ACOUSTIC AND PERCEPTUAL GENDER CHARACTERISTICS IN THE VOICES OF PRE-ADOLESCENT CHILDREN

Cordula Klein

Magisterarbeit – Mai 2004

Institut für Phonetik

Universität des Saarlandes

Zusammenfassung

In der vorliegenden Arbeit wurde untersucht, inwiefern sich die Stimmen deutscher Kinder im Kindergartenalter bezüglich akustischer und perzeptorischer Merkmale hinsichtlich ihres Geschlechts und Alters unterscheiden.

Ausgehend von den Annahme, daß sich Kinder in präpubertärem Alter hinsichtlich körperlicher Maße wie Größe und Gewicht bezogen auf ihr Geschlecht nicht voneinander unterscheiden (vergl. Fitch and Giedd 1999), wurde versucht zu belegen, daß sich auch im Stimmapparat, insbesondere im Kehlkopf als Klanggenerator und somit in der erzeugten Grundfrequenz keine Unterschiede feststellen lassen. Hierzu wurden die Stimmen von siebzehn Kinder im Alter von 3;3 bis 5;10 Jahren in zwei verschiedenen Modi (picture naming task und spontansprachliches Material in Form der Nacherzählung einer Geschichte) auf digitaler Audiokassette aufgenommen. Von diesen Kindern wurden zehn, je fünf 3–4 jährige und fünf 5-jährige, von denen wiederrum fünf männlich und fünf weiblich waren, akustisch bezüglich ihrer Grundfrequenz (F0) und der ersten zwei Formanten (F1 und F2) untersucht. In einem zweiten Schritt wurden zwölf Erwachsenen (sechs Männer und sechs Frauen) einzelne Vokale, Wörter und kurze spontansprachliche Passagen der Kinder in einem Perzeptionstest dargeboten, in dem die Hörer entscheiden sollten, ob es sich bei einem auditiven Stimulus um die Stimme eines Mädchens oder die eines Jungen handelte.

Bezüglich des Geschlechts ließen sich die Annahmen bestätigen, daß Kinder in dem hier untersuchten Alter keine deutlichen Unterschiede im Hinblick auf Körpergröße und -gewicht oder gemessene Grundfrequenz aufweisen. Trennt man die Gruppe hingegen nach Alter in jüngere 3- bis 4-jährige und ältere 5-jährige, zeigt sich eine stärkere Diskrepanz in den Messwerten. In den Formantwerten ließ sich sowohl ein geschlechts- wie auch ein altersspezifischer Trend nachweisen. Während der erste Formant bei Mädchen im Schnitt und besonders

für den Vokal [a] höher liegt als bei den Jungen, stellt sich bezogen auf das Alter ein Trend zur Abnahme des ersten Formantwertes dar. Die Werte für den zweiten Formanten werden hingegen für [o] und [u] angehoben und für [i] abgesenkt, was eine zunehmende Zentralisierung der Vokale impliziert.

Die Ergebnisse des Perzeptionstests belegen, daß deutsche Erwachsene Kinderstimmen über Zufallsniveau dem richtigen Geschlecht zuordnen können, und das bereits bei Kindern im Alter von 3 Jahren. Die Kinder der älteren Gruppe wurden generell besser erkannt, zwischen Jungen und Mädchen gab es nur dann Unterschiede in der korrekten Erkennensleistung der Erwachsenen, wenn sie auch nach Alter getrennt wurden. Die Jungen der älteren Altersgruppe wurden besser erkannt als beide Mädchengruppen, die Gruppe der kleinen Jungen bildete das Ende der Skala. Die Art des Stimulus hatte ebenfalls einen Einfluß auf die Erkennensrate, insofern als daß längere Äußerungen wie Wörter und spontansprachliche Passagen besser dem richtigen Geschlecht zugeordnet wurden als isolierte Vokale. Die Qualität der Vokale wies einen deutlichen Einfluß auf, betrachtet man die beiden Geschlechter separat. Während der [a] Vokal der Mädchen besonders hohe korrekte Erkennensraten erzielte, wurde dieser Vokal von Jungen gesprochen nur äußerst selten dem richtigen Geschlecht zugeordnet. Vorläufige Messungen der Stimmqualität (z.B. Irregularitäten im periodischen Bereich) und Kommentare der Teilnehmer des Perzeptionsexperiments lassen den Schluss zu, daß einige der Kinder, deren Zuordnung zum richtigen Geschlecht den Hörern Schwierigkeiten bereitete, stimmliche Merkmale aufwiesen, die üblicherweise dem anderen Geschlecht nachgesagt werden (ein Junge hatte beispielsweise eine stark behauchte Stimme, ein Mädchen deutliche Anzeichen von *creaky voice*).

Inwiefern anatomische Gegebenheiten oder Aspekte der Sozialisation eine maßgebliche Rolle spielen, kann und soll unter anderem aufgrund der geringen Datenmenge, der rein äußerlichen Messmethoden und vielfältigen Einflussfaktoren nicht abschließend geklärt werden. Die Arbeit stellt aber eine Er-

weiterung der vorhanden Literatur in Bezug auf das Deutsche und die Untersuchung vergleichsweise junger Kinder dar. Desweiteren wurde eine Differenzierung nach Alter unternommen und verschiedene Einflussfaktoren auf die korrekte Erkennensleistung untersucht.

Abstract

The present study examines whether the voices of kindergarten children differ with respect to acoustic and perceptual parameters, depending on the age and sex of the child.

According to Fitch and Giedd 1999, boys do not differ from same aged girls with respect to e.g. body height and weight until puberty. This leads to the assumption that the vocal apparatus, in particular the larynx, should also not show any sex specific variation in anatomy during that period and further, no differences in fundamental frequency between the sexes are to be expected. To find out whether this hypothesis could be confirmed for German, audio recordings of 17 children between the ages of 3;3 and 5;10 were made via a picture naming task as well as the retelling of a story from a picture book. In the first part of the study, ten children (five of them were 3–4 years and five 5 years old; five were girls and five boys) were analysed acoustically with respect to their fundamental frequency (F0) and their first two vowel formants (F1 and F2). In a following perception experiment, twelve adults (six women and six men) listened to the children's voices in form of isolated vowels, single words and passages of spontaneous speech and were asked to identify the sex of the speaker.

Results confirmed that the children did not display any sex dependent differences with regard to body height, weight or F0. When comparing the two age groups, however, clear differences emerged, with the older children being on average taller and heavier, as well as speaking with a lower F0. The formant values

on the other hand showed sex as well as age dependent differences. While F1 (in particular for the vowel [a]) was higher in girls than in boys, the older children (boys and girls alike) displayed lower F1 values than the younger children. The values for the second formant on the other hand, were raised in [o] and [u] but lowered in [i], which implies an increasing centralisation of vowels with age.

The results of the perception experiment revealed that German adults were able to assign the correct sex to a prepubertal speaker above chance level with the older children receiving overall higher correct identification rates (CIRs). A difference between the sexes was only found if boys and girls were gouped according to age. The older boys were correctly identified most often, followed by the two groups of girls. The 3–4-year-old boys were mistaken for the opposite sex more often than any other group. The type of stimulus also had an influence on the identification rate. Generally, the longer material (words and spontaneous speech) received higher rates of correct identification of speaker sex than isolated vowels. When looking at both sexes together, vowel quality seemed not to have played a role in correct identification as all vowel stimuli received similar CIRs. When examining boys and girls seperately, however, it was revealed that [a] is the vowel with the highest correct identification rate in female stimuli, but the vowel with the lowest CIRs among the male stimuli. Drawing on tentative measurements other than F0 - F2 and judging from comments made by the participants of the perception experiment, aspects of voice quality such as breathyness, harshness or creak, as well as variations in intensity levels and subglottal effort might explain some of the more peculiar findings.

It is not the scope of this work to answer the question whether the differences between boys and girls, older and younger children found in this study are due to anatomy, function or socialisation. This would require not only more but also more controlled data, and possibly also invasive methods of measurement. This work, however, presents an addition to the literature in the field, as the language under study was German; the 3-year-old children were among the youngest ones

studied so far in comparable studies; the data was differentiated according to age group; and several factors such as length of stimulus and sex of listener were controlled for and taken into account.

Acknowledgements

Several people deserve my warmest thanks for having accompanied me and my thesis over several months.

First, I would like to thank my supervisor, Prof. William J. Barry, for giving me plenty of rope in the pursuit of my topic and for supporting my progress at all times. I am also grateful to Prof. Dr. Martine Grice for acting as my second examiner.

Next, I would like to express my gratitude to everyone at the Evangelischer Kindergarten Rotenbühl. The director, Mrs. Rhode, for her interest in my project; all kindergarten teachers for their assistance; all parents for unhesitatingly signing the form of consent; and last but not least the boys and girls who provided me (most of the time eagerly and voluntarily) with an interesting and abundant sample of their speech. I am also very grateful to the participants of my perception experiment for supplying their time and energy.

I will always be indebted to Jürgen Trouvain who organised the contact with the kindergarten and helped me to delimit and reassess the workload. Markus Guhe is and was my moral and TeX-support. I thank him for perseverance, proofreading – well, just about everything. I am also obliged to my parents, Margreth and Josef Klein, for their financial support throughout my years at university. A large part I owe to Caren Brinckmann and Martin Kempkes, my task force when it came to structure, technical and moral support, as well as short scripts. They contributed time and expertise especially in the last weeks. I deeply appreciate the help I received from Anne Fürstenberg, Stephanie Köser and Gudrun Schuchmann on questions relating to statistics. Eva Schmitt, Stefan Baumann, Markus Guhe and Jürgen Trouvain were invaluable when it came to proofreading intermediate or final versions of this work. Thanks also to Silke Jarmut who did some tedious typing of numbers.

Contents

1 Introduction

> ``Research has revealed [...] the possibility that pre-pubertal boys and girls can be distinguished on the basis of recordings of their speech alone and, that, after taking account of any gender differences in speech content (vocabulary etc.), boys and girls may also have characteristically different voices in terms of differences in glottal source, vocal tract filter and other parameters''. (Nairn, 1997, 11--12)

It is generally acknowledged that listeners can correctly assign the sex of male and female adult speakers with as some say close to 100% accuracy. This task is assumed to be an easy one as men and women differ strongly with respect to several vocal characteristics, the most prominent of them being fundamental frequency. In general, men tend to have a lower F0 than women, with frequencies centering around 120 Hz and 220 Hz respectively. This clear distinction facilitates the assignment of sex for listeners. This difference in pitch is usually attributed to the anatomical differences of the larynx such as different size, weight and shape between men and women that start to develop with the voice break during puberty.

But what if it were possible for listeners to correctly assign sex to children before puberty? Before the age of eight, the makeup of the vocal apparatus does not differ significantly between the sexes (Fitch and Giedd, 1999), therefore one might expect not to find any significant differences in F0. Indeed, the results of the majority of studies examining children from languages such as Dutch, Finnish and several dialects of English suggest that there are no differences in F0 between the sexes of pre-adolescent children within one age group (Günzburger et al. 1987; Sachs et al. 1973; Moore 1995; Nairn 1997). However, they also

found, that listeners were able to assign correct sex to these voices generally well above chance level.

However, if listeners can nonetheless correctly assign the sex of a speaker, other parameters in the speech signal must serve as acoustic cues. Authors have found that formant patterns and their relation to fundamental frequency, as well as their developmental trends differ between boys and girls (Sachs et al., 1973) but possible anatomical differences in the vocal tract such as size of the the mandible could not be validated (Fitch and Giedd, 1999). Yet again others have concentrated on vocal fold contact behaviour (Robb and Simmons, 1990) but failed to locate variations with respect to gender.

Many aspects of child voice and its acoustic parameters have been studied and sometimes contradictory results could be claimed. The differences in methodology employed and age ranges studied account for a large part of these discrepancies. While some researchers have compared subjects from groups with very narrow age ranges, others have pooled data from children spanning an age range of up to ten years (Sachs et al., 1973). With respect to methodology, some researchers have investigated spontaneous speech (Weinberg and Bennett, 1971), others used read speech Montenegro 2003. While some included developmental changes across age groups, other focused on both children and listeners from different language backgrounds (Karlsson, 1987). Rates of correct identification were also found to vary, depending on the length of the stimulus material (e.g. Günzburger et al. 1987).

The present study aims to replicate some of the findings of these earlier studies, using German speaking pre-pubertal children as speakers and German adult men and women as listeners. Instead of limiting the study to a single aspect, it was attempted to cover a wide range of interesting research questions.

To this end, ten children were used as speakers, a group which was evenly split in boys and girls, as well as two age groups to examine whether any differ-

ences in anatomy as well as speech parameters that might not e.g. surface with respect to sex, possibly do so with respect to age. The first age group comprised five 3- to 4-year-olds, the second group included only 5-year-olds. The children were measured according to height and weight, as no differences in these parameters suggests that vocal tract anatomy does not differ either (Fitch and Giedd, 1999).

Using a picture naming task and a story book, single words and passages of spontaneous speech were elicited with the respective methods. Since read speech was not an option given the age of the subjects, these two methods were deemed appropriate as they controlled for sex specific vocabulary. They were also favoured as they were expected to yield the most natural language output. Parameters that were extracted from the speech of the children, measured and statistically analysed were mean speaking fundamental frequency, extracted from the spontaneous speech passage and the first two formants of the vowels [i], [a], [o] and [u]. Measurements for the latter were obtained from vowels contained in the word list of the picture naming task.

In a subsequent listening experiment, adult males and females listened to single vowels, words and passages of spontaneous speech and then indicated on an answer sheet whether the stimulus they had previously heard was produced by a boy or a girl. Mean correct identification rate was calculated and descriptive statistics were used to examine the differences with respect to this rate depending on sex and age of the speaker, i.e. the child, sex of the listener, length of the stimulus and vowel type.

