑	-	↓	↑	↑	-	-	**72%**	MF3	-	-	↓	-	-	↑	smiling	**64%**
MSE4	↑	-	-	-	↓	↑	-	**88%**	MF4	↑	-	↓	-	-	↑	smiling	39%
MSE5	↑	-	↓	-	↓	-	smiling	**76%**	MF5	↑	↑	-	-	-	-	smiling	47%
MSE6	↑	-	-	-	-	-	smiling	28%	MF6	↑	-	-	-	-	-	smiling	**91%**
ASE	↑	-	-	-	↓	-	smiling	49%									

marise, all the measured acoustic parameters have been used by speakers to express the intended affective states. Nevertheless, group differences in the usage of these parameters have been observed. Results for individual speakers uncovered a considerable amount of variability. Different successful strategies are possible in the same language. Some strategies are found across the two languages: other strategies are more privileged in one of the studied languages.

Chapter 8

Discussion and Conclusion

This chapter summarises perception and production results and addresses the four research questions of the dissertation. The theoretical implications of the findings are discussed. Finally, suggestions for future research in child affective speech are given.

8.1 Summary of results

The perception test results showed that affective expressions encoded by bilingual children and by the majority of monolingual children were identified by adult listeners. The perception test allowed us to validate the methodology developed for the recording of child affective speech. The variability observed in the perception tests was related to the groups of listeners, to the groups of child speakers and to the individual speakers.

The anger of the Scottish actress has the best rate of identification accuracy - 80% for Scottish listeners and 73.3% for French listeners. Anger expressed by bilingual children in French and by French monolingual children was identified with a higher level by both French and Scottish listeners than anger expressed in English. Fear expressed by bilingual children was successfully identified both in English and in French. In contrast, fear encoded by French monolingual children was very poorly identified by both groups of listeners. The Scottish actress's fear was identified by Scottish listeners at 98.2%, while French listeners confused her fear with sadness (53.3% versus 46.7%). For sadness, French listeners showed high identification level for sadness across all groups of encoders: only two significant confusions with fear occurred in their identification of Scottish monolinguals and bilingual children in English. In contrast, Scottish listeners showed significant confusion of sadness with

fear for all the groups of children, except French monolingual children. The Scottish actress has been successfully identified by French listeners at 84.4%, and confused with fear by Scottish listeners (sadness 47.3% versus fear 52.7%). Happiness of bilinguals was identified more accurately in Scottish English than in French by Scottish listeners (70.9% versus 62.9%), but also by French listeners (56.0% versus 44.2%). Note that Scottish listeners showed higher identification levels of bilinguals, both in English and French, than the French listeners. Happiness expressed by both French and Scottish monolingual children was confused with sadness by Scottish listeners, but not by French listeners. Moreover, French listeners could identify happiness of the Scottish actress better (60.0%), than Scottish listeners (40%).

Production results for groups of children show that pitch range measures (span and level) are used by children to differentiate emotions, and that there are differences in the usage of pitch range measures by the four subgroups of children. The results showed that pitch span was narrower in French than in Scottish English, and that pitch level was higher in French than in Scottish English. At the same time, differences in the differentiation of emotions by level were significant between Scottish and French monolingual children, while the differentiation of emotions by span was realised in the same way. The differentiation of emotions by span has the same pattern in Scottish English and French: happiness and anger are realised with a wide span; sadness and neutral are realised with a narrow span; fear may be realised with a narrow or a wide span. The differentiation of emotions by level is similar in the two languages for anger, fear, sadness and neutral: fear has the highest level, it is followed by sadness, and finally anger and neutral have the lowest level. The realisation of happiness is different in Scottish English and French. In French, happiness is realised with a higher pitch level, similar to that of sadness. In English, happiness has a lower level: it is not statistically different from anger and neutral. Among the measured acoustic correlates of rhythm and speech rate, only consonantal variability, or $rINT$, was consistently associated with affective states expressed by children in Scottish English and French. The differentiation of emo-

tions by $rINT$ follows the same pattern across the two languages: anger is realised with a higher $rINT$ than the other affective states.

Group analysis also showed that bilingual children differentiate their affective states by the measured acoustic correlates: moreover, their pattern of differentiation of affective states corresponds with that of monolingual children. Some differences in the affective speech production were observed between monolingual and bilingual children. Bilingual children realised their pitch span in their two languages like Scottish monolinguals: they had a wide pitch span in both English and French. The only exception concerns happiness: this affective state was realised with a wider pitch span in Scottish English and a narrow pitch span in French, like monolingual speakers in the corresponding languages. The differentiation of emotions by span and level was performed according to the realisation pattern of monolingual children. Bilinguals have higher $rINT$ values than monolinguals in the corresponding languages. Bilinguals have a slower speech rate than Scottish monolinguals and a longer final syllable than French monolinguals.

The joint analysis of production and perception data identified a number of acoustic correlates and their particular usage for each studied affective state. The analysis of individual speakers allowed us to observe the scope of existing variability in the usage of the identified acoustic correlates and to describe strategies which may be effectively used by individual speakers to encode affective states. For anger, speakers used wide pitch span, low pitch level, high consonantal variability, low speech rate, creakiness or harshness. Individual variability in the anger realisation was found in the chosen combination of these identified acoustic correlates. Acoustic correlates of anger are used in a consistent way by all the individual speakers. For sadness, similar to neutral pitch span, higher than neutral pitch level, slow speech rate, long final syllable and creakiness were observed across many successfully identified individual speakers of Scottish English and French. Some speakers in French also used similar to neutral pitch level for their sadness. For happiness, one strategy was used by almost all speakers in English and some speakers in French: wide pitch

span, similar to neutral pitch level, sometimes in combination with lower than neutral consonantal variability $rINT$, long final syllable and smiling. Some speakers in French, whose happiness was successfully identified, used other strategies for pitch range: similar to neutral pitch span and high pitch level, wide pitch span and high level. In contrast with other emotions, not all the acoustic correlates identified for fear are used in the same direction by both Scottish and French speakers. While pitch level is realised higher than neutral by almost all the speakers, other acoustic correlates show more variation: pitch span may be wide or narrow; final syllable may be long or short; consonantal variability may be high or low; whisper, creakiness or breathiness may also be used.

8.2 General discussion

The main objective of the thesis was to investigate affective speech of monolingual and bilingual children in a cross-linguistic perspective. Four research questions were addressed: 1) how monolingual Scottish and French adults identify affective states encoded by monolingual and bilingual children; 2) whether acoustic correlates of affective speech could be identified in child productions; 3) whether there are cross-linguistic differences in affective speech of Scottish English and French monolingual children; 4) how simultaneous bilingual children produce affective speech in their two languages and how their affective productions correspond with those of monolinguals. Findings from the perception and production analysis of child affective speech are discussed in the following sections.

8.2.1 Cross-language identification of vocal emotions

The universality of paralinguistic behaviour has been questioned by recent studies Elfenbein & Ambady (2003), suggesting that even if people from different cultural groups can identify each others' emotions above chance, the identification of emotions, expressed by members from one's own cultural group, is generally more accurate than those expressed by members of a different cultural group. This phe-

nomenon is called a within-group advantage in the identification of emotions. The results for the identification of affective states in Scottish English and French present interesting findings in this respect. Apparently, Scottish and French listeners can be equally accurate in the identification of some affective states, like, for example, in the identification of anger expressed by the Scottish actress. The identification of some other affective states may present an in-group advantage: for example, the fear of the Scottish actress was identified at almost 100% by Scottish listeners, and only at 53% by French listeners. Interestingly, listeners can sometimes be more accurate in the identification of affective states in an unknown language than its native speakers: for example, sadness of the Scottish actress was identified by French listeners at 84% and by Scottish listeners at only 47%. This phenomenon has already been reported in a cross-linguistic study of Braun & Katerbow (2005), where Japanese female productions of joy were identified more successfully by German and American listeners than by Japanese listeners. The nature of this phenomenon has not yet been investigated.

Results for the identification of anger and happiness, encoded by children, present a particular interest in this study. They suggest that these affective states may be more successfully identified in one of the languages by the two groups of listeners. Specifically, anger, expressed by bilingual children in French and by French monolingual children, was identified more successfully by both French and Scottish listeners than anger expressed by bilinguals in Scottish English and by Scottish monolinguals. In contrast, happiness of bilinguals was identified with higher accuracy in Scottish English than in French by both Scottish and French listeners. These results look particularly important in the context of the present research where children were recorded according to the same methodology in the two studied languages. These results of the identification test suggest that children may express some affective states in a more expressive way in one of the languages, or, in other words, they may show some emotions more in one of the languages. Similar findings were reported in the study of Koven (2006) (see Chapter 4 for more details), which compared affective

displays of a Portuguese-French bilingual female in her two languages and discovered some differences: the studied bilingual realised more reserved displays of anger in Portuguese than in French. In general, results of the perception test performed on the developed corpus of child affective speech showed higher levels of identification for affective states realised by bilinguals than by monolinguals.

8.2.2 Acoustic correlates of affective speech

As was reported in Chapter 3, the search for stable associations between a group of acoustic parameters and specific affective states has not been very successful: a number of acoustic parameters relating to encoding of affective states were identified, but their association and usage with specific affective states were not consistent across different existing studies. Johnstone & Scherer (2000) suggested several possible reasons: a small number of acoustic parameters and simple methods for their measurement, neglect of arousal types for studied affective states, neglect of individual and cross-linguistic variability in affective speech research.

The present study of the affective speech was performed on a multi-speaker, cross-linguistically comparable corpus. Visual materials were used as the reference of arousal types for the recorded affective states. Measures of pitch range, rhythm and speech rate were taken in the detailed acoustic analysis. Pitch range was not measured automatically (mean F_0, standard deviation of F_0, the difference between maximum and minimum F_0), but according to the method, proposed by Ladd (1996), which allows study of two partially independent measures of pitch range: overall level and span. From the described methodology of the present study the question rises as to whether its results give more consistent associations between acoustic correlates and affective states. The two pitch range measures (span and level) were associated with affective states encoded by children, and they give a better understanding of pitch range variation than more widely used measures of F_0 mean and standard deviation. The value of F_0 mean may be influenced by the variation in F_0 level, or F_0 range, or both. Pitch span and pitch level give more detailed information about pitch

range variability. Production results for groups of children showed that pitch span, pitch level and consonantal variability were used consistently by children for the realisation of their affective states: anger, fear, sadness and happiness. Individual results also showed that some children used other analysed acoustic parameters, i.e. speech rate, final syllable lengthening, vocal variability and voice quality.

The joint analysis of production and perception data shows how individual speakers used acoustic correlates in the realisation of their affective states. For some affective states, like anger, identified acoustic correlates are used in a consistent way by all the individual speakers: wide pitch span, low pitch level, high consonantal variability, low speech rate, creakiness or harshness. For other affective states, like fear, not all the acoustic correlates are used in the same direction by different speakers: pitch span may be wide or narrow; final syllable may be long or short; consonantal variability may be high or low; whisper, creakiness or breathiness may also be used. This realisation variability was also observed in other perception and production studies of fear (Johnstone & Scherer 2000, de Mareuil et al. 2002, Gobl & Chasaide 2003, Burkhardt & Sendlmeier 2000). Apparently, some affective states are characterised by more variable usage of acoustic correlates than others.

The present study identified acoustic correlates of affective states and different strategies of their usage. The analysis of inter-speaker variability indicates that while some speakers use all the measured acoustic correlates to a significantly important extent, other speakers may prefer only some of these acoustic correlates, i.e. pitch range or voice quality. Apparently, usage of all the possible acoustic correlates is not obligatory for successful identification. Thus the question about the existence of consistent associations between acoustic correlates and affective states may be answered in the following way: acoustic correlates of affective states exist, but speakers use them depending on the chosen strategy of their affective realisation. The comparison of studies using a small number of speakers may show that the usage of acoustic correlates is not consistent or even contradictory for the same affective states if these speakers use different strategies in their affective realisations.