The body of this thesis consists of three parts. In Chapter 2, the relevant literature on gender and voice, starting with the differences in adult voice and moving down to kindergarten children will be reviewed, thereby covering anatomical as well as perceptual gender differences. Studies of different aims and scopes will be introduced and the differences in methods and the problems arising from that for the present study will be highlighted.

The second part is twofold and contains the production (Chapter 3) and perception task (Chapter 4) which together constitute the present study. As indicated above, method and analysis of the recordings conducted in the winter of 2003/2004 with ten kindergarten children and the subsequent perception test, where twelve adult listeners were to evaluate the voices of these children will be presented.

The final part, Chapter 5 and 6, will discuss the results of the study in relation to the relevant literature and their implications for future research.

Before starting with the theoretical aspects of gender and voice, the usage of certain conventions as well as the treatment of several sets of quasi-synonyms in the present study need some attention. The synonyms in question are pre-pubertal, prepubescent, pre-adolescent; sex versus gender; and identification versus recognition.

Adolescence is sometimes regarded as the hypernym of puberty (Lippincott et al., 2004), and therefore the more general notion of the transition phase between child- and adulthood, whereas puberty relates to "the period during which the secondary sex characteristics begin to develop and the capability of sexual reproduction is attained" (CancerWEB Project, 2004). Since the present study focuses on the time before puberty, the starting point and not the phase itself is of interest. It therefore does not matter that adolescence according to some terminology covers a longer period in pre-adult life, if the following definition of adolescence is used: "the time period between the beginning of puberty and adulthood" (Webnox Corporation, 2004). These three terms will therefore be used synonymously throughout this work.

The terms *sex* and *gender* will also be treated as synonyms. Although the majority of sources distinguishes sex as relating to biology (structure, function and behaviour) from gender as a social or cultural concept, it would be beyond the scope of this work to discuss the speech of 3- to 5-year-olds in the light of

the nature versus nurture debate. While the term *sex* is agreed to refer primarily to biological aspects, gender is more fuzzy. When referring to *gender* only in the sense of 'behavioural, cultural, or psychological traits typically associated with one sex" (Merriam-Webster, 2004), this will be clearly marked in the text.

Some dictionaries regard the two as synonyms anyway. Oxford Advanced Learners Dictionary e.g. defines *sex* as "the state of being male or female" and *gender* as "the condition of being male or female", as if trying to differentiate the two. However, *condition* is defined as "the present state of a thing" and *state* "the condition in which a person or thing is"(Crowther, 1995). It was therefore considered reasonable to use both terms synonymously.

Similarly handled is the use of the terms *identification* and *recognition*. Throughout the text, they are usually used in the context of listeners' ability to correctly assign speaker sex to an auditory stimulus. Since recognition has the connotation of identifying again (Crowther, 1995), correct identification rate and its abbreviation CIR are used most of the time. However, correct recognition, when used, denotes the same.

One of the conventions used here is that for indicating the age of children. As it is interesting to know the exact age in months, especially for younger children, age is written in the form of "years;months".

Finally, the phonetic symbols used throughout this work are in computer readable Speech Assessment Methods Phonetic Alphabet (SAMPA) format (ESPRIT project 1541, 1987).

2 Theory

2.1 Adulthood

Studies have shown that listeners can correctly identify the sex of a grown-up person by listening to a short sample of speech alone (Lass et al., 1976). The main cue to correct speaker-sex identification appears to be the fundamental frequency of the spoken utterance, which is generally high in females – the literature proposes a mean of 220 Hz – and low – about 120 Hz – in males. Coleman (1976) e.g. recorded forty male and female adults and played five-second samples of their read speech backwards to a group of adult listeners. Since as many supra-segmental differences that might exist between the sexes were eliminated due to reversing the material,

> a correlation coefficient of .94 represents an almost perfect one-to-one correspondence between [...] how male or female sounding a person's voice was judged to be and the frequency of his [or her] laryngeal fundamental (p. 174)

Vocal tract resonances seem to exert a much smaller influence on the maleness-femaleness rating, as implied by a decidedly smaller .59 correlation. Lass also reports decidedly lower correct identification rates for whispered vowels (75%) –which by their very definition lack any information about fundamental frequency except that encapsulated in the harmonics– than for phonated ones (96%).

The difference in adult F0 stems in part from the differences between the sexes in laryngeal anatomy. The following section will therefore give an overview over the make-up and function of the voice-source generator. For a com-

prehensive description of the anatomy, the reader is referred to standard medical textbooks. Here, only those parts of the laryngeal structure are examined that are important in terms of understanding adult gender differences.

Since the most prominent difference in the voices of the sexes is caused by laryngeal activity, possible sub-glottal differences such as pulmonary activity, chest and abdominal muscle tension that might have an impact, are not considered here.

Before addressing the anatomical prerequisites of the male and female voice, it is worthwhile to consider the group of adults all the generalised statements made throughout this work and particularly in this section apply to. These statements relate to the voices of healthy grown-ups. For women, the age range during which the voice fundamental is roughly stable comprises the time between 16 years and the onset of menopause (generally between 40 and 58 years according to the North American Menopause Society 2003), for men that between 20 and 55 years. During menopause, female F0 is known to start decreasing up to the age of 70, whereas the male voice takes the opposite direction from about 50 years on and increases slightly (Traunmüller and Eriksson, 1993). In addition to the afore mentioned authors suggestion that the female voice decreases about 15 Hz during menopause, a study on centenarian women suggests that the voice might lower up to another 35 Hz later in life (Awan and Mueller, 1992). As most of the literature in the field was gathered in Western societies, some caution has to be exerted.

Many factors can influence the voice of the so called healthy adult. Among them are emotions, lifestyle, short term medication or menstruation, to name just a few (Kiesler, 2004). It is e.g. commonly agreed that smoking has a strong influence in that it induces a lowering of F0 in both sexes as well as initiating an earlier start of menopause in women.

A very important factor of course are cultural differences. Yamazawa and

Hollien (1992) report that the mean speaking fundamental frequency of the American women in their study was significantly lower (205 Hz) than that of a group of same aged Japanese women (223 Hz). These assumptions, however, could not always be replicated. For example, Bezooijen (1995) reports on several studies which also found these differences between Japanese and Caucasian women, but she failed to find these differences in her own sample of Japanese and Dutch women's speech. However, she proposes that the high education of her Japanese informants might have induced the low fundamental frequency. Similarly, Henton (1995) reports differences between men and women with respect to the use of vowel space in that women generally had larger vowel spaces, particularly with respect to jaw opening reflected in more extreme values of their first formant. However, the pattern across cultures was similar, in that all women exhibited larger vowel spaces than men. (Compare Section 2.1.1.1 on voice qualities in different societies.)

2.1.1 Anatomy and physiology of the larynx

In both sexes, "the larynx forms the upper end of the windpipe, extending from the trachea to the pharynx" (Nairn, 1997, 15). The three most important cartilages constituting the larynx are the cricoid, the pair of arytenoids and the thyroid. Although all display significant sex differences, the latter of the three, being the biggest, is the one most prone to sex-dependent differences, and therefore the one studied in greatest detail. Maue and Dickson (as cited in Nairn 1997) write that the male thyroid cartilage is with 44 mm and 8 g on average 6 mm higher than that of a female and twice as heavy. Measured from the posterior edges of the cornua's base to the anterior notch where the two laminae join, the male larynx stretches about 20% longer. This anterior protrusion of the *Adam's apple* is the result of a 90° angle, formed by the two plates of the thyroid cartilage. In women

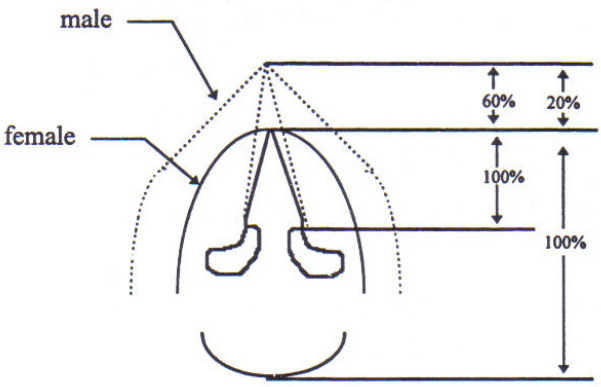

Figure 2.1: Differences between men and women in the length of vocal folds and overall larynx diameter. (Kahane (1978), as cited in Nairn (1997))

this angle is less acute — about 120 ° (Borden et al., 1994). Compare figure 2.1 for the difference between men and women with respect to larynx size.

Due to the bigger overall size of the larynx and the projection of the thyroid, the vocal folds of men and women are non-identical as well. Pompino-Marschall (1995) writes with regard to the connection between vocal fold length and fundamental frequency:

> Bei Frauen, deren Stimmlippenlänge zwischen 13 und 17 mm variiert, liegt die mittlere Grundfrequenz bei ca. 230 Hz [...], wohingegen sie bei männlichen Sprechern mit längeren Stimmlippen zwischen 17 und 24 mm im Mittel 120 Hz beträgt.'' (p. 35)[1]

Clark and Yallop (1995) and Nairn (1997) assume slightly smaller upper values for both sexes.

Since after puberty the vocal folds in boys have not only stretched in length

[1] The quote translates as follows: "Women, whose vocal fold length varies between 13 and 17 mm, have a mean fundamental frequency of about 230 Hz [...], whereas men, with longer vocal folds of about 17 to 24 mm, exhibit a mean F0 of 120 Hz."

but also grown in size, adult males have bigger vocal folds than adult females. Due to a larger amount of tissue in the larynx, the male vocal folds are larger in cross-section, and the walls of their larynges are thicker (Nairn, 1997).

2.1.1.1 Laryngeal voice quality

The anatomical difference as well as a difference in exploiting the functionality of the vocal folds lead to two main voice qualities that women and men use to varying degrees and ends and that are also quite common in children, namely creaky and breathy voice. The following definitions are taken from Laver (1980).

Creaky voice, which is found in both sexes, is a sort of irregular vocal fold behaviour, of which the exact mode of vibration has not yet been firmly established. It is assumed to develop from the fact that only the anterior part towards the thyroid of the vocal folds is vibrating at a very low rate of 24–52 Hz in men. Pitch usually varies strongly from period to period.

Since ventricular and true vocal folds form a thick mass and the pulmonary airstream is rather weak, the movement of the folds is characteristically dampened. This damping is an important perceptual cue: if the amplitude between two periods differs by 42–44 dB, the signal is perceived as creak (Coleman, 1963 as cited in Laver 1980).
The glottal waveform shows short periods, separated by closure phases of different duration, suitably referred to as “train of pulses”. Importantly, the phases of no excitation are longer than those found in modal voice.

Probably due to its low pitch, this voice quality is primarily associated with masculinity. However, it is found in female speech and there usually in utterance-final position (Henton, 1999). It is generally expressed that in English and German, creak is exclusively used para- and extralinguistically, signaling e.g. bored

resignation, initiating turn-taking or serving as a personal marker. According to Laver (1994), in the speech of Copenhagen Danish it may “function as a social marker of upper-class speech” (p.196). Depending on definition, creak can be a form of glottalisation or laryngealisation and may substitute or reinforce glottal stops or oral plosives.

Breathy voice, on the other hand is a phonation type well understood and is generally associated with female speech. Henton and Bladon (1985) found that female British speakers spoke with significantly breathier voices than males. Breathiness accompanies modal voice and is generated by weak adduction of the vocal folds. Muscle tension here is rather weak in general, allowing for a fast airstream flowing through the glottis, thereby creating a slight audible friction noise, which is weaker than in whispery voice.

Fundamental frequency is usually rather low in breathy voice, probably due to the relaxed muscles and, vice versa, a lax voice usually exhibits a breathy component.

Perceptually, breathy voice is associated with intimacy and (sexual) arousal, as well as relaxation in adults of western civilisations, whereas whispery voice is usually related to secrecy. This voice quality amongst other phonatory aspects was examined by Nairn (1997), who found that the children of his study, both boys and girls, tended to have breathiness values close to those of the women studied by Henton and Bladon (1985).

2.1.1.2 Pitch and pitch range

A man can without difficulty speak on a pitch comparable to that of a woman. Equally the pitch range of a male voice can easily overlap with that of a woman, depending e.g. on context or emotional state. Neppert and Pétursson (1986) list

a frequency range of 90–220 Hz for male pitch and 180–450 Hz for that of a female.

It remains to be studied in what way muscles and muscular activity in and around the larynx might differ in the two sexes and to what extent this might influence the perceptual output.

2.1.2 Anatomy and physiology of the supralaryngeal cavity

If the larynx generates the sound, the supralaryngeal cavity can be said to modify it. Since correct gender identification is still well above chance level even in the absence of fundamental frequency (Lass et al. 1976; Schwartz and Rine (1968)), this part of the vocal tract also plays an important part for the present subject matter.

Since men tend to be taller than women (Diverse Populations Collaborative Group, 2004), their vocaltracts are usually longer. The mean vocal tract length of the adult male is estimated at 17.5 cm, that of a woman about 15% shorter, yielding a total length of just under 15 cm (Neppert and Pétursson, 1986). This influences the vocal tract transfer functions in that the female formant values are about 1.2 times higher than those of the males.