8.3 Evidence of cross-linguistic differences in affective speech: Scottish English and French

Results of the perception test have already shown that cross-linguistic differences were observed in the identification of affective states enacted by monolingual Scottish and French listeners. The question was whether cross-linguistic difference in the affective speech of Scottish and French children might be discovered on the production level. The production results analysis for groups of children showed that pitch span was narrower in French than in Scottish English, and that pitch level was higher in French than in Scottish English. At the same time, only differentiation of emotions by level was realised differently by Scottish and French monolingual children. It primarily concerns happiness. In French, happiness is realised with a higher pitch level, similar to that of sadness. In English, happiness has a lower level: it is not statistically different from anger and neutral. The differentiation of emotions by pitch span and by $rINT$ followed the same realisation pattern in the two languages.

The study of affective speech production on the individual level shed more light on the cross-linguistic differences in the strategies used by speakers for the realisation of their affective states. For anger, Scottish and French speakers use the same realisation strategy: wide pitch span, low pitch level, high consonantal variability, low speech rate, creakiness or harshness. For sadness, similar to neutral pitch span, higher than neutral pitch level, slow speech rate, long final syllable and creakiness were observed across many successfully identified individual speakers of both Scottish English and French. Some speakers in French also used similar to neutral pitch level for their sadness. For happiness, one strategy was used in the two languages: wide pitch span, similar to neutral pitch level, sometimes in combination with lower than neutral consonantal variability $rINT$, long final syllable and smiling. Several other strategies of pitch range usage for happiness were observed only for French speakers: similar to neutral pitch span and high pitch level, wide pitch span and high level. Fear results show that apart from pitch level, which is realised very high by French and Scottish speakers, the usage of other acoustic correlates is not

consistent in the two languages: pitch span may be wide or narrow, consonantal variability may be high or low, etc.

Pitch level, consistently used by speakers of the two language as low for anger and as high for fear, may be related in support of the biologically motivated theories developed by Ohala (1983), Gussenhoven (2004) to account for universality and language specificity in pitch variation. The universal connection between F_0 and size of the organism, with smaller larynxes producing higher pitch than larger ones, was identified and defined as 'Frequency Code'. The Frequency Code is claimed to be widely used for the expression of affective meanings linked with power relations: low pitch is associated with masculinity, dominance/assertiveness, aggression, confidence and protectiveness. The results of the child affective speech production show that children in both Scottish English and French realise their anger with a low pitch level, associated with dominance and aggression by the Frequency Code. In contrast, both French and Scottish children use high pitch level for their fear, thus creating an image of a "smaller" person, as suggested by the Frequency Code. The realisation of happiness showed some variability in Scottish English and French, as Scottish listeners tend to realise it with similar to neutral pitch level and a wide span, while French speakers realised it with a high level and wide or similar to neutral span. Similarities, but also cross-linguistic differences, were found in the affective speech production by Scottish English and French monolingual children. The differences were found in the usage of some acoustic correlates and in the preferred strategies for the realisation of affective states.

8.4 Cross-linguistic differentiation of affective speech by bilinguals

Research in acquisition of two languages from birth suggest that simultaneous exposure to two languages has an impact on child phonetic development: bilingual children apparently acquire certain phonetic abilities later than monolinguals. The present study shows that bilingual children can encode affective states in their two languages successfully; moreover, on average, their affective states are identified

even better than those of monolingual children. Rothman & Nowicki (2004) provided substantial support for the association between accuracy in identifying vocal emotions by children and children's social competence. Thus, the ability to successfully encode vocal emotions may also signal high social competence in bilingual children.

Acquisition of affective and linguistic prosody may be different for bilingual children. Slower rate of linguistic prosody acquisition was reported for bilingual children; it was explained by the presence of interaction between the two languages of bilinguals. Based on the results of the present study, the acquisition of affective prosody is not characterised by a slower rate of development, as bilinguals' affective states are well identified. Nevertheless, this finding does not preclude the absence of interaction in the affective speech between the two languages of bilinguals. Both the production and perception results suggest that there is interaction between the two languages. Interestingly, the nature of this interaction is different from that of observed by Gordeeva (2006) for the acquisition of linguistic prosody by Russian-Scottish English bilinguals. Gordeeva (2006) could associate the nature of interaction with the amount of language exposure: language interaction was bi-directional for a balanced bilingual and uni-directional for a dominant bilingual (from the dominant language to the weaker one). In the present study, bi-directional interaction was observed for all the studied bilingual children, independent of their background profile. Thus, this finding comes in support of other studies which call language interaction as a characteristic feature of bilingual phonetic acquisition.

The observed interaction in the affective speech does not preclude the presence of differentiation between the two languages. Bilingual children realise cross-linguistic differences found in the usage of some acoustic correlates and in the privileged strategies for the realisation of affective states, but the specific pattern of the affective speech realisation is individual for each bilingual child, depending on the affective state and on the used language. It was found out that affective speech production allowed considerable variability across individual speakers of the same language: dif-

ferent strategies are possible in the encoding of emotions of the same language. Some bilingual children may use strategies which are found across the two languages, other children may use strategies which are privileged in one of the languages. Moreover, it is not possible to generalise across the affective states and the two languages of the bilingual: each affective state has its specific realisation pattern and level of identification. In this context, the theory of integrated continuum, proposed by Cook (2003) to describe the linguistic organisation in the bilingual mind, looks particularly promising. This model sees the language system of the multilingual speaker as a whole and covers all possible relations between different components of the two languages in one's mind: from the complete separation to the total integration. This model of integrated continuum gives the necessary flexibility for the description of paralinguistic organisation in bilinguals.

8.5 Summary

The results of the thesis showed both group tendencies and individual variability in the child affective speech. The analysed acoustic measures were associated with affective speech. Pitch measures have been found particularly relevant for the cross-linguistic study of affective speech. No fixed pattern of successful encoding of affective states was found. One parameter or a flexible combination of several parameters may be successfully used for affective realisations. Voice quality is widely present in affective corpus of child speech: its detailed analysis is an area which begs further investigation, and it will be the subject of the future exploration of the recorded corpus. In general, the affective expressions encoded by children were successfully identified: observed variability in the identification was related to the language of listeners, the subgroups of children and to the individual speakers. Bilingual children received higher levels of identification than monolingual children. Having a wider range of ways and means for affective realisations, bilinguals may be more successful in communicating their affective states than monolingual children. The present study of affective speech production by simultaneous bilinguals makes a

contribution to the better understanding of the human capacity to acquire and use more than one language, and to adapt to the standards of more than one culture and society.

To conclude, the present study reports on the cross-linguistic and individual variability in affective speech production and advocates their recognition as characteristic of the affective speech. Taking this variability into account may help to address some existing difficulties in theoretical and applied affective speech research.

8.6 Suggestions for future research

The present dissertation reports new findings regarding acoustic correlates of affective speech, cross-linguistic and individual variability in child affective speech production and affective speech of bilingual children. A number of topics arise from these findings. The reported research was conducted in Scottish English and French languages, and it would be beneficial to extend the cross-linguistic study of affective speech production to other languages, particularly to culturally diverse languages where even more variability could probably be found. Moreover, this study was based on a corpus of monolingual and bilingual girls. It would thus be of interest to analyse affective speech realised by boys in order to investigate differences in male and female vocal encoding of emotions. The sex of the speaker may indeed be another factor of variability in affective speech. An even more complete understanding of child affective acquisition should require the collection and analysis of more cross-sectional and longitudinal data of child affective speech. Research in the affective speech of the parents of bilingual children may also give valuable information. In fact, little is known about the parents' influence on the development of individual and language specific characteristics in the affective speech of their bilingual children. This clearly represents another direction in affective speech research that deserves to be explored.

Appendix A

Statistical tests

Table A.1: Four groups of children (MF, MSE, BF and BSE): tests of between subject effects on dependant variables *span*, *level*. Transformed variable - average. Value is significant at $p < 0.05$.

Factor	Measure	Type IV S.S.	df	Mean Square	F	Sig.	Obs. Power
Group	span	7.943	3	2.648	1.306	0.278	0.337
Group	level	11.518	3	3.839	5.162	**0.002**	0.913
Emotion	span	63.005	4	15.751	7.769	**0.000**	0.997
Emotion	level	48.248	4	12.062	16.216	**0.000**	1.000
Interaction	span	14.943	12	1.245	0.642	0.825	0.326
Interaction	level	5.909	12	0.492	0.094	0.783	0.353

Table A.2: MF and BF children: tests of between subject effects on dependant variables *rate*, *PVI*, *rINT*, *syllable*. Transformed variable - average. Value is significant at $p < 0.05$.

Factor	Measure	Type IV S.S.	df	Mean Square	F	Sig.	Obs. Power
Group	rate	0.238	1	0.238	0.454	0.504	0.101
Group	PVI	262.083	1	262.083	0.980	0.328	0.162
Group	rINT	6,871.758	1	6,871.758	31.235	**0.000**	1.000
Group	syl	56,035.806	1	56,035.806	7.073	**0.011**	0.738
Emotion	rate	1.225	4	0.306	0.583	0.677	0.178
Emotion	PVI	978.982	4	244.746	0.915	0.464	0.265
Emotion	rINT	3,767.605	4	941.901	4.281	**0.005**	0.898
Emotion	syl	8,218.449	4	2,054.612	0.259	0.902	0.101
Interaction	rate	0.230	4	0.058	0.109	0.979	0.070
Interaction	PVI	1,094.820	4	273.705	1.023	0.406	0.295
Interaction	rINT	1,408.456	4	352.114	1.601	0.192	6.402
Interaction	syl	2,961.673	4	740.418	0.093	0.984	0.067

Table A.3: MSE and BSE children: tests of between subject effects on dependant variables *rate*, *PVI*, *rINT*, *syllable*. Transformed variable - average. Value is significant at $p < 0.05$.

Factor	Measure	Type IV S.S.	df	Mean Square	F	Sig.	Obs. Power
Group	rate	4.301	1	4.301	7.191	**0.010**	0.745
Group	PVI	382.946	1	382.946	0.366	0.549	0.091
Group	rINT	2,255.979	1	2,255.979	4.609	**0.038**	0.555
Group	syl	16,753.100	1	16,753.100	2.202	0.145	0.305
Emotion	rate	4.012	4	1.003	1.677	0.173	0.470
Emotion	PVI	2,372.758	4	593.190	0.566	0.689	0.174
Emotion	rINT	3,934.354	4	983.588	2.010	0.111	0.553
Emotion	syl	54,445.022	4	13,611.256	1.789	0.149	0.499
Interaction	rate	0.583	4	0.146	0.244	0.912	0.098
Interaction	PVI	4,182.155	4	1,045.539	0.998	0.419	0.288
Interaction	rINT	83.830	4	20.957	0.043	0.996	0.058
Interaction	syl	8,596.707	4	2,149.177	0.283	0.888	0.106

Table A.4: MANOVA results for pitch range (dependent variables - *span* and *level*). The value is significant at $p <0.05$. The first column includes speaker index and factor (L - language, E - emotion, I - interaction).

Subject	Span	Level
B1L	$F(1,76)=16.8$ $p <$**0.0001**	$F(1,76)=0.69$ $p <0.408$
B1E	$F(4,76)=15.3$ $p <$**0.0001**	$F(4,76)=138.8$ $p <$**0.0001**
B1I	$F(4,76)=8.302$ $p <$**0.0001**	$F(4,76)=2.89$ $p <$**0.029**
B2L	$F(1,79)=8.898$ $p <$**0.004**	$F(1,79)=0.267$ $p <0.607$
B2E	$F(4,79)=16$ $p <$**0.0001**	$F(4,79)=66$ $p <$**0.0001**
B2I	$F(4,79)=11$ $p <$**0.0001**	$F(4,79)=6$ $p <$**0.0001**
B3L	$F(1,79)=91.34$ $p <$**0.0001**	$F (1,79)=184.229$ $p <$**0.0001**
B3E	$F(4,79)=31.567$ $p <$**0.0001**	$F(4,79)=34.90$ $p <$**0.0001**
B3I	$F(4,79)=24.480$ $p <$**0.0001**	$F (4,79)=22.882$ $p <$**0.0001**
B4L	$F(1,79)=5.821$ $p <$**0.018**	$F(1,79)=193.060$ $p <$**0.0001**
B4E	$F(4,79)=35.261$ $p <$**0.0001**	$F(4,79)=72.307$ $p <$**0.0001**
B4I	$F(4,79)=8.147$ $p <$**0.0001**	$F(4,79)=24.704$ $p <$**0.0001**
B5L	$F(1,67)=0.0001$ $p <0.994$	$F(1,67)=60.044$ $p <$**0.0001**
B5E	$F(4,67)=57.252$ $p <$**0.0001**	$F(4,67)=121.537$ $p <$**0.0001**
B5I	$F(4,67)=1.21$ $p <0.315$	$F(4,67)=4.669$ $p <$**0.002**
MF1E	$F(4,39)=8.690$ $p <$**0.0001**	$F(4,39)=31.310$ $p <$**0.0001**
MF2E	$F(4,39)=2.852$ $p <$**0.038**	$F(4,39)=7.450$ $p <$**0.0001**
MF3E	$F(4,39)=9.660$ $p <$**0.0001**	$F(4,39)=38.913$ $p <$**0.0001**
MF4E	$F(4,39)=7.927$ $p <$**0.0001**	$F(4,39)=29.728$ $p <$**0.0001**
MF5E	$F(4,39)=9.707$ $p <$**0.0001**	$F(4,39)=8.664$ $p <$**0.0001**
MF6E	$F(4,39)=21.742$ $p <$**0.0001**	$F(4,39)=47.983$ $p <$**0.0001**
MSE1E	$F(4,39)=1.766$ $p <0.158$	$F(4,39)=14.621$ $p <$**0.0001**
MSE2E	$F(4,39)=16.974$ $p <$**0.0001**	$F(4,39)=17.597$ $p <$**0.0001**
MSE3E	$F(4,39)=7.810$ $p <$**0.0001**	$F(4,39)=38.015$ $p <$**0.0001**
MSE4E	$F(4,39)=48.139$ $p <$**0.0001**	$F(4,39)=24.000$ $p <$**0.0001**
MSE5E	$F(4,39)=16.451$ $p <$**0.0001**	$F(4,39)=8.102$ $p <$**0.0001**
MSE6E	$F(4,39)=6.149$ $p <$**0.001**	$F(4,39)=2.722$ $p <$**0.045**
ASE E	$F(4,39)=19.185$ $p <$**0.0001**	$F(4,75)=49.760$ $p <$**0.0001**

Table A.5: MANOVA results for dependent variables - syllable and speech rate. The value is significant at $p < 0.05$. The first column includes speaker index and factor (E - emotion).

Subject	syllable	speech rate
BE1E	$F(4,36)=9.172\ p <$**0.0001**	$F(4,36)=2.469\ p <0.065$
BF1E	$F(4,39)=6.959\ p <$**0.0001**	$F(4,39)=3.198\ p <$**0.024**
BE2E	$F(4,39)=2.642\ p <$**0.050**	$F(4,39)=12.057\ p <$**0.0001**
BF2E	$F(4,39)=0.554\ p <0.698$	$F(4,39)=6.651\ p <$**0.0001**
BE3E	$F(4,37)=10.230\ p <$**0.0001**	$F(4,37)=13.514\ p <$**0.0001**
BF3E	$F(4,39)=1.003\ p <0.419$	$F\ (4,39)=24.901\ p <$**0.0001**
BE4E	$F(4,39)=15.707\ p <$**0.0001**	$F(4,39)=67.306\ p <$**0.0001**
BF4E	$F(4,39)=5.540\ p <$**0.001**	$F(4,39)=10.654\ p <$**0.0001**
BE5E	$F(4,39)=8.292\ p <$**0.0001**	$F(4,39)=19.734\ p <$**0.0001**
BF5E	$F(4,27)=3.421\ p <$**0.025**	$F(4,27)=4.986\ p <$**0.005**
MF1E	$F(4,39)=3.077\ p <$**0.028**	$F(4,39)=5.476\ p <$**0.002**
MF2E	$F(4,39)=1.308\ p <0.286$	$F(4,39)=5.987\ p <$**0.001**
MF3E	$F(4,39)=17.879\ p <$**0.0001**	$F(4,39)=27.789\ p <$**0.0001**
MF4E	$F(4,39)=16.580\ p <$**0.0001**	$F(4,39)=13.453\ p <$**0.0001**
MF5E	$F(4,39)=1.260\ p <0.304$	$F(4,39)=0.673\ p <0.615$
MF6E	$F(4,39)=3.118\ p <$**0.027**	$F(4,39)=9.152\ p <$**0.0001**
MSE1E	$F(4,39)=11.804\ p <$**0.0001**	$F(4,39)=30.278\ p <$**0.0001**
MSE2E	$F(4,39)=7.574\ p <$**0.0001**	$F(4,39)=12.089\ p <$**0.0001**
MSE3E	$F(4,39)=11.729\ p <$**0.0001**	$F(4,39)=18.570\ p <$**0.0001**
MSE4E	$F(4,39)=4.290\ p <$**0.006**	$F(4,39)=8.436\ p <$**0.0001**
MSE5E	$F(4,39)=4.543\ p <$**0.005**	$F(4,39)=5.626\ p <$**0.001**
MSE6E	$F(4,39)=1.195\ p <0.330$	$F(4,39)=1.719\ p <0.168$
ASE E	$F(4,39)=6.405\ p <$**0.001**	$F(4,39)=4.518\ p <$**0.005**

Table A.6: MANOVA results for dependent variables - PVI and $rINT$. The value is significant at $p <0.05$. The first column includes speaker index and factor (E - emotion).

Subject	PVI	rINT
BE1E	$F(4,36)$=9.182 $p <$**0.0001**	$F(4,36)$=3.637 $p <$**0.015**
BF1E	$F(4,39)$=4.257 $p <$**0.007**	$F(4,39)$=1.494 $p <$0.225
BE2E	$F(4,39)$=5.266 $p <$**0.002**	$F(4,39)$=16504. $p <$**0.0001**
BF2E	$F(4,39)$=7.708 $p <$**0.0001**	$F(4,39)$=4.187 $p <$**0.007**
BE3E	$F(4,37)$=1.620 $p <$0.192	$F(4,37)$=4.272 $p <$**0.007**
BF3E	$F(4,39)$=7.192 $p <$**0.0001**	F $(4,39)$=6.499 $p <$**0.001**
BE4E	$F(4,39)$=4.747 $p <$**0.004**	$F(4,39)$=30.798 $p <$**0.0001**
BF4E	$F(4,39)$=1.839 $p <$0.143	$F(4,39)$=6.004 $p <$**0.001**
BE5E	$F(4,39)$=2.376 $p <$0.071	$F(4,39)$=1.539 $p <$0.212
BF5E	$F(4,27)$=3.980 $p <$**0.013**	$F(4,27)$=3.384 $p <$**0.026**
MF1E	$F(4,39)$=3.544 $p <$**0.016**	$F(4,39)$=27.441 $p <$**0.0001**
MF2E	$F(4,39)$=1.913 $p <$0.130	$F(4,39)$=0.946 $p <$0.449
MF3E	$F(4,39)$=1.472 $p <$0.264	$F(4,39)$=7.676 $p <$**0.0001**
MF4E	$F(4,39)$=2.075 $p <$0.105	$F(4,39)$=6.447 $p <$**0.001**
MF5E	$F(4,39)$=0.734 $p <$0.575	$F(4,39)$=3.109 $p <$**0.027**
MF6E	$F(4,39)$=0.926 $p <$0.460	$F(4,39)$=1.459 $p <$0.236
MSE1E	$F(4,39)$=1.948 $p <$0.124	$F(4,39)$=7.229 $p <$**0.0001**
MSE2E	$F(4,39)$=1.313 $p <$0.284	$F(4,39)$=11.849 $p <$**0.0001**
MSE3E	$F(4,39)$=6.185 $p <$**0.001**	$F(4,39)$=14.811 $p <$**0.0001**
MSE4E	$F(4,39)$=1.666 $p <$0.180	$F(4,39)$=6.034 $p <$**0.001**
MSE5E	$F(4,39)$=3.533 $p <$**0.016**	$F(4,39)$=4.702 $p <$**0.004**
MSE6E	$F(4,39)$=2.967 $p <$**0.033**	$F(4,39)$=1.525 $p <$0.216
ASE E	$F(4,39)$=0.160 $p <$0.957	$F(4,39)$=40.272 $p <$**0.0001**

Appendix B

Production data

Production data are presented in the tables below for bilingual and monolingual children as well as adults (mothers and actress). All the analysed acoustic parameters are given as means and standard deviations (in brackets), each calculated over 8 utterances for each affective states of individual speakers. Results of comparison across emotions, computed from Bonferroni post hoc tests of individual multivariate analysis of variance (MANOVA) for each speaker, are displayed in index numbers. For bilingual speakers two series of calculations were performed to allow comparison between emotions in each language. Different index numbers indicate that the difference between values is significant. Values with the same index number are not significantly different from each other. The absence of index numbers shows that values are not significantly different from each other.

For example, let us consider the following list of values: $191(26)^1$ $321(53)^2$ $234(54)^1$ $264(71)^{1,2}$ $278(36)^2$. This means that the values $191(26)^1$ and $234(54)^1$ are significantly different from $321(53)^2$ and $278(36)^2$, but not between each other. The value $264(71)^{1,2}$ does not differ significantly from the values with both 1 and 2 index numbers.

Table B.1: Results for analysed measures of individual bilingual speakers - B1 and B2.

Sp	M	Scottish English					French				
		A	F	S	H	N	A	F	S	H	N
B1	*span*	$41.5(0.7)^1$	$41.1(0.5)^1$	$41.1(0.4)^1$	$42.6(1.1)^2$	$41.2(0.2)^1$	$43.3(1.2)^2$	$42.9(0.5)^2$	$41.6(0.3)^1$	$42.7(0.9)^2$	$40.7(0.5)^1$
B1	*level*	$5.9(0.3)^1$	$8.6(0.3)^3$	$6.9(0.5)^2$	$6.1(0.3)^1$	$5.9(0.1)^1$	$5.6(0.7)^1$	$9.2(0.6)^3$	$7.2(0.4)^2$	$6.2(0.6)^1$	$5.6(0.2)^1$
B1	*rate*	4.9(0.5)	4.4(0.3)	4.9(0.7)	4.8(0.3)	5.0(0.3)	$4.4(0.1)^{1,2}$	$3.9(0.4)^1$	$4.3(0.3)^{1,2}$	$4.4(0.3)^{1,2}$	$4.6(0.6)^2$
B1	*PVI*	$71(8)^1$	$94(11)^2$	$71(12)^1$	$71(10)^1$	$59(15)^1$	$43(10)^{1,2}$	$55(23)^2$	$26(17)^1$	$32(10)^1$	$43(11)^1$
B1	*rINT*	$73(16)^2$	$45(10)^1$	$69(20)^{1,2}$	$58(17)^{1,2}$	$68(18)^{1,2}$	54(18)	41(9)	61(17)	59(14)	55(27)
B1	*syl*	$289(51)^1$	$438(100)^2$	$298(65)^1$	$375(38)^2$	$283(29)^1$	$191(26)^1$	$321(53)^2$	$234(54)^1$	$264(71)^{1,2}$	$278(36)^2$
B1	*VQ*	-	Creaky	-	-	-	Creaky	Creaky	Creaky	Smiling	-
B1	*Id*	8.4%	**51**%	**58**%	**56**%	-	**42.9**%	**62**%	**72**%	**49**%	-
B2	*span*	42.7(1.4)	43.6(2.3)	42.3(1.3)	41.8(0.5)	40.7(0.4)	$43.9(1.5)^2$	$40.7(0.3)^1$	$41.1(0.8)^1$	$41.1(0.5)^1$	$40.6(0.5)^1$
B2	*level*	$5.9(0.4)^1$	$10.2(1.7)^3$	$9.9(1.3)^3$	$8.1(0.4)^2$	$6.2(0.2)^1$	$7.1(1)^2$	$9.9(0.9)^4$	$8.5(0.5)^3$	$8.3(0.4)^3$	$5.9(0.14)^1$
B2	*rate*	$3.6(0.7)^1$	$5.1(0.3)^3$	$3.5(0.7)^1$	$4.6(0.5)^3$	$4.15(0.4)^2$	$2.9(0.7)^1$	$4.2(0.3)^2$	$3.5(0.5)^{1,2}$	$3.5(0.5)^{1,2}$	$3.7(0.3)^{1,2}$
B2	*PVI*	$54(24)^1$	$76(7)^{1,2}$	$83(17)^2$	$75(14)^{1,2}$	$89(16)^2$	$34(11)^1$	$36(16)^1$	$32(18)^1$	$27(11)^1$	$64(15)^2$
B2	*rINT*	$42(13)^2$	$13(5)^1$	$48(12)^2$	$27(12)^1$	$62(19)^2$	$43(23)^1$	$53(15)^{1,2}$	$91(42)^2$	$47(11)^{1,2}$	$72(35)^{1,2}$
B2	*syl*	320(84)	338(46)	414(144)	446(100)	373(34)	498(567)	385(44)	387(89)	527(156)	383(76)
B2	*VQ*	Creaky	-	-	Smiling	-	Creaky	-	-	-	-
B2	*Id*	**98**%	**73.8**%	**39**%	**94**%	-	**97**%	**70**%	**62**%	**39**%	-

Table B.2: Results for analysed measures of individual bilingual speakers - B3, B4 and B5.