However, Nairn (1997) points out that "there are differences other than size between the vocal tract anatomies of men and women". Whereas the oral cavity in females is only about 1.3 cm shorter compared to men, "the pharynx of females is approximately 2.3 cm shorter than the male measurement". This would amount to a much larger overall size difference (3.6 cm) as opposed to the one reported by Neppert and Pétursson (1986), which was about 2.7 cm. An explanation might be that the data cited by Nairn stems from the results of a study on Russian, conducted by Fant in 1966. Data collected over thirty years later shows yet again

a different picture. Fitch and Giedd (1999) used magnetic resonance imaging to measure the vocal tracts of 129 persons between 2.8 and 25 years. The statistic analysis conducted "at each age revealed no significant sex differences before age 15 years" with respect to body weight and height as well as vocal tract length. Their results show that "postpubertal subjects [n=33, age 14.7–25.1 years, mean 17.9 years] showed a highly significant sex difference, with male [vocal tract length] averaging 12.9 mm longer than females".

These differences in findings are explained by Fitch and Giedd with differences in methodological approaches. Firstly, Fant's subjects were recorded sitting upright, while Fitch and Giedd's participants were lying down. Secondly, Fant's measurement started at the bottom of the larynx cavity which Fitch and Giedd assume to mean "an origin at the inferior margin of the cricoid". This would at least account for a difference of 1 cm between their measurements and those of Fant. Thirdly, Fant did x-ray vocal tracts in different vowel positions and not during quiet respiration. Some caution is therefore necessary when comparing measurements of vocal tract lengths.

As intriguing and evident as all these anatomical differences are, "the configuration of the vocal tract [. . .] depends on two factors – the organic makeup and the precise articulations of the speaker" (Nairn, 1997). The articulation can be (and is) frequently and persistently manipulated deliberately. The vocal tract is e.g. lengthened – and formants lowered – through lip protrusion and the opposite effect can be achieved through lip-spreading, as is the case when smiling. Mattingly (1966) correlated vowel formants of children, women and men and reports that "the separation between male and female distributions for some vowel formants is much sharper than variation in individual vocal tract size can reasonably explain".

There are other parameters contributing to the male-femaleness distinction than the ones mentioned so far, e.g. speaking rate or the frequency of glottal stops (Byrd, 1992); an extensive list of which can e.g. be found in Henton (1999).

2.2 Puberty

With regard to chronological age, it is difficult to narrow puberty down to a certain age range. It cannot be predicted at what age an individual will enter puberty, as this can be influenced by factors outside the body, such as certain types of medication, as well as e.g. cerebral dysfunctions due to brain injury during childhood. It is generally assumed, however, that girls enter puberty around 10 to 15 years of age, boys approximately 2 years later. If boys develop secondary sexual characteristics before the age of 10 or do not develop them up to 16 years, their development is considered to be deviant (Bürger, 2002a). The lower limit of abnormal development for girls is 8 years, the upper limit for normal development in the growth of pubic hair and breasts is 13 years, that for menarche 16 years (Bürger, 2002b).

Fitch and Giedd (1999) locate the peri-pubertal stage in the subjects of their study from the ages of 10.3 to 14.54 years, which covers thirty-nine of their subjects; one child age 8.1, as well as some others younger than the lower limit mentioned above, reported a Tanner stage 2 (cf. explanation below) and were classified peri-pubertal accordingly. In Fitch's study, pubertal status "was quantified using a self-administered questionnaire, yielding Tanner ratings of pubertal stage from 1 (pre-pubescent), 2 to 4 (intermediate stages), or 5 (fully mature)". This classification is based on Tanner (1962), who proposed fives stages of development of the secondary sexual characteristics genitals[2], pubic hair growth and breast maturation.

Puberty is a period in which the child experiences massive anatomical and psychological development. Of the two, the anatomical change is more influential on speech and voice, although the psychological aspect should not be overlooked. The *Mutationsfistelstimme* is one example of such a result of a mental influence.

[2]Nowadays, genitals along with the internal reproductive organs are considered primary sexual characteristics.

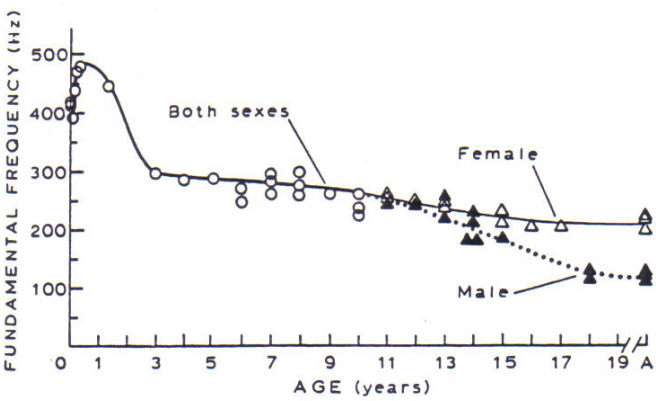

Figure 2.2: Fundamental frequency development according to Kent (1976).

This deviant voice development pertains to boys only and is presumed to have its origin in an abnormal mother-child bond or to bespeak a missing identification with the male social role. Auditorily, this voice remains child-like, high pitched and the vocal fold vibration is irregular (Borjazin, 2003).

The distribution of certain hormones at the beginning of puberty induces the growth of the body in general, entailing strong growth in the vocal apparatus, especially the thyroid cartilage and the vocal folds. Not being able to control the changing anatomy immediately, the sudden growth of the vocal folds in length and thickness and the enlargement of the thyroid results in a perceptual correlate, usually referred to as *voice break*. Although this transformation occurs in both sexes, the changes are more dramatic and therefore more easily perceived in boys than in girls. Whereas the female fundamental frequency only drops by a third, boys experience a lowering of F0 of about an octave during this developmental period (cf. figure 2.2).

As puberty's defining feature is transition, namely that from the small vocal apparatus which presumably does not yet exhibit any sex-dependent diversific-

ation to the large grown-up one which differs markedly between the two sexes, it is not feasible to generalise as to what size and dimensions the pre-pubertal larynx and vocal tract per se have.

Depending on the theoretical approach of what curve the change of the larynx and vocal tract takes in the different sexes, the development is either similar, i.e. linear, in the two sexes with the boys growth-rate being a little higher than that of the girls or the development is quite different, as the quantal theory would predict. For the two approaches, compare Section 2.3 below. In the latter case, the female curve is almost linear during puberty, whereas the male curve only rises slightly up to the age of fifteen and then abruptly accelerates to reach adult dimensions around the age of twenty.

2.3 Childhood

The earliest comparative studies of children's speech date back as far as the 1940s. Since then, the area has been scrutinised from many different angles and a plethora of methods and experimental designs has been employed. Similarly to the two sections on adulthood and puberty, the following section will first present an overview over the anatomical structure of the child's vocal apparatus. Then, the studies which have been conducted in the area and which relate to the present one will be discussed with respect to their research objectives and results obtained. This chapter will close with the hypotheses that have evolved for the present study from the inspiration of the already existing literature.

2.3.1 Anatomy of the larynx and the supralaryngeal cavity

Since "the consensus seems to be that there are no significant structural differences between the vocal organs of males and females prior to puberty" (Nairn, 1997) the following section covers the child in general, from the neonate to puberty, and usually does not distinguish between the sexes. This constraint is somewhat imposed as earlier studies often pooled data for both sexes up to puberty. Where necessary, disparate findings for the two sexes will be presented and discussed. Most of the aspects cited and described here are taken either from Nairn (1997), Borden et al. (1994) or Tanner (1962).

The vocal apparatus of the newborn child is more similar to that of earlier developmental stages of the human being – even closer to that of other mammals – than to that of an adult. Since the larynx is not yet needed for speech production, it is located directly behind the tongue and can be completely sealed off with the epiglottis, which e.g. facilitates the prevention of food diversion into the trachea. Within the first year of life, the larynx descends into the trachea primarily due to the growth of the back of the tongue, whereby a pharyngeal cavity is created. This marks the transition from the primary function of swallowing and breathing to that of producing speech. Researchers have come to different conclusions as to whether the anatomical structure of children differs with respect to sex. Tanner (1962) still stresses the differences in anatomical structure between the sexes, which according to him, are present from birth on and, which are not only maintained but even intensified up to puberty and beyond. However, he also admits that at least in the 1960s "Geschlechtsunterschiede durchweg sehr schlecht belegt sind [...] der Leser sollte sich darüber klar sein, daß dieses Kapitel deshalb viel eher trügerische Angaben enthalten könnte als andere Abschnitte [...] manches [muss] heute noch etwas spekulativ bleiben".[3]

[3]The original quote translates as follows: "gender differences are generally poorly documented [...] the reader should be aware that this chapter might contain more misleading data than other

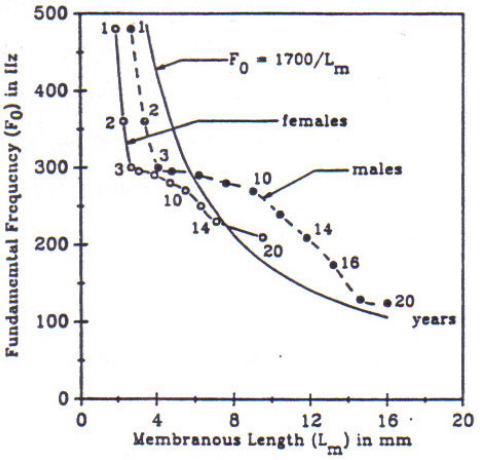

Figure 2.3: The development of fundamental frequency as a function of vocal fold length, according to Titze (1985).

Kent (1976) collected the results of several studies dealing with fundamental frequency development, which he plotted against age of boys and girls (cf. figure 2.2). If F0 development mirrors laryngeal development, one may assume that both sexes undergo the same morphological changes roughly simultaneously up to puberty (in the 1970s this was expected to start around the age of 11 years).

Nowadays, the consensus seems to be that up to three years of age, the larynx continues to grow considerably; afterwards, however, "there is little further development until puberty" (Nairn, 1997). However, some researchers assume a difference between the development of the larynx in general and that of the vocal folds in particular. Especially the length development of the vocal folds is an issue. Two theories are at conflict here: the linear versus the quantal theory of growth.

The linear theory predicts a steady growth-rate of the vocal folds. By this assumption, the folds of females grow .4 mm, those of males .7 mm per year, starting at 3 mm in the neonate and reaching adult length at around 20 years of

sections [...] much speculation remains at this point".

age. The fundamental frequency would therefore be predicted to decline steadily, but slightly faster in males, up to adulthood.

Others have supported the notion of a quantal theory of vocal fold length increase. This theory assumes that the vocal folds undergo growth spurts shortly after birth and again at the beginning of puberty. The first spurt during the first 3 years is identical for both sexes, whereas that during puberty is more pronounced in males and would account nicely for the extensive drop in F0 during that period. Nairn (1997) compares the two theories with respect to their differences in males and females. In both of the sexes, the quantal theory predicts slightly longer folds than the linear theory up to the age of eight. In the female, this curve is steeper than would be predicted by the linear theory and this positive difference remains until the age of fifteen, where the two curves converge. In the male, the quantal theory predicts no growth between the ages of eight and fifteen, followed by rapid growth to twice their length by the age of twenty. The male folds would therefore be shorter between the ages of 8 and 20 than predicted by the linear theory.

However, others have assumed that although the years before puberty do experience a growth in vocal fold length, this might not show an effect on F0 as vocal fold stiffness increases accordingly (Titze, 1985). The author, nonetheless, emphasises that although this stiffness "could counteract the effect of a length increase, [...] this is speculation" (cf. figure 2.3).

To summarise, one may say that "[t]he evidence for or against either the linear or quantal theories is not yet conclusive, however, [...] the quantal theory best seems to fit the available measurements"(Nairn, 1997).

The age group analysed in the present study falls into a range where the curve of growth is roughly linear and the steepness of the curve is very similar for boys and girls. From the data referred to by Nairn, the length of the vocal folds can be estimated at around 5.5 to 7.5 mm.

2.3.1.1 The vocal tract

As has been stated earlier in this chapter, adult men and women differ with respect to the length of their vocal tracts, with men having on average longer vocal tracts than women. This difference cannot yet be found in pre-pubertal children.

Fitch and Giedd (1999) had among their subjects 54 children who were classified as average pre-pubertal. Vocal tract length was measured during quiet respiration and no significant sexual dependent differences were found for vocal tract length in the pre-pubertal age group. In addition to average vocal tract length, Fitch and Giedd also divided the vocal tract up into the five sections pharynx, velum, tongue dorsum, tongue blade and lip and measured each segment separately. There was a significant interaction between sex and age group only for the pharyngeal section, showing that the pharynx section displayed differences amongst boys and girls within the peri- and post-pubertal groups, but no sex-related differences could be found for the pre-pubertal children.

Another significant finding pertains to the male lip segment which is slightly longer than the female one in all age groups.

In the following section, the influence of the voice source generator and the anatomy of the supralaryngeal cavity on fundamental frequency and formant frequencies is examined and the relevant literature in the field of pre-pubertal voices is reviewed.

2.3.2 Effects on fundamental and formant frequencies

As described earlier, vocal tract anatomy appears to be quite similar in pre-pubertal children, yet sex differences can be observed with respect to e.g. vocal fold length or lip size. Whether these differences, however, need to surface in the

acoustic and perceptual output remains to be discussed.

Sachs et al. (1973) indeed found differences in F0 in children before puberty. They examined fourteen boys and twelve girls between the ages of 4 and 14 years. Sustained vowels [i], [a] and [u] were elicited and formed the basis for fundamental frequency extraction. The difference of 25 Hz between boys and girls was significant, yielding the result that girls had lower F0 values than boys.

In addition to F0, the first two formant frequencies were measured as well and it surfaced that girls had on average higher formant frequencies than boys. This combination of low F0 and high formants in girls and high F0 and low formants in boys is assumed to be a major cue for the correct identification of the children's voices by adult listeners. Sachs et al. ruled out that listeners based their judgement on the basis of formants alone, as some girls with values close to those of the boys were still clearly identified as girls. However, given that the age range in this study is quite large (4 to 14 years), and F0 was found to be different for boys and girls, the fundamental might after all have been the most important cue, as it is in the identification of speaker sex in adults. The authors, however, did not evaluate spontaneous speech.

Busby and Plant (1995) found significant differences in F0 between 5 and 7 year old children. They examined five of each sex in the four age groups 5, 7, 9 and 11 years to find out whether F0 and the values of the first three formants decrease with age. One result amongst others was that in the group of 5- and 9-year olds, boys exhibited a significantly higher F0 than girls, a finding that was the reverse in the age group of the 7 year old children. The data in the group of 11-year olds remains inconclusive, but the girls seem to have a slightly higher F0, this, however, not being significantly so.

For the formant values, it was found that in general all formants for all of the investigated eleven monophthongs of Australian English decrease with age in both sexes, the girls generally displaying higher formant values. In particular,

F1 of the open vowels and F2 for all vowels except [U] are higher in girls than in boys.

These findings partly replicate those of Hasek et al. (1980) 15 years earlier, who studied American children, 15 of either sex in the six age groups 5 to 10. They measured F0 of the sustained cardinal 5 vowel [A] and found significant differences between boys and girls from age 7 on. Whereas fundamental frequency in boys decreased from age 6 onwards, girls seemed to exhibit more of an alternating falling-rising pattern, with F0 falling between the ages of 5 to 6, then rising between 6 and 7, etc. The significance at age 7 pertains probably to the fact that whereas boys' F0 falls from 6 to 7 years, girls' F0 rises, yielding 234 Hz and 262 Hz respectively. In addition to that, a significant age effect was found, in that the groups 5 and 6 differed from the older ones in terms of a higher F0, which according to Hasek et al. is primarily due to the sharp drop in F0 which occurs in boys between age 6 and 7.

The work of Günzburger et al. (1987) on the speech of Dutch children was

> ''based on the assumption that pre-pubertal boys and girls do not differ with respect to larynx size, nor in size of the vocal cavities and, therefore, maybe expected to have approximately the same fundamental and formant frequencies''.

They tested eleven boys and six girls between 7;6 and 8;9 years. Although the boys in their study showed higher values in fundamental frequency than the girls, the difference between 246 and 243 Hz was not significant.

Another interesting finding comes from Robb and Simmons (1990) with respect to differences in vocal fold contact behaviour for the vowels [i], [u] and [a]. They studied twenty-six US-American children between 4;4 and 6;6 years and were able to show that the sexes did differ in contact quotients with boys exhibiting higher values and therefore less contact of the vocal folds than girls.

However, this did not significantly influence fundamental frequency, which was 282 Hz for boys and 273 Hz for girls across vowels, but it supports the notion of breathiness in children's voices.

Many other researchers, who did not find any sex differences in F0 in prepubertal children, clearly associate correct identification with formant frequency values.

Weinberg and Bennett (1971) experienced a large overlap in fundamental frequency values, taken from thirty seconds of spontaneous speech in their population of sixty-six 5- and 6-year-old children. Several years later, the same authors examined seventy-three US children between 6;1 and 7;10 years (Bennett and Weinberg, 1979). These children produced isolated phonated and whispered vowels, whose overall correct identification rates with 65% and 66% were not significantly different from each other. With F0 missing in the whisper condition, Bennett and Weinberg came to the conclusion that "average fundamental frequency characteristics did not provide critically relevant sexual information" and that cues about a child's sex are primarily encoded in the formants.

Whiteside conducted several studies with groups of children who were separated by two year intervals, starting at 6 years up to 10 years (cf. Whiteside and Hodgson 1998; Whiteside and Hodgson 1999; Whiteside et al. 2002). The children within each age group did not differ with respect to F0 according to their sex. In the first of the studies mentioned, Whiteside and colleagues also experienced a decrease in F0 for both sexes with age. This general decrease could at first not be replicated by a study conducted one year later, where girls emerged to show different developmental patterns from boys. They were found to have higher formants F1 to F3 than boys, however, the development of formant change was similar for both sexes.

A subset of fifteen of the same children were used for a longitudinal comparison roughly three and a half years later, where a general decrease in F0-F2 was

observed in the same group of children between the earlier measurement and the one in 2002. However, a clear decrease was again only found in the boys, when not only comparing one group at two points in time, but also comparing the three groups at the time of the second measurement. Here, the girls showed a falling pattern from age 9 to age 11 and a rising of F0 from 11 to 13. When compared to a mean F0 of a group of young adults, the F0 of girls as young as 9.5 years was found to be within 1 standard deviation of the female adult mean, suggesting that by then, unlike their male peers, several girls may have reached adult-like fundamental frequency values.

However, Whiteside et al. (2002) stress the individual developmental patterns, as several children showed definite rising of their formants, contrary to the general trend of lowering formants with age.

With a focus on social class differences, Edwards (1997) and Montenegro (2003) compared the speech of working and middle class children. While Edwards examined twenty 10-year-old British pupils, Montenegro conducted a similar study with sixteen 6-year-olds in the Philippines. Both elicited read passages of English and tested correct identification of child sex by adult listeners. In terms of the number of errors made by the judges, the sex of middle class children was generally better identified than that of the working class children. However, this is only half of the truth. Judges made the fewest mistakes on the group of working class boys. Working class girls received the highest error rate, but middle class girls and boys did not exhibit such a strong clustering of errors on either end of the correctness scale.

Overall correct identification rate was 84% in Edwards' study, 69% in that of Montenegro. For the latter, it is important to keep in mind that English was not the subjects' mother tongue.

However, not all researchers share the view that correct sex identification can be primarily attributed to information in the formants. Karlsson (1987) and Nairn

(1997) remain doubtful as to the role of formants. Karlsson examined children aged 3 to 8 years from English, Finnish and Swedish language backgrounds. She did not find correlations between speaker sex and either F0 or formant frequencies.

Nairn looked at the speech of eighty-nine 4.5–5.5 year olds from a Scottish-English language background. These children did not yield any significant sex differences with respect to F0 and of the eighteen formant comparisons, only few were significant. These included F1 and F2 of [a] and F2 of [i].

2.3.3 Other parameters

Apart from F0 and formant frequencies, researchers have also investigated other parameters as possible sources for differences between the sexes and an influence on correct identification of gender in children.

Bennett and Weinberg (1979) additionally looked at the effect of monotonicity on correct sex identification, where subjects spoke sentences in a normal and a monotonous fashion. This proved to be detrimental to the correct identification of girls, who, in the monotone condition were frequently misidentified as boys.

Whiteside and Hodgson (2000) examined voice onset time (VOT) in 7, 9 and 11 year old children. The time lapse between the release of a plosive, through the aspiration phase, up to the beginning of periodicity of the following vowel was measured. The authors showed that, boys displayed greatest phonemic contrast between [p]-[b] and [t]-[d] at age 9 but then developed a reduction in contrast up to the age of 11, while girls exhibited a rather continuous trend towards strong phonemic contrast. Whiteside and Hodgson take these results to account for the strong impact the anatomical changes of puberty – which generally occur after the age of 9 – have on boys voices and the difficulties they have in making ad-

justments on the one hand side, as well as the sociophonetic reality of the male tendency towards reduction and centralisation in the vowel space on the other.

Several authors have examined whether speech rate differs between the sexes. Both Karlsson (1987) and Haselager et al. (1991) report no sex dependent differences among their groups of 3-8 for the first study and 5, 7, 9 and 11 year old subjects of the second study respectively. However, the latter authors state that speech rate generally becomes faster with age in both sexes.

2.3.4 Perceptual identification of pre-pubertal voices

Despite the disparate findings on which parameters might influence gender identification in pre-pubertal voices, correct speaker sex identification rate (CIR) has generally been high (cf. table 2.1). With the exception of Moore (1995), who experienced rates as low as 32%, all other authors cited here report correct identification rates that lie clearly above chance level: around 64% for overall recognition (Günzburger et al., 1987) up to 87% correct identification of male voices in the study of Bennett and Weinberg (1979).

Generally speaking, studies involving – amongst other age groups – very young children around 4 years, tend to have worse identification rates than those with primarily older subjects. The boundaries are blurred in studies such as that of Sachs et al. (1973), where participants spanned the ages 4 to 14, with no evident subgrouping.

Of all studies cited here, boys were usually better identified than girls. Seven cases reported equal or higher correct identification rates for boys, whereas four tests revealed higher rates for girls. However, this does not necessarily mean that listeners clearly identified a boy's voice because of certain parameters. It could as well be the case that whenever listeners were uncertain of the sex of a speaker,

Investigators	Age of children	CIR vowel	sentence	spontaneous	read
Bennett and Weinberg (1979)	6;1-7;10	66%	80%	–	–
Edwards (1997)	10	–	–	–	84%
Günzburger et al. (1987)	7;6-8;9	55%	74%	–	–
Montenegro (2003)	6	–	–	–	69%
Nairn (1997)	4;6-5;6	66%	76%	73%	–
Sachs et al. (1973)	4-14	66%	81%	–	–
Weinberg and Bennett (1971)	5, 6	–	–	74%	–

Table 2.1: Correct identification rates (CIR) for vowels, sentences, spontaneous and read speech.

by default they marked his or her sample as male.

In those studies where females were better recognised, the default answer for both sexes might have been female. Nairn (1997) attends to this problem in that he calculated whether his subjects showed a bias to respond *girl* or *boy*. Indeed, his listeners tended to prefer the answer *female*. An explanation for this might be that Nairn's study included only very young subjects, who were 4.5 to 5.5 years old, and that young voices with generally high fundamental frequencies tend to be associated with the female sex.

Not only are there differences with regard to which sex is better identified, but the ability to judge voices correctly appears to be also dependent on the sex of the listener. Word is generally passed on that women are better judges than men and achieve higher correct identification rates. This is often assumed since women used to stay at home to raise the children and kindergarten as well as elementary school teachers are at least in Western societies by a majority female. This brought about the notion of the *expert listener* (see below). Since many studies used either only female judges or did not make the sex of the listeners explicit, it is at this point difficult to judge the validity of this assumption.

Other authors, e.g., found the opposite to be true for the population they studied. Montenegro (2003) discovered her male listeners to produce better results, committing fewer errors than the female judges.

Some authors also report a sex of speaker and sex of listener interaction where listeners' ability to identify the sex of a given speaker depended on whether that speaker was of the same or opposite sex of that of the listener. This effect is not significant in the study of Nairn (1997), but a tendency is visible due to males' poor performance on recognising boys correctly.

Bennett and Weinberg (1979) circumvented the possibility of a listener-sex influence on correct identification by using only women as judges. Interestingly, male speakers were more often correctly identified than female speakers, yielding cross-sex advantage. Again, Montenegro experienced the opposite, with male judges scoring best (i.e. committing least errors) on the most male-like children, aka working-class boys, whereas female judges arrived at highest correct identification scores for most female-like children, namely middle-class girls.

The notion of one sex producing higher correct results than the other leads to the concept of expert versus lay judges which underlies several other studies (cf. Günzburger et al. 1987; Karlsson 1987; Moore 1995). Possibly inspired by the fact that women as primary caregivers have extensive contact with children and therefore might do better than men, the question arose of whether certain groups of people might show a better judging ability due to either extensive contact with children (e.g. kindergarten teachers) or special auditory abilities (such as blind subjects). It was therefore expected by Moore that kindergarten teachers should score higher than other people when judging the voices of young children, the likes of whom they have daily contact with. These teachers had to judge utterances of 2-8 syllables and utterances longer than 8 syllables. Surprisingly, kindergarten teachers obtained both, the worst and the best results of all participating groups, depending on the length of the utterance. In the condition of less than 8 syllables they achieved the afore mentioned 32% CIR, on the longer ut-

terances 59%. Another group of experts were 9-12 year old children, who were chosen on the assumption that

> children who are sure of their own gender identification, and who are in daily contact with children, would themselves be able to correctly identify voice qualities that differentiate boys from girls (p. 300)

They performed similarly to the kindergarten teachers, reaching only 39% CIR on the less than 8 syllables condition, but 57% when judging longer utterances. One might assume that of the remaining two groups, the Finns, who were judging children speaking the judges' mother tongue would perform better than those speakers of American English, who did not know Finnish. However, the ones with more expertise in the language at hand performed slightly worse.

Moore's findings are mirrored by Karlsson (1987) who had Finnish, Swedish, English and Chinese adult listeners evaluate the voices of English, Finnish and Swedish children. She did not find any "strong tendency for a higher percentage of right answers for children speaking the listener's own language".

Günzburger et al. (1987) chose fourteen visually handicapped pupils between 14.6 and 19.5 years as experts "on the presumption that these people, due to the forcibly developed extended attention for auditory cues, may have above normal skill in auditory tasks" (p. 52). Against expectations, these children performed slightly worse on this task, and the fact that they did recognise female voices better than male ones suggests a bias towards responding *girl*. Günzburger et al. explain the poor results on behalf of their blind subjects, who according to their own testimony are very accurate in judging a person's sex in everyday situations, by assuming that other "hearing skills such as spatial hearing" aid them with the detection of speaker sex in their everyday lives.

In the preceding sections, the possibility of utterance length having an influ-

ence on correct identification rate has been implied, e.g. in the study of Moore (1995). Several authors offered stimuli of differing length to their listeners to evaluate whether longer utterances were better recognised than short ones.

Moore operated on syllable level[4] and revealed that correct identification rate increases when utterances are longer. All listener groups except the Americans who did not speak Finnish, performed better on longer material.

Similarly, both groups of listeners in the experiment of Günzburger et al. (1987) performed better (69.5%) on the sentences than in the vowel condition (57.5%). The divergence between the two conditions was greater for the normally sighted adults (74% over 55%) than for the blind listeners (65% over 60%).