Sp	M	Scottish English					French				
		A	F	S	H	N	A	F	S	H	N
B3	$span$	$44(0.7)^2$	$41.6(0.8)^1$	$42.1(0.7)^1$	$45.8(0.8)^3$	$42.5(1)^1$	$42.3(0.7)^2$	$41.4(0.4)^1$	$41.3(0.2)^1$	$41.6(0.3)^1$	$42.1(0.5)^1$
B3	$level$	$5.7(0.2)^{2,3}$	$6.0(0.4)^3$	$5.8(0.4)^{2,3}$	$5.6(0.1)^{1,2}$	$5.2(0.2)^1$	$6.7(0.5)^2$	$6.4(0.4)^2$	$7.01(0.3)^2$	$7.9(0.6)^3$	$5.5(0.2)^1$
B3	$rate$	$3.8(0.3)^1$	$3.9(0.3)^1$	$3.9(0.3)^1$	$4.4(0.1)^2$	$4.6(0.2)^2$	$2.8(0.2)^1$	$3.2(0.2)^1$	$2.9(0.3)^1$	$3.7(0.2)^2$	$2.7(0.2)^1$
B3	PVI	$101(10)$	$88(14)$	$100(10)$	$92(13)$	$92(16)$	$88(11)^1$	$84(15)^1$	$74(13)^1$	$74(12)^1$	$102(10)^2$
B3	$rINT$	$83(25)^2$	$78(12)^{1,2}$	$78(18)^{1,2}$	$55(14)^1$	$57(17)^1$	$94(32)^2$	$51(22)^1$	$73(28)^2$	$51(12)^1$	$108(38)^2$
B3	syl	$492(42)^2$	$480(52)^2$	$463(65)^2$	$441(27)^2$	$360(26)^1$	$521(68)$	$498(49)$	$526(94)$	$472(41)$	$483(65)$
B3	VQ	Creaky	Whisper	Creaky	-	-	Harsh	Whisper	Harsh	-	-
B3	Id	**98%**	**43%**	**54%**	**87%**	-	**100%**	**43%**	**47%**	**70%**	-
B4	$span$	$42.2(0.6)^1$	$42.2(0.5)^1$	$41.5(0.4)^1$	$43.7(0.7)^2$	$40.8(0.3)^1$	$42.6(0.6)^2$	$41.7(0.6)^1$	$41.1(0.3)^1$	$42.3(0.7)^2$	$41.5(0.2)^1$
B4	$level$	$6.1(0.2)^1$	$6.7(0.5)^2$	$6.4(0.3)^{1,2}$	$6.2(0.1)^1$	$6.1(0.1)^1$	$6.8(0.2)^{1,2}$	$8.9(0.5)^3$	$7.2(0.2)^2$	$6.9(0.4)^2$	$6.3(0.1)^1$
B4	$rate$	$3.7(0.3)^1$	$3.0(0.2)^1$	$3.4(0.2)^1$	$3.9(0.2)^1$	$4.7(0.2)^2$	$3.4(0.2)^1$	$3.3(0.3)^1$	$3.3(0.3)^1$	$4.2(0.4)^2$	$3.8(0.3)^1$
B4	PVI	$61(18)^2$	$64(12)^2$	$39(10)^1$	$64(13)^2$	$60(11)^2$	$79(8)$	$74(11)$	$70(13)$	$67(21)$	$83(8)$
B4	$rINT$	$117(24)^3$	$95(14)^2$	$74(11)^1$	$60(8)^1$	$48(4)^1$	$40(14)^1$	$72(28)^{1,2}$	$61(32)^1$	$58(12)^1$	$99(30)^2$
B4	syl	$412(66)^1$	$551(61)^2$	$502(56)^2$	$456(53)^1$	$348(39)^1$	$422(76)^{1,2}$	$494(79)^2$	$429(81)^{1,2}$	$332(46)^1$	$450(64)^{1,2}$
B4	VQ	Creaky	-	-	Breathy	-	-	-	Creaky	Smiling	-
B4	Id	**37%**	**45%**	**78%**	**55%**	-	**61%**	**35%**	**77%**	**65%**	-
B5	$span$	$44.3(1.7)^2$	$43.5(0.8)^2$	$41(0.5)^1$	$44.4(1.3)^2$	$39.7(0.7)^1$	$44.3(1)^2$	$44.3(0.5)^2$	$40.6(0.4)^1$	$43.7(2.8)^2$	$40.3(0.2)^1$
B5	$level$	$5.8(0.5)^1$	$6.4(0.4)^1$	$8.4(0.5)^2$	$5.5(0.4)^1$	$5.9(0.2)^1$	$6.2(0.4)^1$	$7.9(0.5)^2$	$9.5(0.5)^3$	$7.1(1.3)^{1,2}$	$6.3(0.1)^1$
B5	$rate$	$3.7(1.6)^1$	$3.8(0.4)^1$	$3.5(0.2)^1$	$3.9(0.2)^1$	$4.6(0.2)^2$	$3.5(0.2)$	$3.2(0.2)$	$3.3(0.4)$	$3.4(0.3)$	$3(0.2)$
B5	PVI	$77(13)^1$	$66(8)^1$	$77(13)^1$	$73(13)^1$	$82(8)^2$	$73(16)^1$	$61(11)^1$	$66(11)^1$	$63(12)^1$	$82(10)^2$
B5	$rINT$	$53(15)$	$39(11)$	$45(26)$	$44(7)$	$34(15)$	$78(20)^1$	$84(20)^{1,2}$	$66(14)^1$	$80(29)^1$	$94(22)^2$
B5	syl	$485(28)^2$	$492(51)^2$	$516(47)^2$	$470(65)^2$	$383(50)^1$	$303(74)^1$	$387(71)^{1,2}$	$408(50)^2$	$309(59)^{1,2}$	$336(56)^{1,2}$
B5	VQ	Creaky	Whisper	Creaky	-	-	Creaky	-	-	Smiling	-
B5	Id	**46%**	**79%**	**69%**	**29%**	-	**63%**	**49%**	**47%**	**49%**	-

Table B.3: Results for analysed measures of individual French speakers - MF1, MF2 and MF3.

Sp	M	Anger	Fear	Sadness	Happiness	Neutral
MF1	*span*	$42.5(0.7)^2$	$40.9(0.2)^1$	$41.6(0.7)^1$	$41.2(0.7)^1$	$41.4(0.5)^1$
MF1	*level*	$6.5(0.5)^1$	$8.5(0.3)^3$	$7.4(0.5)^2$	$7.3(0.6)^2$	$6.5(0.2)^1$
MF1	*rate*	4.36(0.3)	4.36(0.2)	4.26(0.1)	4.5(0.3)	4(0.2)
MF1	*PVI*	$39.2(16.1)^1$	$60.9(7.3)^2$	$58(13.2)^2$	$62.1(9.1)^2$	$49.6(21.6)^{1,2}$
MF1	*rINT*	$77.1(16.2)^2$	$30(4.8)^1$	$30.5(7.6)^1$	$31.2(8.4)^1$	$59.6(16.4)^1$
MF1	*syl*	305(42.4)	356(5)	310(47.7)	310(35.9)	310(35.9)
MF1	*FL*	1.27(1.2)	6.86(0.7)	3.65(1.5)	3.52(1.6)	1.18(0.4)
MF1	*VQ*	-	-	-	-	-
MF1	*Id*	**39**%	**37**%	**62**%	**40.3**%	-
MF2	*span*	$42(0.3)^2$	$41.7(0.3)^{1,2}$	$41.3(0.2)^1$	$41.6(0.4)^{1,2}$	$41.5(0.6)^{1,2}$
MF2	*level*	$6.4(0.2)^1$	$6.61(0.3)^{1,2}$	$6.7(0.2)^2$	$6.3(0.3)^1$	$6.9(0.2)^2$
MF2	*rate*	$3.3(0.1)^2$	$3.3(0.2)^2$	$3.2(0.1)^2$	$3.5(0.2)^2$	$3(0.4)^1$
MF2	*PVI*	45.7(13.8)	57.7(10.4)	46(12.7)	40.6(10.1)	45.1(16.9)
MF2	*rINT*	49.9(19)	57.1(29.4)	39(15.1)	52.2(10.3)	57.68(29.3)
MF2	*syl*	342(44.8)	345(7)	389(51.7)	354(61.3)	0(1)
MF2	*VQ*	-	-	Creaky	Smiling	-
MF2	*Id*	**26**%	7%	**81**%	**55**%	-
MF3	*span*	42.2(0.6)	41.8(0.5)	41(0.36)	42.6(1)	41.1(0.3)
MF3	*level*	$6.1(0.2)^1$	$7.9(0.4)^3$	$7(0.3)^2$	$6.1(0.6)^1$	$6(0.3)^1$
MF3	*rate*	3.5(0.2)	3.9(0.1)	3.4(0.2)	3.8(0.4)	4.5(0.2)
MF3	*PVI*	66.7(9.8)	56.1(12)	57.5(10.4)	60.9(11.6)	55.9(10.8)
MF3	*rINT*	$47.4(12.7)^{2,3}$	$23.9(10.9)^1$	$43.3(18.8)^2$	$38.4(13.1)^2$	$61.2(12.3)^3$
MF3	*syl*	472(58.6)	428(39)	479(45.1)	398(76.8)	282(35.4)
MF3	*VQ*	Creaky	Breathy	Creaky	Smiling	-
MF3	*Id*	**85**%	**31**%	**76.3**%	**64**%	-

Table B.4: Results for analysed measures of individual French speakers - MF4, MF5 and MF6.

Sp	M	Anger	Fear	Sadness	Happiness	Neutral
MF4	*span*	$41.4(0.3)^2$	$41.2(0.3)^2$	$41.1(0.3)^2$	$41.3(0.4)^2$	$40.6(0.2)^1$
MF4	*level*	$5.8(0.3)^1$	$7.7(0.7)^2$	$6.1(0.3)^1$	$6.2(0.7)^1$	$6.1(0.2)^1$
MF4	*rate*	$3.6(0.2)^2$	$3.5(0.2)^2$	$3(0.2)^1$	$3.7(0.2)^2$	$3.9(0.3)^2$
MF4	*PVI*	55.2(11.2)	54.1(10.5)	67.6(11.1)	56.3(7)	65.5(2)
MF4	*rINT*	$37.6(11.8)^3$	$19.2(5.1)^1$	$29.6(12)^2$	$16.3(5.4)^1$	$32.6(13)^2$
MF4	*syl*	329(46.6)	400(42.4)	440(64.9)	391(67.4)	248(29.5)
MF4	*VQ*	-	-	-	Smiling	
MF4	*Id*	**78**%	27%	**88**%	**39**%	-
MF5	*span*	$41.8(0.5)^2$	$41.5(0.5)^2$	$41.7(0.6)^2$	$42(0.5)^2$	$40.5(0.3)^1$
MF5	*level*	$5.8(0.4)^1$	$6.7(0.2)^2$	$6.5(0.4)^2$	$6.5(0.3)^2$	$6.2(0.1)^1$
MF5	*rate*	4.5(0.2)	4.3(0.5)	4.3(0.3)	4.4(0.3)	4.4(0.4)
MF5	*PVI*	63.3(13.5)	72.8(13.2)	67.3(8.6)	64(15.2)	64(14.1)
MF5	*rINT*	29.7(9.9)	27.9(4)	42.6(9.4)	31.4(11.4)	31.2(10.1)
MF5	*syl*	265(53.6)	276(29)	240(49)	239(31.8)	246(36.3)
MF5	*VQ*	Creaky	-	-	Smiling	-
MF5	*Id*	**62**%	13.1%	**66.3**%	**47.5**%	-
MF6	*span*	$42.6(0.79)^2$	$43.7(0.84)^3$	$42.1(0.8)^2$	$43.9(0.9)^3$	$40.8(0.3)^1$
MF6	*level*	$5.85(0.19)^1$	$7.33(0.45)^2$	$9.31(0.93)^3$	$6.57(0.64)^{1,2}$	$6.3(0.14)^1$
MF6	*rate*	$4.1(0.41)^{1,2}$	$4.9(0.3)^2$	$4(0.2)^1$	$4.4(0.1)^{1,2}$	$4.3(0.47)^{1,2}$
MF6	*PVI*	43(12)	55(9.3)	45.6(15.6)	50.3(10.9)	51.2(19.4)
MF6	*rINT*	60(33.7)	36.6(15.4)	49.1(16.7)	43.1(12.3)	48(15.1)
MF6	*syl*	313(76)	258(36.65)	294(45.5)	306(37.5)	339(24.8)
MF6	*VQ*	Harsh	-	-	Smiling	-
MF6	*Id*	**80**%	22%	26.1%	**91**%	-

Table B.5: Results for analysed measures of individual Scottish speakers - MSE1, MSE2 and MSE3.

Sp	M	Anger	Fear	Sadness	Happiness	Neutral
MSE1	*span*	40.6(0.7)	40.9(0.5)	40.2(0.4)	40.4(0.5)	40.5(0.3)
MSE1	*level*	$6.2(0.5)^1$	$7.2(0.5)^2$	$7(0.6)^2$	$6.1(0.4)^1$	$5.8(0.2)^1$
MSE1	*rate*	$4.6(0.3)^{1,2}$	$4.7(0.3)^{1,2}$	$4.2(0.3)^1$	$5.5(0.5)^2$	$5.8(0.2)^2$
MSE1	*PVI*	85.7(19)	83.4(9.8)	88.9(11)	89.6(8.4)	73.4(14)
MSE1	*rINT*	$57.6(25)^2$	$42.3(12.3)^{1,2}$	$56.1(16)^2$	$31.2(8.7)^1$	$25.6(4.4)^2$
MSE1	*syl*	382(68)	404(64)	417(104)	295(33)	233(18)
MSE1	*VQ*	-	-	Creaky	-	-
MSE1	*Id*	**32**%	**35**%	**58**%	**32.3**%	-
MSE2	*span*	$41.7(0.3)^2$	$43.1(0.4)^3$	$41.2(0.4)^{1,2}$	$42.5(1.3)^{2,3}$	$40.5(0.4)^1$
MSE2	*level*	$5.7(0.2)^1$	$6.5(0.3)^3$	$6.1(0.1)^2$	$5.8(0.3)^{1,2}$	$6.04(0.2)^2$
MSE2	*rate*	$3.8(0.1)^1$	$4.9(0.4)^2$	$4.1(0.3)^{1,2}$	$4.6(0.5)^{1,2}$	$4.3(0.3)^{1,2}$
MSE2	*PVI*	95.6(11)	82.3(23)	91.9(11)	93(11)	98.4(14)
MSE2	*rINT*	$75.1(21.4)^2$	$42.9(14)^1$	$43(13)^1$	$32(7.4)^1$	$24.9(17)^1$
MSE2	*syl*	489(56)	371(35)	455(51)	403(54)	451(38)
MSE2	*VQ*	-	Breathy	-	Smiling	-
MSE2	*Id*	23%	13.8%	**44.9**%	**88.8**%	-
MSE3	*span*	$42.4(0.6)^{1,2}$	$42.1(1.1)^{1,2}$	$41.1(0.8)^1$	$42.6(1.4)^2$	$40.5(0.3)^1$
MSE3	*level*	$5.7(0.3)^1$	$7.9(0.8)^2$	$7.7(0.4)^2$	$6.2(0.3)^1$	$6.01(0.1)^1$
MSE3	*rate*	$6.4(0.4)^{1,2}$	$6.4(0.4)^{1,2}$	$5.2(0.4)^1$	$6.6(0.6)^2$	$5.4(0.3)^{1,2}$
MSE3	*PVI*	$116(12)^2$	$102(10)^2$	$106(12)^2$	$109(4.2)^2$	$89(16)^1$
MSE3	*rINT*	$44.6(9.9)^2$	$21.4(4.9)^1$	$41.6(9.01)^2$	$25.5(4.8)^1$	$41.5(8.9)^2$
MSE3	*syl*	190(41)	297(42)	321(37)	243(63)	306(33)
MSE3	*VQ*	Creaky	-	-	-	-
MSE3	*Id*	**48**%	**88**%	**48**%	**72**%	-

Table B.6: Results for analysed measures of individual Scottish speakers - MSE4, MSE5 and MSE6.

Sp	M	Anger	Fear	Sadness	Happiness	Neutral
MSE4	*span*	$45.7(0.8)^2$	$45.3(3.05)^2$	$43.9(1.3)^2$	$51.5(0.6)^3$	$41.1(0.6)^1$
MSE4	*level*	$5.6(0.2)^1$	$8.9(1.6)^2$	$6.5(0.6)^1$	$5.7(0.2)^1$	$5.8(0.2)^1$
MSE4	*rate*	3.3(0.3)	3.4(0.3)	3.4(0.3)	3.6(0.2)	3.9(0.2)
MSE4	*PVI*	110(8.2)	108(9.3)	111(10.5)	104(9.8)	116(11)
MSE4	*rINT*	$111(19)^2$	$69(34)^1$	$90.2(19)^{1,2}$	$66.7(19)^1$	$68.7(14)^1$
MSE4	*syl*	540(105)	603(131)	539(103)	598(63)	431(43)
MSE4	*VQ*	Creaky	-	Creaky	-	-
MSE4	*Id*	**98**%	**68**%	**46**%	**88.3**%	-
MSE5	*span*	$42.3(0.8)^2$	$43.6(1.1)^3$	$42.1(0.7)^2$	$43.4(0.7)^{2,3}$	$40.8(0.2)^1$
MSE5	*level*	$5.7(0.3)^1$	$6.4(0.5)^2$	$6.3(0.2)^2$	$6.1(0.2)^2$	$5.9(0.1)^{1,2}$
MSE5	*rate*	3.8(0.2)	3.9(0.3)	3.9(0.2)	3.9(0.1)	4.2(0.2)
MSE5	*PVI*	$66.2(4.2)^{1,2}$	$62.4(9.1)^1$	$74(7.2)^2$	$73.1(11)^{1,2}$	$62.9(9.3)^{1,2}$
MSE5	*rINT*	$28.2(5.1)^{1,2}$	$28.5(5.1)^{1,2}$	$32.7(3.4)^2$	$23.9(2.9)^1$	$31.7(5.3)^2$
MSE5	*syl*	415(28.8)	430(64)	411(30)	404(19)	354(33)
MSE5	*VQ*	Creaky	Breathy	Creaky	Smiling	-
MSE5	*Id*	22%	14%	24%	**75.8**%	-
MSE6	*span*	$40.7(0.2)^2$	$40.9(0.4)^2$	$40.5(0.2)^2$	$40.7(0.2)^2$	$40.3(0.1)^1$
MSE6	*level*	$5.9(0.1)^{1,2}$	$6.1(0.2)^2$	$6.04(0.3)^{1,2}$	$5.8(0.2)^1$	$5.9(0.1)^{1,2}$
MSE6	*rate*	4.9(0.3)	4.9(0.3)	4.7(0.3)	4.6(0.3)	4.9(0.4)
MSE6	*PVI*	75(13.3)	65.8(9.2)	73.9(12)	61.9(11)	79.5(13)
MSE6	*rINT*	34.8(8.5)	39.5(8.8)	40.1(11)	36.3(12)	21.2(9.2)
MSE6	*syl*	373(49)	371(46)	361(29)	412(51)	381(64)
MSE6	*VQ*	Creaky	Breathy	Breathy	Smiling	-
MSE6	*Id*	11%	23%	**57**%	28%	-

Table B.7: Results for analysed measures of the Scottish actress - ASE.

Sp	M	Anger	Fear	Sadness	Happiness	Neutral
ASE	*span*	$41.9(0.3)^2$	$43.9(1.2)^3$	$41.4(1.2)^{1,2}$	$42.6(0.5)^2$	$40.5(0.4)^1$
ASE	*level*	$5.6(0.1)^1$	$8(0.7)^2$	$6(0.5)^1$	$5.1(0.4)^1$	$5.7(0.2)^1$
ASE	*rate*	$4(0.1)^1$	$5.3(0.2)^3$	$4.58(0.3)^2$	$4.5(0.2)^2$	$5.26(0.3)^3$
ASE	*PVI*	69.2(7.4)	68.3(10)	71.17(23.5)	70(7)	73.1(11.3)
ASE	*rINT*	$81.8(18.7)^2$	$22.7(5.8)^1$	$31.6(9.2)^1$	$29.3(9.3)^1$	$28.1(5.6)^1$
ASE	*syl*	$388(24.2)^3$	$316(17.6)^1$	$370(39)^3$	$339(45.9)^2$	$335(25.2)^2$
ASE	*VQ*	Harsh	-	Creaky	Smiling	-
ASE	*Id*	**77**%	**78**%	**64**%	**49**%	-

Appendix C

Recording procedure and illustrations

The recording was performed either in a recording studio or in a quiet schoolroom. Parents of the child were not present. The child and the experimenter were sitting at the table in front of each other. The child had a tiny clip-on microphone attached to her clothing. The experiment was also recorded on the video camera.

First of all, the child performed a picture-naming test. A pile of randomised cards was put on the table in front of the child, and the instructions were: 'Here are cards with different objects: apples, strawberries, bananas, etc. Please, take one card and say what you see on the picture. For example, if you see an apple, you say: 'I see an apple there'. You must always finish your sentence with 'there'. Let us try several cards.' Then, the child did several cards, and if there was a mistake with wording (the child could add other words, or change the word order, or forget to say 'there' at the end.), she was corrected and told that the same sentence always had to be used and that only the name of the object changed. When the child understood the test, she was asked to start the test and work through the whole pile of cards. All the cards were randomised. The child took a card from the pile, said 'I see a * there' and put it aside.