The listeners in the study of Bennett and Weinberg (1979) compared to those of other researchers, achieved the highest correct identification rate on sentences (80%). This rate was also higher than that of the vowel condition (66%).

Nairn (1997), on the other hand, does not fully agree with the assumption that the longer the utterance, the better the identification rate. He examined the identification rate on isolated vowels, sentences and spontaneous passages. Whereas CIR was significantly lower on vowels (66%) than on the other two conditions, he did not find any significant difference between the correct identification rate of sentences (76%) and spontaneous speech (73%). This is surprising, since "the sponaneous speech sample is the most life-like in that it resembles actual communicative output the most closely". Furthermore, in such a "free-choice situation [...] we would expect many of the child's sociolinguistic markers to rise to the surface [of which] gender is a prime exemplar" (p. 134). However, spontaneous speech does not fare better than single sentences. Nairn explains this with the *additional cognitive load* of spontaneous speech. In the vowel and sentence condition, all children produced the exact same samples, which did not differ in content across children. In the spontaneous speech condition, however, only the topic of

[4]Unfortunately, it is not clear whether these syllables formed words or not.

conversation is the same for all children, the actual words used to describe it vary greatly. Nairn assumes that when evaluating spontaneous speech, the listener has to perform the additional task "of normalising or extrapolating the child's gender from continually varying phonetic input". This extra effort might counteract a better recognition rate that could be gained from longer speech signal.

2.3.5 Present study

Based on the literature reviewed so far, the aim of the present study was in part to replicate earlier findings as well as adding some new aspects. The scope was to test, whether results similar to the ones reported in the literature could be obtained for German speakers and listeners. As the studies cited investigated children from several different language background, it was expected to find similar performances in the German population studied here.

As some studies have possibly confounded pre- and peri-pubertal age groups, the starting point was to examine the speech of children before puberty. As Fitch and Giedd (1999) reported peri-pubertal stages in children as young as 8 years, the age for the present study was fixed at below that age. To rule out as much gender specific speech behaviour as possible, which might have developed by the first school years, the group of subjects was limited to children in kindergarten. However, it is naturally understood that even children at such a young age display certain gender specific behaviour. As parents and kindergarten teachers will tell, it is not a stereotype to claim that boys are for example usually louder and wilder than girls.

Based on the findings of Fitch and Giedd (1999), it was assumed that young children of a similar age, if they do not differ considerably with respect to height and weight, can safely be regarded as having approximately the same laryngeal

dimensions. This assumption had to be taken for granted as no other means of measurement were available for the present study. If the larynx indeed was not significantly different in the two sexes, fundamental frequency should also be similar. This does not take into account different sub-glottal energy effects, such as boys speaking louder, possibly yielding a higher fundamental frequency.

As most studies concerned themselves with vowels as the smallest entity of investigation, it was decided to do the same in the present study to be able to examine formant frequencies. It was expected that differences between the sexes might be found here, as was implied by the result of such studies as the one by Sachs et al. (1973), who found that boys exhibited lower formant values than girls. As it is known from the speech of adults that men tend to centralise their vowels whereas women make more use of the available vowel space (Henton, 1995), it would be interesting to see whether this is already evident in young children.

In accordance with the literature, the perceptual relevance of these parameters was to be investigated. It was assumed that even voices of children as young as 3 years can be correctly identified with respect to their sex.

A group of evenly distributed female and male adult listeners was to be gathered to evaluate the voices of the young boys and girls. This was done to reflect a possible difference in correct identification rate depending on the sex of the listener. Another difference in correct identification was proposed for the auditory evaluation if stimuli are of different length. As seen before, most studies used vowels, sentences and passages of spontaneous or read speech.

As the children in this study were quite young compared to those tested by other investigators, it was assumed to be especially difficult to elicit exactly the same sentence from each child, whereas vowels and spontaneous speech was deemed less problematic. To obtain an intermediary stage of utterance duration between vowels and spontaneous speech, roughly comparable to the series of

vowels, sentences and spontaneous speech used in the studies mentioned earlier, words were to be elicited via a picture naming task and presented to the listeners.

Contrary to the results of Nairn (1997), who found neither an increase nor a significant difference between the identification rates of sentences and spontaneous speech, it was expected that words would differ strongly from spontaneous speech but probably not as much from vowels. It was further to be examined whether the vowel quality had an influence on correct gender identification. As the open vowels offer the best opportunity for gender specific modulation of the vowel, it was tentatively formulated, that correct identification would be highest for this vowel.

Since the kindergarten classes comprise children between 3 years up to school age, and since others had found age effects before, a subgrouping was intended, yielding the same amount of 3- to 4- and 5-year-old children in the sample population. All testing was to be carried out therefore with respect to sex, as well as age, where suitable.

Summing up, the following hypotheses were formulated:

1. German boys and girls at the age of 3-5 years do not display sex dependent differences in fundamental frequency.

2. German boys and girls at the age of 3-5 years do display sex dependent differences in F1 and F2.

3. German boys and girls of 3–4 years do not differ with respect to fundamental frequency from boys and girls of 5 years.

4. German boys and girls of 3–4 years do differ with respect to formant frequencies from boys and girls of 5 years.

5. Adult listeners are able to correctly identify (i.e above chance level) the gender of a given child of both age groups.

6. There is no difference in the correct identification of girls and boys.
7. Older children are better identified than younger ones.
8. Listeners exhibit sex-dependent differences in correct identification rates. Women perform better than men.
9. Correct identification increases with increasing length of utterance. For the present study, this means spontaneous speech is best identified, followed by words and vowels.
10. Correct identification varies with vowel type.

3 Production: Data collection and evaluation

The next two chapters contain a description of the present study. In accordance with some of the literature in this area, this study has a production as well as a perception part. The following chapter comprises the procedure of collecting the data, and the measurements and statistical analyses conducted thereon. Later, Chapter 4 describes twelve adult listeners' auditory evaluation of the children's voices.

3.1 Child subjects

Between November 2003 and March 2004, eight boys and fifteen girls between the ages of 3;3 years and 5;10 years were recruited for recordings in a local kindergarten in south-west Germany. It was established in advance that all participants had to have been born in Germany and had to have parents who both spoke German as their first language. This was done to guarantee as much homogeneity as possible amongst the group of participants. In addition, it was ensured that none of the children had any reported hearing or learning difficulties and that none was attending a speech therapist.

However, three children had to be exempted from further investigation after it emerged during the course of the recordings that one boy and one girl exhibited noticeable speech impairments and another girl had a parent who was not a native speaker of German.

The parents' written consent was obtained in advance, in response to a very general explanation of the nature of the recordings. The spontaneous decision to participate in the study or not, however, was left entirely to the children.

3.2 Recordings

3.2.1 Material

Some of the material which was collected did not prove suitable for either acoustic or perceptual analyses. This includes isolated vowels, sentences and a diadochokinetic task, administered to determine possible differences in speaking rate. The problems arising from these tasks and the reasons for abandoning them after the first recording session, will be covered under section 3.2.1.3.

The material which was established to be suitable for F0 and formant measurements, as well as subsequent evaluation by adult listeners were single words, vowels which were isolated from these words, and passages of spontaneous speech.

3.2.1.1 Words

The word list which was compiled consisted of 33 pictures displaying everyday objects, such as a cow, a scarf or a cradle (see figure 3.1).[1] The words were carefully chosen to satisfy several constraints.

Firstly, the four German vowels [i], [a], [o] and [u] had to be covered, as they were to be used in isolation later in the auditory identification task and also served as the basis for formant measurements. *Secondly*, the pictures needed to be easily discernible for 3-year olds from black-and-white computer printouts. As much as the youngest participants needed to be able to recognise the object, it was also crucial to pick those objects, whose names they could actually produce.

[1] The complete list of words and the corresponding pictures can be found in Appendix B of the original thesis.

Figure 3.1: Examples of two pictures used in the picture naming task. “Hut” [hu:t] was one of the monosyllabic words, “Igel” ['i:g@l] a bisyllabic one.

Thirdly, the relevant vowels had to be in stressed position in the word. This was to ensure that the material in all cases would be long enough to yield proper analysis. Monosyllabic words would have been most straightforward, but since not enough of those could be found, the gap was filled with bisyllabic ones. Words longer than two syllables were not considered, as this would have made potential influences from the number of syllables more difficult to control.

It was further ensured that the vowels were surrounded by non-nasal context as nasality is known to influence the first formant by weakening it Neppert and Pétursson, 1986.

As mentioned before, the words of the picture naming task were to serve two purposes. Firstly, they were to be presented as complete words to adult listeners. Secondly, the vowel of the initial stressed syllable was to be isolated and used as a substitute for the missing sustained vowels which most children who participated in the first session in November 2003 failed to produce, and which were subsequently removed from the list of tasks.

The words were considered to constitute an intermediary category between single vowels and spontaneous speech, regarding length of utterance.

3.2.1.2 Spontaneous Speech

As mentioned before, the primary means of eliciting the spontaneous speech material was the children's book "Der Grüffelo".[2] It is a picture book which includes short passages of verse and all children who were to participate in this study were familiar with the story.

This procedure of having children recount a story instead of completely unconstrained conversation was directly copied from Nairn (1997), who explains that

> [T]he resulting speech sample has the advantage of being of a very similar topic across all children but leaves the final choice of wording open to each individual child. (p.65)

3.2.1.3 Other recordings

Initially, three other tasks were intended to be carried out in the present study: the elicitation of isolated vowels; the production of two sentences related to a picture; and a diadochokinetic task, where syllables such as [fa] are produced in rapid repetition.

If one intends to measure mean fundamental frequency, F0 range and formant values, subjects are usually asked to sustain certain vowels on a comfortable pitch. This is common practice with adult listeners and should render values that can easily be measured if the vowel is sustained long and steadily enough. Since several researchers, who worked with children as young as four years, have employed this method in the past Nairn, 1997; Robb and Simmons, 1990; Sachs

[2] The original is an English story book called *The Gruffalo*.

et al., 1973, it seemed reasonable to assume a successful application of it to the population studied here. However, as the first three children of around 3;4 years were recorded and failed to produce the isolated vowels even after being verbally prompted by the experimenter, this task was removed from the list of tasks and it was decided to use the vowels from the words of the picture naming task instead.

The production of complete sentences proved to be an obstacle as well. Again, it was youngest children who had the most trouble with this task. During the spontaneous speech task, some of them only answered in isolated words or short noun phrases and were not able or too shy to produce complete sentences. This task, as it was not expected to be realised by all subjects, was therefore also removed from further analysis.

To determine whether the sexes differ in speaking rate, Haselager et al. (1991) asked their Dutch child subjects to repeat syllables such as *fa, ta, sa, ka* as fast as they could, yielding outputs such as [fafafafafafa]. Haselager et al. fixed these series at six syllables minimum. Although several different approaches to this task were sampled in the present study, not all children could or would produce these syllables quickly and over the required amount of syllables. Since this task could not easily be mastered and the different approaches proved to be rather time consuming – time which was too valuable as the attention span of the younger subjects was very short anyway – it was later decided to refrain from subjecting the participants to this task. Theoretical support for not including the measurement of speech rate comes from Haselager et al. themselves, when they summarise: ”Boys and girls are to be regarded as equivalent in their speech rate performance”. Speech rate does increase with age but the sexes “do not exhibit different developmental patterns as the interaction age x sex is not significant”.

Although this conclusion is drawn from utterances of children between five and eleven, it will be assumed to apply to the age group at hand as well. Nairn (1997) only briefly touches on this subject, simply stating that “sex-dependent cues might be found in the time-domain, e.g. such factors as [. . .] speech rate”,

which leads to the conclusion that either speaking rate is not a statistically relevant indicator of sex or that not much investigation into the aspect of sex dependent speech rate has been administered to pre-adolescent children. For the time being, it is presumed that there exists no difference between boys and girls between three and six years with regard to this aspect. For this reason, administering a diadochokinetic task was ignored.

3.2.2 Data collection procedure

All recordings were conducted during the morning hours, always in the quietest, though not sound treated room in the kindergarten. It was not feasible to take the children out of the kindergarten into a separate building as this would have included dressing the children and having one kindergarten teacher accompany each child every time. The drawbacks in terms of noise in the kindergarten itself were therefore estimated as less severe than those of relocating the children.

The equipment used was an AKG D330BT microphone, as well as an AIWA HHB 1 Pro digital audio tape deck. The data was recorded onto several TDK 120 and 180 EC digital audio tapes and later copied to computer hard disc.

The recording of one child lasted on average about 20 minutes. However, the actual durations of the recordings varied considerably. While some children needed more time to familiarise themselves with the equipment and the experimenter (the author of this study), others were eager to talk and could hardly be stopped. Again others tried to rush through, so that a recording took somewhere between 15 and 40 minutes. Six recording sessions of about 3.5 hours each were necessary to collect the whole data set.

The children were not informed about the nature of the task but the whole procedure was simply integrated into the daily routine of so-called “open play”,

where children were allowed to generally do what they please. They were asked whether they would like to look at some pictures and where then taken individually to the separate room. Some children wanted to participate but felt uncomfortable to be alone with the experimenter. These were accompanied by one of the kindergarten teachers.

The child was seated opposite the microphone at a table and on the corner with the experimenter, so that the latter could at the same time monitor the equipment and keep eye-contact with the child, without the microphone interfering. It was hoped that this way, the child would not be too conscious of it and would talk most naturally.

Each child was measured in height and weight after his or her recording session was finished.

The sessions usually commenced with showing the above mentioned pictures to the children, accompanied with the question “Was ist das?” (What is this?) or a similar one. It proved to be feasible to start the session that way for two reasons. First of all, it was an easy task that all children mastered. A lot of praise could be given, which in most cases encouraged the children to produce a lot of material afterwards. Secondly, the word list was to be presented twice to the children to get repetitions of the same word. It was therefore used once at the beginning of the session and again at the end.

The decision to use one word list twice instead of developing a long one which is presented only once was based on several considerations.

Concerning the developmental stage, it was necessary to ensure that all pictures could be recognised by children between three and five years of age. This, together with the other constraints mentioned above, greatly reduced the number of possible pictures.

The linguistic capabilities of the youngest subjects also had to be considered. Although they might recognise an object on the picture, they might not be able to

name it, especially not by its intended name. As expected, and, as was confirmed, this was e.g. the case with the word *Globus* "globe". A synonym in German of it is *Weltkugel*, which most children – apart from the ones who were altogether unable to name the object – used as their first choice. Even though some of the pictures bore a potential of not being named properly at the first trial, the children could usually name the object when it was circumscribed (in the case of a heap of coal e.g., it was mentioned that one puts these black lumps into the barbecue grill and lights them, the intended word being "Kohle"). Two words however posed repeated problems and therefore most of the children had to be prompted on those by the experimenter whispering those particular words. One was "Tube" (tube) and the other "Wiege" (cradle).

The nature of the elicitation task was another issue. The same material and tasks were to be administered to all children; yet, on the one hand, the younger ones had problems concentrating on a single task for a longer period of time and the older ones on the other hand, appeared to be easily bored by the simplicity of the task. Both problems could be solved by splitting the "word-task" in two parts: the 3-year olds got new energy from focusing their concentration onto a different task after about five minutes and the older children were not yet bored after such a short amount of time.

The recounting of the Gruffalo-story was generally carried out after the first run of the word list. The book lay in front of the child and he or she was animated to tell the story to the experimenter. If a child did not produce enough material (some only uttered single words and then moved on to the next page), he or she was encouraged by the experimenter asking about the pictures and the story which sometimes involved flipping back and forth. If still not enough material could be gained, the children were asked questions unrelated to the story. These might include questions about family members, friends, pets, etc., while topics which in all probability would have produced straightforward sex specific vocabulary, such as "What did you do yesterday?" were avoided.

Some of the children were recorded on two different days. This was due to technical problems in some cases; in others the child wanted to discontinue at some point during the first session, but was eager to participate during another session a few weeks later. This did not pose a problem, as it was ensured that no child crossed from the younger age group into the older one in between the two recording sessions.

These children were usually presented with the story of the Gruffalo during their first session and on the second meeting asked to name the objects. The word list in these cases was also presented twice, but obviously not separated by the Gruffalo-task. Here, the experimenter tried to involve the subjects in a short conversation after having presented the word list once, to insert a gap similar to the procedure with the rest of the children and to distract them. If this, however, seemed to make them uneasy or, if they did not respond, the pictures were simply shuffled and presented to them again in a different order. The topics were the same as the ones used with the other children.

3.3 Challenges in recording small children

When conducting speech recordings with healthy adults, the procedure of the recording session is usually quite straightforward. People who agree to partake usually approach the task with a certain level of seriousness and their cooperation is generally guaranteed. Instructions can be formulated on an adult level of understanding, participants are aware of the scientific nature of the experiment and there is an unspoken consensus on appropriate behaviour.

Without intending to imply the opposite to be true of children, several other considerations are important when recording children, especially very young ones. Some of the challenges that were encountered in the present study will

be presented in this section. They serve to clarify the reasons for preferring e.g. one method of data collection over another and might help those interested in recording children of a similar age with the necessary decisions beforehand.

3.3.1 Microphones

A decision had to be made on what microphone would best be used with the children of the present study. The three available options were a clip-on microphone, a headset or a stand-alone microphone.

Clip-on microphones or headsets seemed most desirable, as it was expected that children in the present age group might find it difficult to keep still in front of a separate stand-alone microphone. However, both types of recording equipment were soon dismissed. The clip-on microphone proved to be unsuitable, as children frequently leaned over the table and would touch the table with the microphone. Even when sitting still, the frothing of clothes would have been picked up by it. The headsets on the other hand were unsuitable as well. The ones available were either too heavy or too big for the small heads of the three-year-olds and the lighter ones were too unstable. Some children were also very uncomfortable with the headset, refused to keep it on or moved it around, when it was in their way. Since the whole recording session was based on interaction, the more stable headsets, of which the earpieces covered the whole ear, were rejected, because they made it difficult for the children to understand the experimenter.

The stand-alone microphone seemed to be the only feasible alternative. It was possible to monitor the distance of the children to it, it was not interfering with the agility of the younger children and it was stable and stayed in one place. This was helpful, as children could be made aware of it when they were completely forgetting about the microphone, which was not always desirable. All children

recorded here were comfortable with this solution and, additionally, it was technically most compatible with the DAT recorder used.

3.3.2 Recording location

It was also important to choose where to conduct the recordings. The main contrasting aspects, which tend to stand in direct opposition to each other, are the quality of the recordings and the naturalness of the environment.

For the recording of adults, the environment – depending on the task – might not influence their cooperation and motivation much. On the contrary, a laboratory setting might help focus on the task and the scientific nature of the recordings. The technical quality of the data is usually best under these circumstances.

For the recording of children, however, it has been common practice to choose a location familiar to them. The assumption behind this is that children behave most normally in an environment they know. The data collected will therefore hopefully be quite natural and the cooperation higher. Small children might additionally feel uncomfortable in unknown surroundings, which might result in a refusal to talk. A recording at home was considered, but soon discarded due to time constraints, as well as the fact that it was not believed to add to the naturalness; on the contrary.

The difficulties of finding a quiet room in a kindergarten are self-evident. A separate building on the same premises posed problems of a very practical kind during wintertime. Relocating single children or small groups into a neighbouring building would have involved dressing and a teacher accompanying them each time. This would have caused too much disruption and would have been too time consuming, therefore the trade-offs between ease of use and recording in a somewhat noisy environment were deliberately accepted. Arrangements with the

kindergarten teachers helped to keep the noise in the adjacent rooms and hallways to a minimum during recording sessions.

3.3.3 Material and tasks

To allow for the best possible comparison across all children, the same material was presented to each of them and the method chosen was a picture naming task (see above). The pictures had to be chosen so that, ideally, they could be recognised and named by all children. This put a rather strong constraint on the collection and initial ideas about e.g. minimal pairs had to be quickly discarded. The usual constraints, e.g. excluding words with vowels in nasal context added to the difficulty.

The pictures were also selected to be as unambiguous as possible, to avoid having to resort to verbal prompting. The literature is undecided on this topic, but, if possible, prompting is avoided.

McRoberts and Best (1997) studied the vocal interactions of a child and her caregivers over a period of 1;2 years and they did not find that the child would accommodate her F0 to that of her parents. On the contrary, they found "evidence for consistent adjustments by the parents"(p. 719). However, they "do not rule out the possibility that other aspects of vocal behaviour might be imitated by infants" (p. 734). Therefore, if possible, prompting was avoided in the present study. For some words, however, the experimenter had to resort to prompting. Interestingly, these words differed, as children had strongly varying vocabularies. Additionally, not necessarily the younger children had to be prompted but also e.g. one of the older boys (SEA).

As mentioned several times, it was problematic to elicit sustained vowels as well as complete sentences from the children. Even when resorting to imitation

by verbally prompting the child, most either fell silent when it came to such artificial tasks or they started to engage in free language play that much, that the task became obsolete. As it was still important to elicit viable material from which to extract speaking fundamental frequency and formant frequencies, it was necessary that children did produce language. Especially the younger ones had problems with those tasks that actually focused on language, whereas looking at pictures and describing them, those of the picture naming task as well as those in the story book, never posed a problem. In retrospect, consulting e.g. a speech and language therapist on to how to elicit the relevant material from small children would have been a wise decision.

Depending e.g. on the time of day, physical shape and emotional state of the child on a particular day, motivation, curiosity, concentration or fear had a strong impact on their performance. Children who at one point agreed to participate would refuse one hour later. Not that this is surprising, but it sometimes proved to be very time consuming.

It was further found to be essential to avoid yes/no-questions, as at least the children in the present study usually did not elaborate on an answer if yes or no sufficed. Sometimes they simply nodded or shook their heads, even when told to say it out loud. These questions, however, proved to be a useful starter at the beginning of narrating the story *The Gruffalo*, when a question of the kind "And what happens next?" often prompted an answer like "The snake" instead of "The Gruffalo meets the snake". However, children again varied greatly as to how much information they would submit verbally, with one of the youngest (LAB) being the most articulate child.

3.4 Measurements

Of the seventeen children from whom the complete dataset was obtained, ten were chosen for further analysis. In the following, I will outline those parameters that were measured and subjected to statistical analysis. A list of the measurements obtained can be found in Appendix C of the original thesis.

3.4.1 Fundamental frequency

The underlying considerations based on the existing literature as to the choice of parameters for this study have been discussed in chapter 2. The measurement of speaking fundamental frequency was obtained from the children's passages of spontaneous speech.

The relevant material was digitised (16-bit, mono, 22 kHz sampling rate) from the original DAT and saved to separate audio files, which for analysis were uploaded into PRAAT, a free software tool for speech analysis and synthesis, developed at the University of Amsterdam, Holland (Boersma and Weenik, 2004). The sound file together with a visible pitch contour for the voiced sections of the speech sample were displayed, allowing for a pitch range of 75 to 900 Hertz. Apart from the voice of the child, this sound file naturally also contained e.g. backchannel utterances and feedback-questions of the experimenter. Together with wrong pitch estimations calculated by the program, especially in plosives, these parts were devoiced using the unvoice function in PRAAT. Material that was pitch-halved or doubled could often be corrected by calculating the actual pitch of the relevant stretch of speech by hand and moving wrong pitch points up or down to the right level, a procedure which is easily done in PRAAT. Voice information in those portions containing creaky voice was also deleted. Although

this might be regarded as a distortion of the actual pitch range a child had produced, creak is not reliably calculated, hence it distorts measurements. Outliers in the higher F0 regions of over 800 Hz were rare and either pertained to obvious miscalculations or vocalisation other than speech, e.g. laughter and could therefore safely be deleted. Except for all truly voiced portions of speech produced by the child, material was devoiced. Consequently, these devoiced portions of the signal were excluded from the calculation of mean speaking fundamental F0.

Initially, mean F0 was intended to be calculated from interpause stretches. As, however, speaking rate and pausing behaviour varied considerably between children and no clear definitions to this respect could be obtained particularly for the speech of children between three and five, mean F0 was calculated over the complete passage of spontaneous speech. As this included several minutes of speech, this step was deemed appropriate.

Another possible method of obtaining speaking fundamental frequency, by isolating the vowels from the words produced in the picture naming task and extracting fundamental frequency from there, did not prove feasible. This was due to the fact that the children varied their F0 often extremely on the single words, especially on monosyllabic ones, which at times resulted in F0 measurements of outliers around 900Hz in the first part of the vowel and creaky voice in the second half.

3.4.2 Formant frequencies

Although the words of the picture naming task were unsuitable for the calculation of fundamental frequency, they were used for the extraction of the first two formants. This was again done in PRAAT through marking the steady state portion of a vowel and then extracting the relevant formant with the *Get the first/second*

formant-command in the editing window. The steady state portions used were generally kept to a duration of 100-160 ms, and were only shorter if the material was too short from the outset. Formants were measured for each vowel, yielding 38 measurements for [i] and [o] and 40 each for [a] and [u].

3.5 Results

3.5.1 Height, weight and F0

As a first step, it was necessary to ascertain whether the children studied here displayed any differences with respect to their height, weight and fundamental frequency. As the sample consisted of only five subjects per sex, yielding ten values for height, weight and fundamental frequency each, only descriptive statistics were conducted. In accordance with the findings of Fitch and Giedd (1999), it was expected that there should not be a difference between 3 to 5 year old boys and girls with respect to any of the three parameters, a finding which could in part be replicated and which is presented in figures 3.2, 3.3 and 3.4. Table 3.1 lists the measurements for height, weight and fundamental frequency for all ten subjects.

In general, boys are slightly taller (115 cm) than girls (109 cm). The smallest child is a 3-year-old girl; however, as there were no 3-year-old boys in the sample, this measurement is lacking direct comparison. When the group is split according to age, the divergence in height between 3- to 4-year-olds (106 cm) and 5-year-olds (118 cm) is much greater (cf. 3.2).

In terms of weight (cf. 3.3), a similar pattern emerges. Whereas boys and girls on average do not differ much from each other (21 versus 18 kg), the 3- to 4-year-olds are much lighter (16 kg) than the older children (23 kg).

Speaker	Sex	Age	Height(cm)	Weight(kg)	F0(Hz)
ALE	male	70	122	23.6	244
SEA	male	67	116	21.0	335
JAC	male	61	123	26.5	263
PAU	male	48	105	17.9	288
ERI	male	53	109	15.6	292
JOS	female	62	116	21.1	303
ALS	female	62	114	22.4	255
MAR	female	53	111	18.8	281
LAB	female	46	102.5	14.7	326
RIC	female	39	100	14.8	307

Table 3.1: List of speaker sex, age in months, height, weight and mean fundamental frequency.

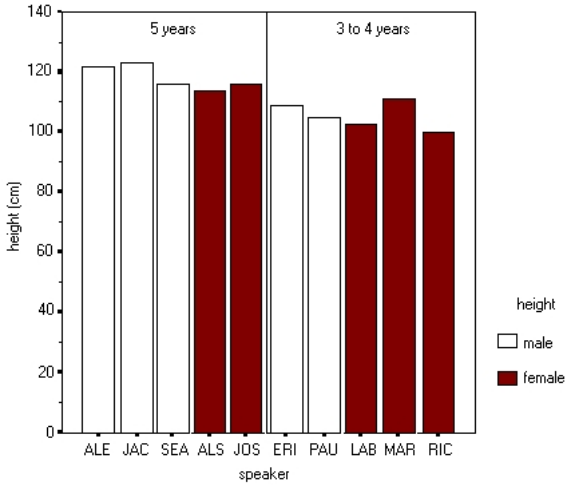

Figure 3.2: Height (cm) of boys and girls.