When the picture-naming test was finished, cards with facial expressions of four affective states (anger, fear, sadness and happiness) were put on the table. First of all, the experimenter checked that the child identified the emotions expressed on the pictures. The experimenter showed four cards, one by one, and asked: 'What do you think about the girl in the picture? How does she feel?' All the children identified the emotions without difficulty. Then the following instruction was given: 'Now, you will need to say one and the same sentence, 'I see a banana there', but in different ways. Please, look at the picture and say it in the same way as the child

159

in the picture would do it: in a happy way, in a sad way, in an angry way or in a scared way. For example, you look at the girl who is happy (the experimenter points at the picture with a happy girl.). And you say, 'I see a banana there' in a happy way, as if you were happy like the girl in the picture. When you look at the girl who is sad (the experimenter points at the picture with a sad girl.), you say, 'I see a banana there' in a sad way, as if you were sad like the girl in the picture. You always say the same sentence 'I see a banana there', but in different ways, in a sad way, in a happy way, in an angry way, in a scared way, like the girl in the picture.' Then the child was asked to do one card for each of the four emotions to check that she understood the instructions. Finally, the child was asked to do the whole pile of cards, one by one. The pile of cards was prepared in the following way: the cards were first randomised and then numbered. The numbering of the cards served for the association of the child vocal productions with intended affective states. Thus, the order of randomised cards was the same for all the recorded children.

Below are examples of illustrative drawings used for the experiments with children and adults.

Figure C.1: Child - fear.

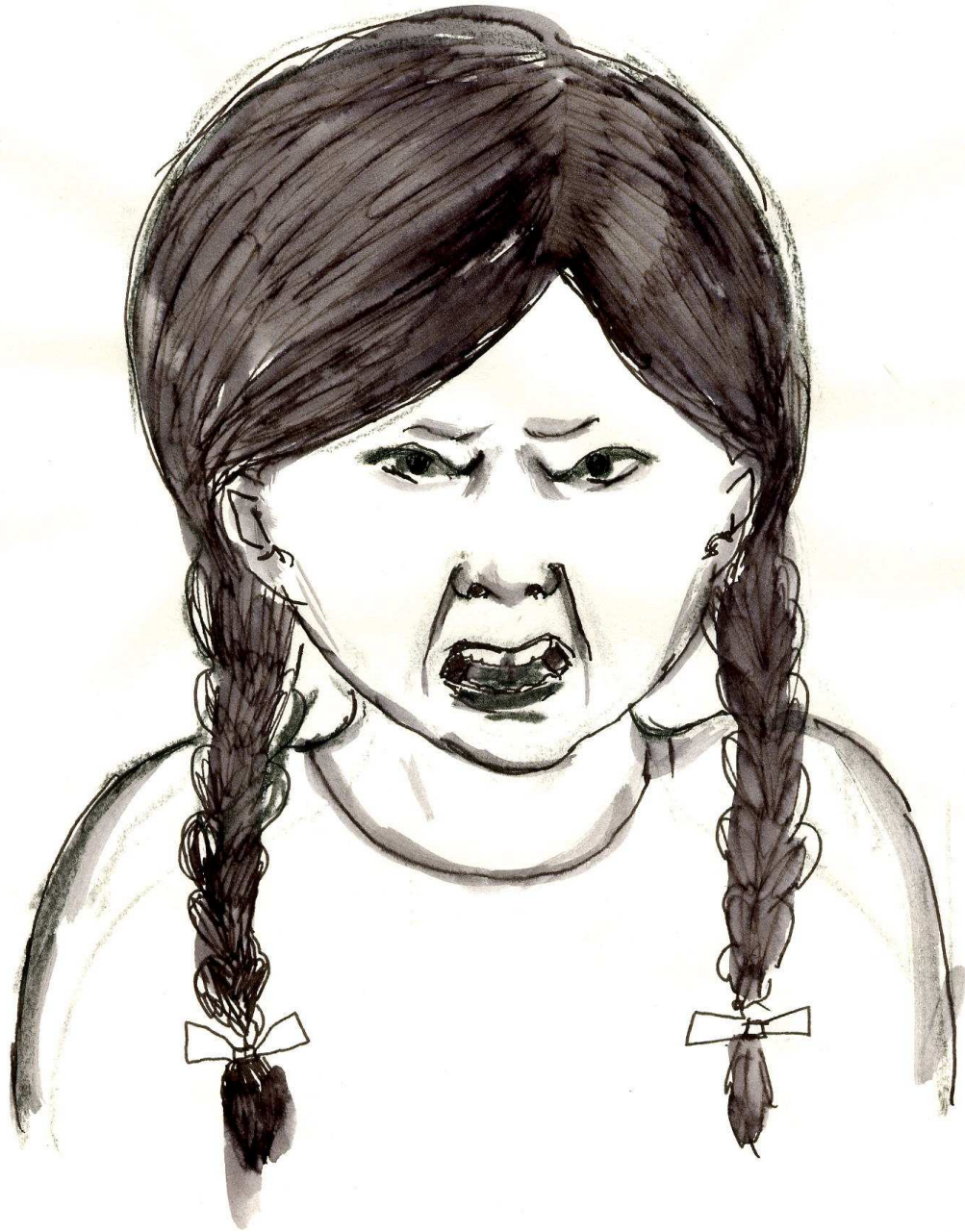**Figure C.2**: Child - anger.

Figure C.3: Child - sadness.

Figure C.4: Child - happiness.

Figure C.5: Female adult - fear.

Figure C.6: Female adult - anger.

Figure C.7: Female adult - sadness.

Figure C.8: Female adult - happiness.

Appendix D

Script for Perceval software

Below is an example of script that can be used with the Perceval software for perception tests.

```
[INFORMATION]
AUTHOR=Ioulia
DATE=14/11/2005
VERSION=3.0.2.2004
TITLE=test

[TRIAL_DATA]
TRIAL1=<mon4_sc0051_happiness.wav><4>
TRIAL2=<mon4_sc0074_sadness.wav><3>
TRIAL3=<mono_girl1_fr_0027_happiness.wav><4>
TRIAL4=<mono_girl1_fr_0054_anger.wav><1>
[...]
TRIAL495=<ma_f_a_2.wav><1>
TRIAL496=<ma_f_a_3.wav><1>
TRIAL497=<ma_f_a_4.wav><1>
TRIAL498=<ma_f_a_5.wav><1>

[TRIAL_EVENTS]
X10=BEGIN
X20=DISPLAY_FILEBMP<File5.bmp>
X30=PLAY_SOUND<#1>
X40=GET_INPUT<DELAY 15000>
X50=END

[SETTINGS_GROUP1]
INSTRUCTION_FORMAT=<exp1.txt><FONT Arial><SIZE 28><BKCOLOR 0xffffff>
<TXTCOLOR 0x000000><POSITION HCenter|Top>
RESPONSE_FORMAT=<$SUBJECT><$TITLE><$FILESCRIPT><$GROUP><$TRIAL>
<#1><#2><$RESPONSE><$ERROR><$RTIME>
IMAGE_FORMAT=<BKCOLOR 0xffffff>
TRAINING_ORDER=<RANDOM 1-4>
TRIAL_ORDER=<RANDOM>
INPUT=<1 CK_1><2 CK_2><3 CK_3><4 CK_4>
CORRECT=<#2>

PAUSE=60
```

Appendix E

Language background questionnaire

The questionnaire is to assess the language background of both parents and their bilingual child.

Child name:

1. Parents (mother and father)

   - Place of birth

   - Age

   - Education (please, specify place)

   - Profession

   - Languages (age of language learning, level)

   - Language used at work

   - Language used at home

2. Child

   - Place of birth

   - Age

   - Language history: Which languages and language skills (writing, reading, speaking, understanding) were acquired, when and how? Relate also childcare and school to languages.

   - Language proficiency: What is the child's proficiency in each of the four skills (writing, reading, speaking, understanding) in each language?

   - Function of languages: Which languages and language skills are used currently, in what context, for what purpose and to what extent?

- Language modes: How often and for how long is the child in a monolingual mode (when only one language is active) and in a bilingual mode (when both languages are active)? When in a bilingual mode, how much code-switching and borrowing is taking place?

3. Duration of residence in Scotland

4. Dialect awareness (father, mother, child)

5. Predominant home language use, parents' language use policy

6. Parent's view on the child language use and competence

7. Child's attitude to the fact of being bilingual

8. Any other comment(s)

References

Abercrombie, D. (1967), *Elements of General Phonetics*, Edinburgh University Press, Edinburgh.

Amir, N., Ron, S. & Laor, N. (2000), Analysis of an emotional speech corpus in hebrew based on objective criteria, *in* R. Cowie, E. Douglas-Cowie & M. Schröder, eds, 'Proceedings of the ISCA Workshop on Speech and Emotion', Belfast, Ireland.

Andersen, P. A. & Guerrero, L. K., eds (1998), *Handbook of Communication and Emotion: Research, Theory, Applications and Contexts*, Academic press.

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R. & Yamada, T. (2004), 'Perceived phonetic dissimilarity and l2 speech learning: the case of japanese /r/ and english /l/ and /r/', *Journal of Phonetics* **32**, 333–250.

Arvanti, A., Ladd, D. & Mennen, I. (2006), 'Phonetic effects of focus and "tonal crowding" in intonation: Evidence from Greek polar questions', *Speech Communication* **48**, 667–696.

Aubergé, V. & Cathiard, M. (2003), 'Can we hear the prosody of smile?', *Speech Communication* **40**.

Audibert, N., Aubergé, V. & Rilliard, A. (2004), Ewiz : contrôle d'émotions authentiques, *in* 'Actes des Journées d'Etudes sur la Parole', Fès, Maroc.

Bachorowski, J.-A. (1999), 'Vocal expression and perception of emotion', *Current Directions in Psychological Science* **8**, 53–57.

Banse, R. & Scherer, K. R. (1996), 'Acoustic profiles in vocal emotion expression', *Journal of Personality and Social Psychology* **70**(3), 614–636.

Bänziger, T. & Scherer, K. R. (2005), 'The role of intonation in emotional expressions', *Speech Communication* **46**, 252–267.

Blanc, M. (2000), *Bilinguality and Bilingualism*, Cambridge University Press, New York.

Bosch, L. & Sebastián-Gallés, N. (2003*a*), Language experience and the perception of a voicing contrast in fricatives, *in* 'Proceedings of the 16th International Congress of Phonetic Sciences', Barcelona, Spain, pp. 1987–1990.

Bosch, L. & Sebastián-Gallés, N. (2003*b*), 'Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life', *Language and Speech* **46**, 217–243.

Braun, A. & Katerbow, M. (2005), Emotions in dubbed speech: an intercultural approach with respect to f_0, *in* 'Proceedings of Interspeech-2005', Lisbon, Portugal, pp. 521–524.

Burkhardt, F. & Sendlmeier, W. F. (2000), Verification of acoustical correlates of emotional speech using formant-synthesis., *in* R. Cowie, E. Douglas-Cowie & M. Schröder, eds, 'Proceedings of the ISCA Workshop on Speech and Emotion', Belfast, Ireland.

Burnham, D. (2003), 'Language specific speech perception and the onset of reading', *Reading and Writing: An Interdisciplinary Journal* **16**, 573Ű–609.

Campbell, N. (2000), Databases of emotional speech, *in* R. Cowie, E. Douglas-Cowie & M. Schröder, eds, 'Proceedings of the ISCA Workshop on Speech and Emotion', Belfast, Ireland.

Campbell, N. & Mokhtari, P. (2006), Voice quality: the 4th prosodic dimension, *in* 'Proceedings of the XVth International Congress of Phonetic Sciences', Barcelona, Spain, pp. 2417–2420.

Camras, L. (1991), Conseptualizing early infant affect: View II. A dynamical systems perspective on expressive development, *in* K. Strongman, ed., 'International review of studies on emotion', Vol. 1, Wiley, New York.

Camras, L. A. (1992), 'Expressive development and basic emotions', *Cognition and Emotion* **6**, 269–283.

Camras, L. A., Holland, E. A. & Patterson, M. J. (1993), Facial expression, *in* M. Lewis & J. M. Haviland, eds, 'Handbook of Emotions', The Guilford Press, New York.

Camras, L., Malatesta, C. & Izard, C. E. (1991), The development of facial expressions in infancy, *in* R. Feldman & B. Rimé, eds, 'Fundamentals of nonverbal behavior', Cambridge University Press, Cambridge, UK.

Caramazza, A. & Miozzo, M. (1997), 'The relation between syntactic and phonological knowledge in lexical access: evidence from the "tip-of-the-tongue" phenomenon', *Cognition* **64**, 309–343.