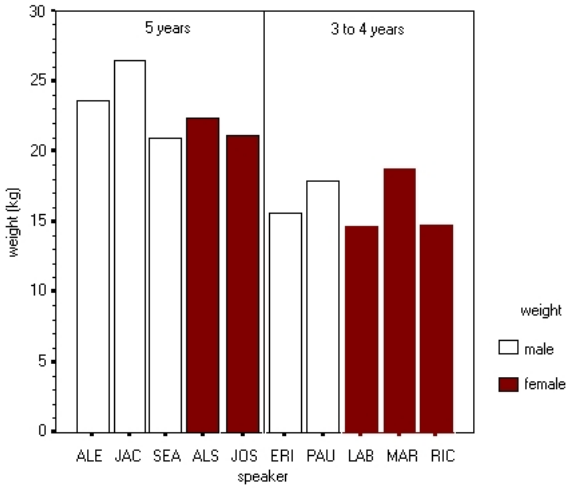

Figure 3.3: Weight (kg) of boys and girls.

Average fundamental frequency measured for five 3- to 5-year-old boys is about 10 Hz lower (284 Hz) than for girls of the same age group. However, the variance for boys is much greater, spanning almost 100 Hz, from the lowest frequency of 244 Hz (ALE) to 335 Hz (SEA) — both produced by a 5-year-old. Table 3.1 lists all ten values.

From looking at the figure, it seems that this difference in F0 between boys and girls could be significant. However, the variance in the male group is very high and the highest F0 value measured was 335 Hz and was produced by a boy (SEA). On the other hand, it cannot be said that all low values were produced by boys, as one of the 5-year-old girls (ALS) produced the second lowest value of 255 Hz.

In general, the fundamental frequency of the younger children is higher than that of the older ones. Nevertheless, it is interesting to note that the person with the highest F0 was a 5-year-old boy.

It was confirmed that the younger children differ from the older ones in terms

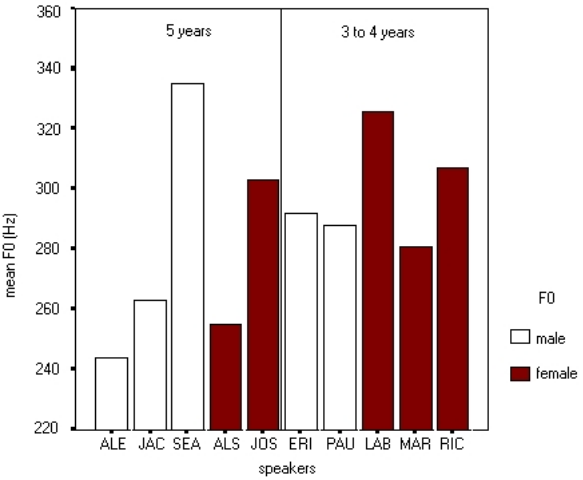

Figure 3.4: Mean fundamental frequency of each child subject, grouped by age.

of bodily measurements. The 3- to 4-year-olds were generally smaller and lighter than the five year olds. The two age groups also differ with respect to F0 in that the younger age group displayed a mean F0 of 299 Hz, whereas the older children's mean lay about 20 Hz lower at 280 Hz. Again, caution needs to be exerted, as mean F0 was calculated over just ten values of again mean F0 of spontaneous speech. The large within group variance should not be overlooked.

3.5.2 Formant frequencies

The formant frequencies were obtained from the long vowels in word-initial stressed syllable position. Rotated scatterplots giving an impression of the vowel triangle were produced as a means of descriptive statistics (cf. figures 3.5 and 3.6 for differences with respect to sex and figures 3.7 and 3.8 for those relating to age). A multivariate analysis of variance (MANOVA) was then conducted, to see whether the first two formants differed with age and sex, as well as with vowel

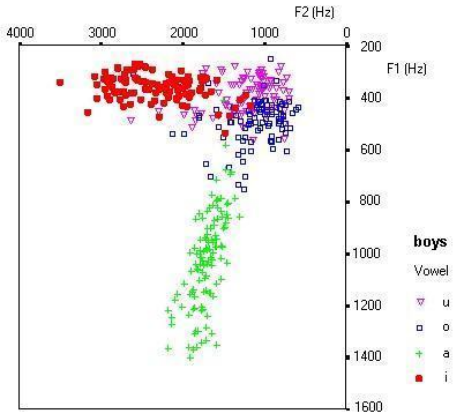

Figure 3.5: Scatterplot for F1 and F2 measurements of the vowels [i], [a], [o] and [u] of five boys.

type. For this purpose, the measured data was converted into a format readable for the statistics software program SPSS 10.0 (Statistical Package for the Social Sciences).

The third variable, vowel type, was not expected to bring forth any surprising significances. *Type of vowel* was only included to ensure, that the actual measurements obtained here reflect very basic assumptions about vowel quality. That is to say that the vowels presented to the listeners were sufficiently distinguishable in quality. It was expected, that the vowels would differ significantly from each other, apart from F1 for [i] and [u] and possibly F2 for [o] and [u].

The scatterplots give a first impression of the vowel triangle of the four groups of children. Especially F1 for [a] seems to be quite different for boys and girls, with girls displaying on average higher values than the boys. F2 seems to be slightly higher for the [a] of girls as well, although this difference is not as pronounced.

With respect to age, F1 for [a] is clearly lower for the older children than the younger ones and the F1 of [o] seems to be higher for younger children. For the

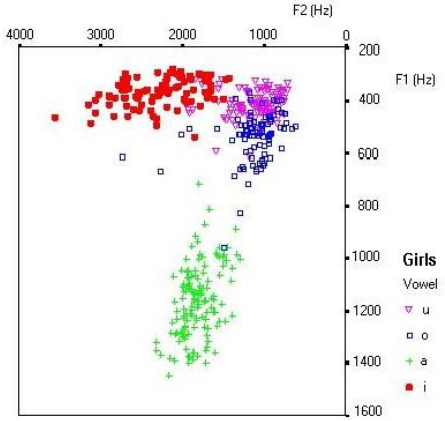

Figure 3.6: Scatterplot for F1 and F2 measurements of the vowels [i], [a], [o] and [u] of five girls.

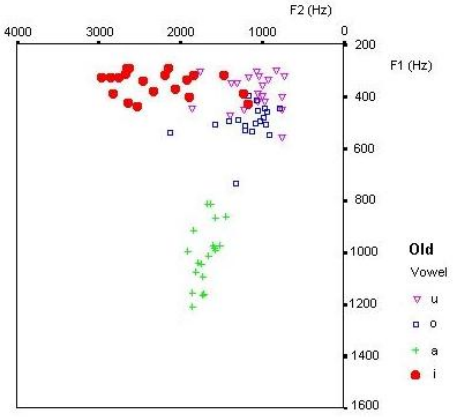

Figure 3.7: Scatterplot for F1 and F2 measurements of the vowels [i], [a], [o] and [u] for the older group (five 5-year-olds).

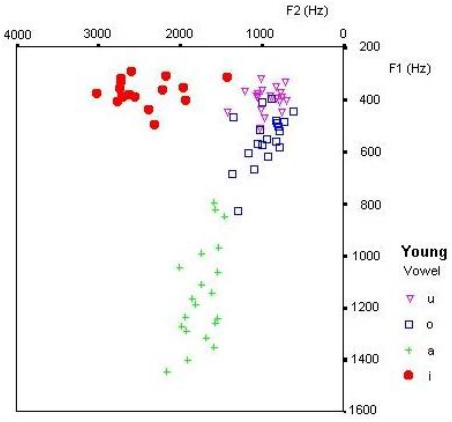

Figure 3.8: Scatterplot for F1 and F2 measurements of the vowels [i], [a], [o] and [u] for the younger group (five 3- to 4-year-olds).

other two vowels, no clear tendencies could be observed from the scatterplots alone.

The results of the multivariate analysis of variance are shown in table 3.2. Since all factors except type of vowel have too few groups to perform Post Hoc tests, the direction of interactions are made accessible through diagrams of interaction.

As expected, a significant interaction between type of vowel and the two formants was observed and a subsequent Post Hoc test (Scheffé) revealed, that F1 differs significantly for all vowels except [i] and [u] and F2 differs for all except [o] and [u]. As this information simply forms the basis for further analysis and does not add any new information, diagrams are dispensed with here.

Generally, it can be observed that the differences in F1 are more pronounced than those for F2. When looking at F1 and F2 separately, it can be observed that F1 shows significant differences with respect to the main effects of sex and age of a speaker, whereas F2 does not exhibit significant differences for these two factors. To reveal which vowels were particularly prone to significant differences in F1 with respect to sex and age, diagrams highlighting interactions were pro-

Main effect	dependent variable	F-value	significance
sex	F1	22.114	**.000
	F2	0.021	.884
age	F1	24.104	**.000
	F2	0.817	.367
vowel	F1	1242.289	**.000
	F2	343.804	**.000
Interactions	**dependent variable**	**F-value**	**significance**
sex*age	F1	10.134	*.002
	F2	4.078	*.044
sex*vowel	F1	5.163	*.002
	F2	0.850	.468
age*vowel	F1	7.073	**.000
	F2	7.216	**.000

Table 3.2: Results of the multivariate analysis of variance (MANOVA). The differences in the formants F1 and F2 depending on sex, age and vowel type are displayed. One asterisk marks significant results ($p \leq 0.05$), two asterisk mark highly significant results ($p \leq 0.001$).

duced. As these frequencies are averaged across all vowels, these effects have to be put into relation to the findings on speaker age and vowel interaction, as well as speaker sex and vowel interaction.

Figure 3.9 shows the interaction of sex of the speaker with vowel type for F1. Whereas [i], [o] and [u] only exhibit marginal sex dependent differences, the female F1 value for [a] is decidedly higher than the mean F1 for the group of boys, which supports the findings of e.g. Nairn (1997) or Busby and Plant (1995) who found sex dependent differences in F1 for open vowels. The differences found in F2 are not significant.

The pattern for age and vowel type looks similar, in that the main difference between the group of 3- to 4-year olds and 5-year-olds can be observed for F1 of [a], which is lower for the children in the older group (cf. figure 3.10). However, all vowels express this tendency of being lower in the older age group.

F2 of [a] is basically stable across age, whereas F2 of [o] and [u] appear to

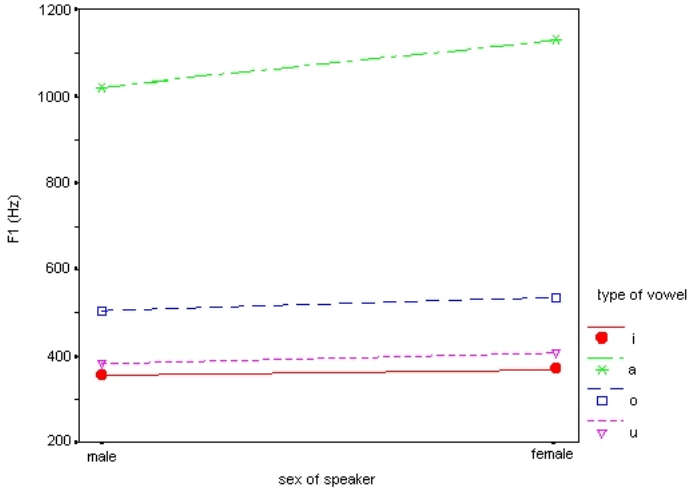

Figure 3.9: Diagram of interaction between sex of speaker and type of vowel with respect to the first formant (F1).

rise and F2 of [i] appears to fall. In other words, those vowels with the lowest F2 values ([o] and [u]) experience a rise in F2, and [i], the one with the highest value is lowered with age. The F2 values for the older age group are therefore closer together than those of the younger ones. This centralisation of formant values might either be a general tendency in all older children or it might reflect a gender specific behaviour, since this age group contained more boys than girls. The tendency of reduction typically found in adult males might be already evident at this age. A further discussion of this finding is discussed in Chapter 5.

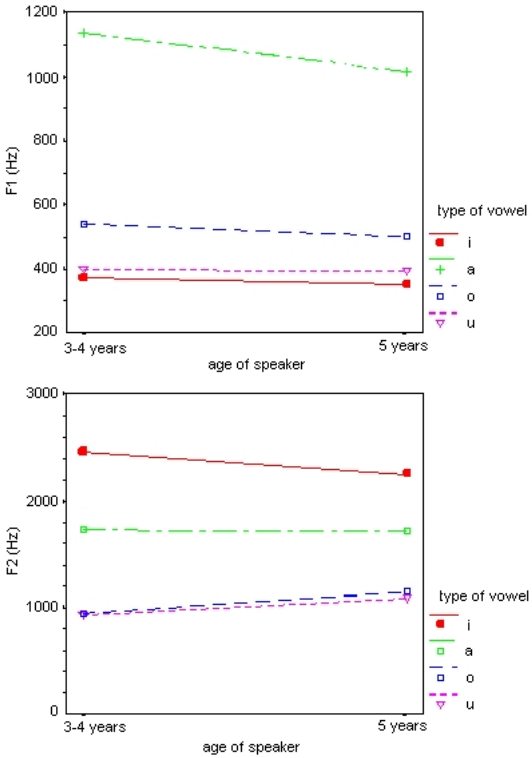

Figure 3.10: Diagrams of interaction between age of speaker and vowel type with respect to the first and second formant.