Caspi, A., Elder, G. H. & Bem, D. J. (1987), 'Moving against the world: life-course patterns of explosive children', *Developmental Psychology* **21**.

Catford, J. C. (1977), *Fundamental problems in phonetics*, Edinburgh University Press.

Chang, W. C., Koh, J. B. K., Tan, P., & Neo, A. (2002), Fusion or alternation - two models of bilingualism and psychological consequences in Singapore, *in* 'The fourth International conference on Psychology of the Chinese People', Academia Sinica, Taipei, Taiwan.

Chen, A., Gussenhoven, C. & Rietveld, T. (2002), Language-specific uses of the effort code, *in* 'Proceedings of the International Conference Speech Prosody 2002', Aix-en-Provence, France.

Chen, A. J. (2005), Universal and language-specific perception of paralinguistic intonational meaning, PhD thesis, Radboud University, Nijmegen, the Netherlands.

Chen, A., Rietveld, T. & Gussenhoven, C. (1999), Language-specific effects of pitch

range on the perception of universal intonational meaning, *in* 'Proceedings of the 9th Eurospeech Conference', pp. 1403–1406.

Clément, J. (1999), Structure des représantations prosodiques. Développement normal et pathologique du traitement de la prosodie, PhD thesis, Université de Paris 5.

Cohen, J. (1960), 'A coefficient of agreement for nominal scales', *Educational and Psychological Measurement* **20**, 37–46.

Cohn, J. & Katz, G. S. (1998), Bimodal expression of emotion by face and voice, *in* 'Workshop on Face / Gesture Recognition and Their Applications, The Sixth ACM International Multimedia Conference'.

Cole, P. M., Bruschi, C. J. & Tamang, B. L. (2002), 'Cultural differences in children's emotional reactions to difficult situations', *Child Development* **73**(3), 983–996.

Cook, V. J. (2002), Bilingual congnition, *in* 'Proceedings of EUROSLA Annual Meeting', Basel.

Cook, V. J. (2003), Introduction: the changing L1 in the L2 user's mind, *in* V. J. Cook, ed., 'Effects of the Second Language on the First', Multilingual Matters, Clevdon, pp. 1–18.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W. & Taylor, J. I. (2001), 'Emotion recognition in human-computer interaction', *IEEE Signal Processing Magazine* **18**, 32–80.

Daniels, H. (1996), *Introduction to Vygotsky*, Routledge, London.

Dauer, R. M. (1983), 'Stress-timing and syllable-timing reanalysed', *Journal of Phonetics* **11**, 51–69.

de Bot, K. (1992), 'A bilingual processing model: Levelt's 'speaking' model adapted', *Applied Linguistics* **13**, 1–24.

de Mareuil, P. B., Célérier, P. & Toen, J. (2002), Generation of emotions by a morphing technique in English, French and Spanish, *in* 'Proceedings of the International Conference Speech Prosody 2002', Aix-en-Provence, France.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M. & Gagnon, D. A. (1997), 'Lexical access in aphasic and nonaphasic speakers', *Psychological Review* **104**, 801–838.

Denham, S. A. (1997), "When I have a bad dream, mommy holds me': Preschoolers' conceptions of emotions, parental socialization and emotional competence', *International Journal of Behavioral Development* **20**.

der Meulen, S. V., Janssen, P. & Os, E. D. (1997), 'Prosodic abilities in children with specific language impairment', *Journal of Communication Disorders* **30**, 155–170.

Douglas-Cowie, E., Cowie, R. & Schröder, M. (2000), A new emotion database: Considerations, sources and scope, *in* R. Cowie, E. Douglas-Cowie & M. Schröder, eds, 'Proceedings of the ISCA Workshop on Speech and Emotion', Belfast, Ireland.

Dromey, C., Silveira, J. & Sandor, P. (2005), 'Recognition of affective prosody by speakers of English as a first or foreign language', *Speech Communication* **47**(3), 351–359.

Dunn, J. & Brown, J. (1991), Relationships, talk about feelings, and the development of affect regulation in early childhood, *in* J. Garber & K. Dodge, eds, 'The development of emotion regulation and dyregulation', Cambridge University Press, Cambridge.

Ekman, P. (1992), 'An argument for basic emotions', *Cognition and Emotion* **6**, 169–200.

Ekman, P. (1994), 'Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique', *Psychological Bulletin* **115**, 268–287.

Ekman, P. (2003), *Emotions revealed*, Times books.

Ekman, P. & Friesen, W. V. (1971), 'Constants across cultures in the face and emotion', *Journal of Personality and Social Psychology* **17**, 124–129.

Elfenbein, H. A. & Ambady, N. (2003), 'Universal and cultural differencies in recognizing emotions', *Current Directions in Psychological Science* **12**, 159–164.

Emde, R. N., Plomin, R., Robinson, J., Corley, R., Fries, J. D., Fulker, D. W., Reznick, J. S., Campos, J., Kagan, J. & Zahn-Waxler, C. (1992), 'Temperament, emotion and cognition at fourteen months: the MacArthur longitudinal twin study', *Child Development* **63**, 1437–1455.

Ferrand, L. (2001), 'La production du langage: Une vue d'ensemble', *Psychologie Française* **46**, 3–15.

Field, T. M. & Walden, T. A. (1983), 'Production and discimination of facial expressions by preschool children', *Child Development* **53**, 1299–1300.

Flege, J. E. (1995), Second-language speech learning: Theory, findings and problems, *in* W. Strange, ed., 'Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research', Timonium, MD, York Press, pp. 233–272.

Friedlmeier, W. & Trommsdorff, G. (2002), Negative emotional reaction and interactive regulation in 2- and 5-year old German and Japanese children with their mothers, *in* 'The XVIIth Meetings of the ISSBD', Ottawa, Canada.

Frijda, N. H. (1993), Moods, emotion episodes, and emotions, *in* M. Lewis & J. M. Haviland, eds, 'Handbook of Emotions', The Guilford Press, New York.

Gardiner, H. W. (2001), Culture, context, and development, *in* D. R. Matsumoto, ed., 'Handbook of Culture and Psychology', Oxford University Press, Cary, NC, pp. 101–117.

Gay, T. (1981), 'Mechanisms in the control of speech rate', *Phonetica* **38**, 148–158.

Genesee, F. (1989), 'Early bilingual development: One language or two?', *Journal of Child Language* **16**, 161–179.

Genesee, F. (2003), Rethinking bilingual acquisition, *in* J.-M. Dewaele, ed., 'Bilingualism : Beyond Basic Principles', Multilingual Matters 123, Clevedon, pp. 204–228.

Gérard, C. & Clément, J. (1998), 'The structure and development of French prosodic representations', *Language and Speech* **41**, 117–142.

Gobl, C. & Chasaide, A. N. (2003), 'The role of voice quality in communicating emotion, mood and attitude', *Speech Communication* **40**, 189Ű–212.

Goldsmith, H. H. (1993), Temperament: variability in developing emotion systems, *in* M. Lewis & J. M. Haviland, eds, 'Handbook of emotions', Guilford, New York, pp. 353–364.

Goldsmith, H. H. & Campos, J. (1982), Toward a theory of infant temperament, *in* R. N. Emde & R. J. Harmon, eds, 'The development of attachment and affiliative systems', Plenum Press, New York.

Gordeeva, O. B. (2006), Language interaction in the bilingual acquisition of sound structure: A longitudinal study of vowel quality, duration and vocal effort in pre-school children speaking Scottish English and Russian, PhD thesis, Queen Margaret University College.

Grabe, E. & Low, E. L. (2002), 'Durational variability in speech and the rhythm class hypothesis', *Papers in Laboratory Phonology* **7**.

Green, D. W. (1998), 'Mental control of the bilingual lexico-sematic system', *Bilingualism: Language and Cognition* **1**, 67–81.

Grichkovtsova, I. (2002), Acquisition des émotions/attitudes prosodiques du français par des enfants russes scolarisé en France, Ms, Université Stendhal Grenoble 3.

Grosjean, F. (1998), 'Studying bilinguals: Methodological and conceptual issues', *Bilingualism: Language and Cognition* **1**, 131–149.

Grosjean, F. (2001), The bilingual's language modes, *in* J. Nicol, ed., 'One Mind, Two Languages', Oxford: Blackwell, pp. 1–22.

Guion, S. (2003), 'The vowel systems of quichua-spanish bilinguals: Age of acquisition effects on the mutual influence of the first and second languages', *Phonetica* **60**(2), 98–128.

Gussenhoven, C. (2004), *The phonology of tone and intonation*, Cambridge University Press, Cambridge.

Gussenhoven, C. (2005), Transcription of Dutch intonation, *in* S.-A. Jun, ed., 'Prosodic typology and transcription: A unified approach', Oxford University Press, Oxford.

Hadding-Koch, K. & Studdert-Kennedy, M. (1964), 'An experimental study of some intonation contours', *Phonetica* **11**, 175–185.

Harris, P. (1993), Understanding emotion, *in* M. Lewis & J. M. Haviland, eds, 'Handbook of Emotions', Guilford, New York, pp. 237–246.

Heise, D. R. & O'Brien, J. (1993), Emotion expression in groups, *in* M. Lewis & J. M. Haviland, eds, 'Handbook of Emotions', The Guilford Press, New York.

Hermes, D. J. & Rump, H. H. (1994), 'Perception of prominence in speech intonation induced by rising and falling pitch movements', *Journal of the Acoustical Society of America* **96**, 83–92.

Holodynski, M. & Friedlmeier, W. (2006), *Development of emotions and emotion regulation*, Springer, Berlin.

Izard, C. (1971), *The Face of Emotion*, Appleton-Century-Crofts, New York.

Izard, C. E. (1979), *The maximality discriminative facial movement coding system (Max)*, Information Technologies and University Media Services, University of Delaware, Newark.

Izard, C. E., Dougherty, L. M. & Hembree (1983), *A system for identifying affect expressions by holistic judgements (Affex)*, University of Delaware, Computor Network Services and University Media Services, Newark.

Johnstone, T. & Scherer, K. R. (2000), Vocal communication of emotion, *in* M. Lewis & J. M. Haviland, eds, 'Handbook of emotions', Guilford, New York, pp. 220–235.

Kehoe, M. M. (2002), 'Developing vowel systems as a window to bilingual phonology', *International Journal of Bilingualism* **6**, 315–334.

Khattab, G. (2002), Sociolinguistic competence and the bilingual's adoption of phonetic variants: Auditory and instumental ata from English-Arabic bilinguals, PhD thesis, University of Leeds.

Kopp, C. B. (1993), Emotional distress and control in young children, *in* N. Eisenberg & R. A. Fabes, eds, 'Emotion and its regulation in early development', Jossey Bass, San Francisco, pp. 41–56.

Koven, M. (2006), Feeling in two languages: A comparative analysis of a bilingual's affective displays in French and Portuguese, *in* A. Pavlenko, ed., 'Bilingual minds: Emotional experience, expression, and representation', Multilingual Matters, Clevedon, UK.

Kroll, J. F. & de Groot, A. M. B. (1997), Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages, *in* A. M. B. de Groot & J. F. Kroll, eds, 'Tutorials in Bilingualism : Psycholinguistic Perspectives', Lawrence Erlbaum Associates, New Jersey, pp. 169–199.

Ladd, D. R. (1996), *Intonational phonology*, Cambridge University Press.

Laukkanen, A.-M., Vilkman, E., Alku, P. & Oksanen, H. (1996), 'Physical variation related to stress and emotional state: a preliminary study', *Journal of Phonetics* **46**, 313–335.

Laver, J. (1994), *Principles of Phonetics*, Cambridge University Press, Cambridge.

Lazarus, R. S. (1991), *Emotion and Adaptation*, Oxford University Press, New York.

Levelt, W. J. M. (1989), *Speaking: From intention to articulation*, MIT Press, Cambridge, MA.

Lewis, M. & Haviland, J. M., eds (1993), *Handbook of Emotions*, The Guilford Press, New York.

Low, E. L., Grabe, E. & Nolan, F. (2001), 'Quantitative characterisations of speech rhythm: 'syllable-timing' in Singapore English', *Language and Speech* **43**, 377–401.

Makarova, V. & Petrushin, V. A. (2003), Phonetics of emotion in Russian speech, *in* D. R. M.J. Sole & J. Romero, eds, 'Proceedings of the XVth International Conference of Phonetic Sciences', Barcelona, Spain.

Mejvaldova, J. (2001), Expressions prosodiques de certaines attitudes en tchèque et en français. Etude comparative, PhD thesis, Univerzita Karlova, Université de Paris 7.

Milroy, L. & Muysken, P. (1995), Introduction: Code-switching and bilingualism research, *in* L. Milroy & P. Muysken, eds, 'One speaker two languages: Cross-disciplinary perspectives on code-switching', Cambridge University Press, New York, pp. 1–14.

Mozziconacci, S. (2002), Prosody and emotions, *in* 'Proceedings of the International Conference Speech Prosody 2002', Aix-en-Provence, France.

Mozziconacci, S. & Hermes, D. (1999), Role of intonation patterns in conveying emotion in speech, *in* 'Proceedings of the International Congress of Phonetic Sciences', San Francisco, USA.

Mozziconacci, S. J. L. (1998), Speech variability and emotion: Production and perception, PhD thesis, Technical University Eindhoven.

Murray, I. R. & Arnott, J. L. (1993), 'Towards the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion', *Journal of Acoustic Society of America* **93**(2), 1097–1108.

Nicoladis, E. (2001), Finding words in the input: Evidence from a bilingual child, *in* J. Cenoz & F. Genesee, eds, 'Trends in Bilingual Acquisition', Benjamins, Amsterdam, pp. 131–148.

Nicoladis, E. & Genesee, F. (1996), 'A longitudinal study of pragmatic differentiation in young bilingual children', *Language Learning* **46**, 439–464.

Oatley, K. & Johnson-Laird, P. N. (1987), 'Towards a cognitive theory of emotions', *Cognition and Emotion* **1**.

Ohala, J. J. (1983), 'Cross-language use of pitch: An ethological view', *Phonetica* **40**, 1–18.

Ohara, Y. (1992), Gender-dependent pitch levels in Japanese and English: A comparative study, *in* K. Hall, J. Nichols & O. J. J. Ohala, eds, 'Proceedings of the Second Berkeley Women and Language Conference: Locating Power.', Berkeley, CA, pp. 469–477.

Paeschke, A. & Sendlmeier, W. F. (2000), Prosodic characteristics of emotional speech: Measurements of fundamental frequency movements., *in* R. Cowie, E. Douglas-Cowie & M. Schröder, eds, 'Proceedings of the ISCA Workshop on Speech and Emotion', Belfast, Ireland.

Palmer, J., ed. (2001), *Fifty Modern Thinkers on Education : From Piaget to the Present Day*, Routledge, London, UK.

Panksepp, J. (1993), Neurochemical control of moods and emotions: Amino acids to neuropeptides, *in* M. Lewis & J. M. Haviland, eds, 'Handbook of Emotions', The Guilford Press, New York.

Paradis, J. (2000), Beyond 'one system or two': Degrees of separation between the languages of French-English bilingual children, *in* S. Döpke, ed., 'Cross-linguistic structures in simultaneous bilingualism', Benjamins, Amsterdam, Philadelphia, pp. 175–200.

Paradis, J. & Genesee, F. (1996), 'Syntactic acquisition in bilingual children: Autonomous or interdependent?', *International Journal of Bilingualism* **18**, 1–25.

Patterson, D. (2000), A linguistic approach to pitch range modelling, PhD thesis, University of Edinburgh.

Patterson, D. & Ladd, R. D. (1999), Pitch range modelling: New demensions of variation, *in* 'Proceedings of the International Congress of Phonetic Sciences', San Francisco, USA.

Pavlenko, A. (2002), 'Bilingualism and emotions', *Multilingua* **21**.

Pavlenko, A., ed. (2006), *Bilingual minds: emotional experience, expression and representation*, Multilgual Matters.

Peperkamp, S., Dupoux, E. & Sebastian-Galles, N. (1999), Perception of stress by French, Spanish, and bilingual subjects, *in* 'The seventh European Conference on Speech Communication and Technology', Aalborg, Denmark.

Peters, J. (2002), 'The role of intonation in emotional expressions', *Linguistische Berichte* **189**, 27–57.

Pfitzinger, H. R. (2006), Five dimensions of prosody: Intensity, intonation, timing, voice quality, and degree of reduction, *in* 'Proceedings of the third International Conference Speech Prosody 2006', Dresden, Germany.

Piaget, J. (1968), *Six Psychological Studies*, Vintage Books, New York.

Pierrehumbert, J. (1980), The phonology and phonetics of English intonation, PhD thesis, MIT.

Piske, T., Flege, J. E., MacKay, I. R. A. & Meador, D. (2002), 'The production of english vowels by fluent early and late italian-english bilinguals', *Phonetica* **59**(1), 49–71.

Plutchik, R. (1993), Emotions and their vicissitudes: Emotions and psychopathology, *in* M. Lewis & J. M. Haviland, eds, 'Handbook of Emotions', The Guilford Press, New York.

Porter, R. E. & Samovar, L. A. (1998), Cultural influences on emotional expression: Implications for intercultural communication, *in* P. A. Andersen & L. K. Guerrero, eds, 'Handbook of Communication and Emotion: Research, Theory, Applications and Contexts', Academic press.

Ramus, F., Nespor, M. & Mehler, J. C. (1999), 'Correlates of linguistic rhythm in the speech signal', *Cognition* **73**, 265–292.

Ratner, C. (2004), Vygotsky's conception of psychological development, *in* R. Rieber & D. Robinson, eds, 'The Essential Vygotsky', Kluwer Academic/Plenum Publishers, New York, pp. 401–413.

Redlinger, W. & Park, T. (1980), 'Language mixing in young bilinguals', *Journal of Child Language* **7**, 337–352.

Richardson, K. (1998), *Models of Cognitive Development*, Psychology Press, London, UK.

Rose, P. (1987), 'Considerations in the normalization of the fundamental frequency of linguistic tone', *Speech Communication* **6**, 343–351.

Rothman, A. D. & Nowicki, S. (2004), 'A measure of the ability to identify emotion in children's tone of voice', *Journal of Nonverbal Behavior* **28**, 67–92.

Russell, J. A. (1991), 'Culture and categorization of emotions', *Psychological Bulletin* **110**, 426–450.

Saarni, C. (1984), 'An observational study of children's attemps to monitor their expressive behavior', *Child Development* **53**.

Schachter, S. & Singer, J. (1962), 'Cognitive, social and physiological determinant of emotional state', *Psychological Review* **69**.

Scherer, K. R. (1986), 'Vocal affect expression: A review and a model for future research', *Psychological Bulletin* **99**, 143Ű–165.

Scherer, K. R. (1993), 'Neuroscience projections to current debates in emotion psuchology', *Cognition and Emotion* **7**.

Scherer, K. R. (2000), Vocal emotions, *in* M. Lewis & J. M. Haviland-Jones, eds, 'Handbook of Emotions', second edn, The Guilford Press, New York.

Scherer, K. R. (2003), 'Vocal communication of emotions: A review of research paradigms', *Speech Communication* **40**.

Scherer, K. R., Banse, R. & Wallbott, H. G. (2001), 'Emotion inferences from vocal expression correlate across languages and cultures', *Journal of Cross-Cultural Psychology* **32**(1).

Scherer, K. R. & Ceschi, G. (2000), 'Studying affective communication in the airport: The case of lost baggage claims', *Personality and Social Psychology Bulletin* **26**(3), 327–339.

Schröder, M. (2004), Speech and emotion research: An overview of research frameworks and a dimensional approach to emotional speech synthesis, PhD thesis, University of Saarlandes.

Shochi, T., Aubergé, V. & Rilliard, A. (2006), How prosodic attitudes can be false friends: Japanese vs. French social affects, *in* 'Proceedings of the third International Conference Speech Prosody 2006', Dresden, Germany.

Shweder, R. A. (1993), The cultural psychology of the emotions, *in* M. Lewis & J. M. Haviland, eds, 'Handbook of Emotions', The Guilford Press, New York.

Slaney, M. & McRoberts, G. (1998), Baby ears: A recognition system for affective vocalizations, *in* 'The International Conference on Acoustics, Speech, and Signal Processing', Seattle, WA, USA.

Smiljanić, R. & Hualde, J. I. (2000), Lexical and pragmatic functions of tonal alignments in two Serbo-Croatian dialects, *in* 'Proceedings from the Main Session of the 36th Regional Meeting of the Chicago Linguistic Society', pp. 469–482.

Staats, A. W. & Eifert, G. (1990), 'The paradigmatic behaviorism theory of emotions: Basis for unification', *Clinical Psychology Review* **10**.

Stibbard, R. (2000), Automated extraction of ToBI annotation data from the Reading/Leeds emotional speech corpus, *in* R. Cowie, E. Douglas-Cowie & M. Schröder, eds, 'Proceedings of the ISCA Workshop on Speech and Emotion', Belfast, Ireland.

Stibbard, R. M. (2001), Vocal expression of emotions in non-laboratory speech: An investigation of the Reading/Leeds emotion in speech project annotation data, PhD thesis, University of Reading.

Strongman, K. T. (1996), *The psychology of emotion: theories of emotion in perspective*, John Wiley and Sons, Chichester.

Sundara, M. & Polka, L. (2007), 'Discrimination of coronal stops by bilingual adults: The timing and nature of language interaction', *Cognition* **doi:10.1016/j.cognition.2007.01.011**.

Sundara, M., Polka, L. & Genesee, F. (2006), 'Language-experience facilitates discrimination of /d- ð/ in monolingual and bilingual acquisition of English', *Cognition* **100**, 369–388.

't Hart, J., Collier, R. & Cohen, A. (1990), *A perceptual study of intonation*, Cambridge University Press.

Tomioka, N. (2002), A bilingual language production model, *in* 'Proceedings of the International Conference on the Multimodality of Human Communication: theories, problems and applications', University of Toronto, Victoria College.

Tsai, J. L. & Chentsova-Dutton, Y. (2003), 'Variation among european americans in emotional facial expression', *Journal of Cross-Cultural Psychology* **34**, 650Ű–657.

Vihman, M. (1985), 'Language differentiation by the bilingual infant', *Journal of Child Language* **12**, 297–324.

Volterra, V. & Taeschner, T. (1978), 'The acquisition and development of language by bilingual children', *Journal of Child Language* **5**, 311–326.

Vygotsky, L. S. (1978), *Mind in Society: The Development of Higher Psychological Processes*, Harvard University Press, Cambridge, MA.

Vygotsky, L. S. (1981), The genesis of higher mental functions, *in* J. V. Wertsch, ed., 'The concept of activity in Soviet psychology', Sharp, Armonk, p. 163.

Wadsworth, B. J. (2004), *Piaget's Theory of Cognitive and Affective Development/Foundations of Constructivism*, 5th edn, Allyn and Bacon, New York.

Weinreich, U. (1953), *Languages in Contact*, Mouton, the Hague.

Whitworth, N. (2003), Bilingual acquisition of speech timing. Aspects of rhythm production by German-English families, PhD thesis, University of Leeds.

Wierzbicka, A. (1992), 'Talking about emotions: Semantics, culture, and cognition', *Cognition and Emotion* **6**.

Wierzbicka, A. (1999), *Emotions across languages and cultures: Diversity and universals*, Cambridge University Press, New York.

Wilson, I. L. (2006), Articulatory settings of French and English monolingual and bilingual speakers, PhD thesis, University of British Columbia.

Yanushevskaya, I., Gobl, C. & Chasaide, A. N. (2006), Mapping voice to affect: Japanese listeners, *in* 'Proceedings of the third International Conference Speech Prosody 2006', Dresden, Germany.

Zetterholm, E. (2003), Voice imitation: A phonetic study of perceptual illusions and acoustic success, PhD thesis, University of Lund.