4 Perception: Listening Experiment

This chapter describes the perception experiment conducted with adult listeners on a subset of the child data collected in the kindergarten. First, a small pretest will be described, which was run to determine the presentation of the vowel stimuli. Afterwards, the results obtained via descriptive statistics are presented and discussed.

4.1 Preparation and pretest

Since the children here did not produce sustained, isolated vowels like the ones other researchers had used in their studies, the vowel stimuli that were supposed to be offered to listeners in the present study posed a problem. Vowels had to be digitally cut out of the words the children had produced in the picture naming task. However, embedded in words – at times even in complete sentences – these vowels were assimilated, sometimes centralised or shortened, therefore very unlike the possibly near-cardinal-like quality of the sustained vowels used in other studies. It was expected that this might lead to severe identification difficulties for listeners, if these vowels were presented one at a time. In the present data, some vowels had steady states as short as 70 ms; other researchers who used sustained vowels, usually presented material of about 1 second to their listeners. In the present case it was feared that correct identification results might be artificially lowered, simply due to extremely short stimulus durations. On the other hand, these vowels were produced very naturally (though presented without their consonantal context), and might therefore in fact enhance overall recognition rate because they are more typical representatives of what people might actually hear

from children in everyday life.

A small pretest was run to establish, whether recognition rates were decidedly higher when more vowel material was offered as a stimulus as opposed to a single vowel condition. Listeners were presented with stimuli consisting of one child's vowel followed by a new stimulus and in the second half of the pretest four vowels spoken by one child were presented as one stimulus followed by another group of four vowels and so on. Additionally, to compensate for the shortness of some vowels, not just the steady state portions were used, but transitions from the preceding consonant were included, yielding the percept of a syllable rather than that of a vowel. This makes them more comparable to the isolated sustained vowels of one second in duration, which are arguably as much syllable-like as the vowels including their transitions used here.

Sixteen single vowel stimuli and 7 four-vowel-groups where used for this purpose. As the mode of presentation was identical to the one for the actual perception test, the description of the presentation of stimuli can be read in section 4.3 below. Four listeners, one male and three females, who were trained phoneticians, judged these utterances as to whether they were spoken by a boy or a girl. The test results did reveal that some listeners performed better on the single vowel stimuli (two listeners scored 50% and 56% correct identification rate) and others did worse, with rates lower (31% and 38%) than the 43% all of them reached in the four-vowel-groups. It was concluded that more material in this case did not necessarily produce a better recognition rate, a finding which was also reflected in the comments listeners made about the task. Most of them reported that they did not consider it easier to correctly judge the stimuli which included more material.

From the results of this short pretest, it was expected that listeners should generally be able to judge single vowel stimuli and a forced choice test, with one vowel stimulus presented at a time, was chosen. An additional advantage of presenting single vowel stimuli was the possibility to compare correct recognition

rates with respect to vowel type. This effect would have been lost, if listeners had judged four vowels as one single stimulus. It was further believed that the more time people have to react to a stimulus, especially such a short one as a vowel, the less intuitively they judge it. It might even induce people to revise a choice.

4.2 Subjects

Twelve adults, six women and six men between the ages of 24 and 36 years (mean = 29.75 years) served as listeners. They were all affiliated with the university and apart from two of the males had no regular contact with children.

4.3 Procedure and stimuli

The stimuli were presented in three blocks: vowels, words and spontaneous speech. Vowels were presented one at a time, with a beep (a pure tone of 1 kHz) preceding the stimulus presentation and a three second interval response time. First, all [i] were presented, followed by [a], [o] and [u]. Vowels within such a sub-block were randomised. In the second block, single words had to be evaluated and the procedure was like the one for vowels. Both, the vowel and the word block contained 158 stimuli. The last block contained twenty short passages of spontaneous speech, which were taken from children's narrative of the story *Der Grüffelo.* All listeners received the same order of stimuli.

The stimuli were compiled into three WAV audio files, with each block started separately by the experimenter. Stimuli were presented via headphones and each block was preceded by training material, so that the listeners could adjust to the voices and tasks. These dummies were excluded from further analysis. The very

first block and sub-block, the group of [i] was preceded by four such dummies, all other blocks, the three remaining vowels, words and spontaneous speech were each preceded by two dummies. Vowels were, naturally, preceded by vowels of the respective quality, words by words, etc. Listeners received an answer sheet (cf. Appendix A of the original thesis) and were asked to mark *boy* or *girl* after each stimulus, depending on what sex they assumed the speaker of the utterance had. Total running time of the experiment including instructions was approximately forty minutes. There was the possibility of a short break between the three blocks of vowels, words and spontaneous speech, but non of the listeners felt the need to take it.

With regard to the spontaneous utterance condition, two passages of approximately 10 seconds each were taken from the recordings of the ten children. Where possible, this material was collected en bloc from a longer passage spoken by the child, uninterrupted by the researcher. Sometimes, however this was not possible, as some children did not produce ten-second-material in one take. In these cases, several short utterances were assembled to constitute one stimulus. When asked afterwards whether they experienced one or two stimuli to be awkward in some sense, none of the listeners reported these assembled ones. This was taken as an indication that this assemblage did not stand out as disturbing or different.

For each child, except one girl, two samples of spontaneous speech, each 10 seconds long, were selected, yielding twenty short passages that had to be evaluated by the listeners. To make up for the second sample of the one girl three samples of another girl were used, so that an equal number of boys and girls samples was present in the spontaneous speech condition. The listeners were presumably not aware of the distribution of stimuli. The sound files containing the stimuli used in the perception experiment can be found on CD-ROM (Appendix D of the original thesis).

4.4 Results

The results of the perception experiment data were subjected to several descriptive statistics which can be found below. Correct responses for each listener and each stimulus were calculated and transformed into mean correct identification rate (CIR) in percent. This mean CIR was then used for further analyses.

4.4.1 Listeners

The most general hypothesis, that it is possible for listeners to determine the sex of a child above chance level, was confirmed. Listeners of the present study were able to assign correct sex in 61.73% of the cases. Highest overall correct identification rate was 66.67%, lowest 55.36%. Both values were produced by females. Listeners' responses were entered into Microsoft Excel spreadsheets and mean correct identification rate was calculated by counting all cells containing correct responses, dividing them by the total number of cells and multiplying this by 100. Contrary to what was predicted, women did not achieve overall higher correct identification rates than men, but the values of the two groups only differed marginally, men achieving a slightly higher rate (62%) than women (61.46%).

Since the group of males included two experts, TB and JT, fathers of children below school age, the mean was first calculated excluding them, since experts are known to perform worse than lay judges (cf. Moore 1995). Without listeners TB and JT, correct identification rate, however, was at 61.91% very similar to the mean of the group when those two listeners were included. It was therefore assumed that these males did not exert a detrimental effect on overall male correct identification rate and they were henceforth included in all of the investigations. In contrast to the men, five out of the six female listeners produced the three

Listener	Sex	Vowel	Word	Passage	Average1	Average 2
MK	male	56.33	60.13	85.00	67.15	59.82
RK	male	65.19	58.86	60.00	61.35	61.90
SB	male	56.96	68.35	65.00	63.44	62.80
SG	male	62.03	62.03	80.00	68.02	63.10
JT	male	54.43	65.82	75.00	65.08	61.01
TB	male	62.66	61.39	85.00	69.68	63.39
UP	female	52.53	56.96	65.00	58.16	55.36
GS	female	63.92	53.80	55.00	57.57	58.63
SJ	female	58.86	56.96	80.00	65.27	59.23
CB	female	56.96	69.62	95.00	73.86	65.18
AF	female	64.56	65.82	90.00	73.46	66.67
MS	female	60.13	66.46	70.00	65.53	63.69
Average	–	59.55	62.18	75.42	65.72	61.73

Table 4.1: Correct identification rate (CIR) according to listener and stimulus type. Listener sex is also indicated. Average 1 is the mean calculated by summing-up the CIR of vowels, words and spontaneous speech, divided by three. Average 2 is calculated by accumulating correct responses of all speakers into mean CIR per stimulus, divided by number of stimuli.

worst overall recognition rates and the two highest ones; men centered around the mean value.

Table 4.1 summarises the correct identification rates obtained from each listener, according to stimulus type (vowel, word and spontaneous speech) as well as two values of overall correct recognition rate for each listener. The value Average 1 is a simple mean of the three columns before, and therefore does not take into account that there were fewer spontaneous speech samples than vowels or words in the data. Average 2 was calculated as indicated above. Since these values are slightly different from each other, and since for some comparisons one or the other is more suitable, both are listed here.

The general trend mirrors the hypothesis that people perform better, the longer the utterance is. However, four listeners (RK, TB, GS and SJ) performed better in the vowel condition than in the word condition. These listeners include

	Vowels	Words	Passage
average	59.55	62.18	75.42
male	59.60	62.76	75.00
female	59.49	61.60	75.83
worst	52.53	53.80	55.00
best	65.19*	69.62	95.00

Table 4.2: Correct identification rate (CIR) in percent with respect to stimulus type. The number followed by the asterisk indicates the one single value produced by a male in the categories best and worst identified.

two males and two females. The possibility that these four listeners had a different strategy than the other listeners for extracting sex from a speech sample will be discussed in Chapter 5. Also against the trend, one male (SB) achieved higher results in the word condition than when evaluating spontaneous speech.

In the spontaneous speech condition, one female listener (CB) made only one mistake, yielding a correct identification rate of 95%. Interestingly, this mistake was made on the stimulus of a 5-year-old male (JAC), who was not the one with the overall highest mean fundamental frequency. However, he yielded the two worst results in the spontaneous speech task and in this particular case reached the lowest CIR score of all (cf. figure 4.6). On a previous sample of spontaneous speech by speaker JAC, the same listener, CB, had assigned the correct male sex to that speaker.

The highest and lowest recognition rate for each stimulus type was usually produced by a female listener, except for the best overall vowel identification, which was performed by a male listener (RK).

Table 4.2 partly summarises the results shown in table 4.1. Here, percentages are given for male and female listeners together, according to stimulus type. In addition, the best and worst absolute recognition rate obtained by a listener are given for each stimulus type.

The variance in the spontaneous speech sample appears fairly large as the

lowest correct identification rate was 55% and the highest 95%, with a mean of roughly 75%. It is important here to remember that the spontaneous speech sample contained only twenty utterances and one error induced the correct identification rate to be reduced by 5%.

4.4.2 Speakers

As mentioned earlier, children were correctly identified with respect to their sex in 62% of the cases. Contrary to most findings in the literature, where girls were better identified than boys, this study reveals mean correct identification rates being about equal for the sexes, with boys' rates being at 61% marginally lower than those of girls at 63%. The child, who was best identified by adult listeners was a 5-year-old boy (ALE), who was correctly identified in 85% of the cases. The one receiving the lowest percentage of correct votes was a 3-year-old boy (PAU), with 39% correct sex allocation. Since there were no missing values, i.e. all listeners judged all stimuli, the reverse can be stated in that PAU was overwhelmingly misidentified as a girl. This result is also consistent with the hypothesis that the older a child is, the easier it is to assign the correct sex.

As can be seen in figure 4.1 and the corresponding table 4.3, three of the older children (ALE, JAC and JOS) and three of the younger ones (ERI, LAB and RIC) were correctly identified above chance. In the older group, two of the children (ALE and JAC) are male, in the younger group, two of the children (LAB and RIC) are female. The children, who were identified below chance level, one boy and one girl from either age group, reached correct identification rates between 39% and 50%. The second young boy (ERI) just barely reached a rating above chance level and was misidentified as a girl in 47% of the cases.

The boxplot in figure 4.2 indicates that the variance in both sexes spans the

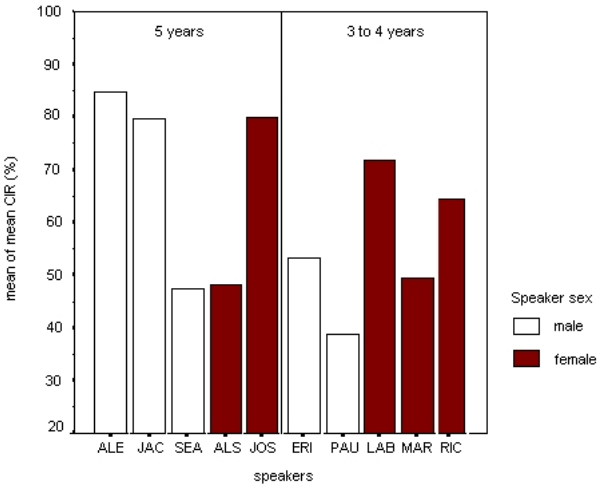

Figure 4.1: Mean CIR for each speaker across all twelve listeners.

whole continuum, showing that for both sexes, some stimuli were unanimously judged correctly (100%) as well as incorrectly (0%). As boxplots in SPSS display the median only and not the arithmetic mean, values for the latter will be found in the text. In most cases, however, the relative position of mean and median of the relevant parameters are comparable, and since variance is displayed, this form of diagram was preferred over others.

Fourteen stimuli were misidentified by all twelve judges, meaning they yielded an error rate of 100%. Thirteen of these stimuli were produced by three children: two 3- to 4-year olds (PAU, male, and MAR, female) and one 5-year-old girl (ALS). All of these children were misidentified with respect to their sex over 50% of the time; both, MAR and ALS had decidedly lower mean fundamental frequency than the other girls in their respective age group.

In table 4.3, mean recognition rates for each speaker are listed, starting with the highest rate awarded to a speaker. For better comparison, sex and age are also indicated. Together with figure 4.3, this table shows that mean correct identification is higher for the group of 5-year-olds than for the 3- to 4-year-olds, with

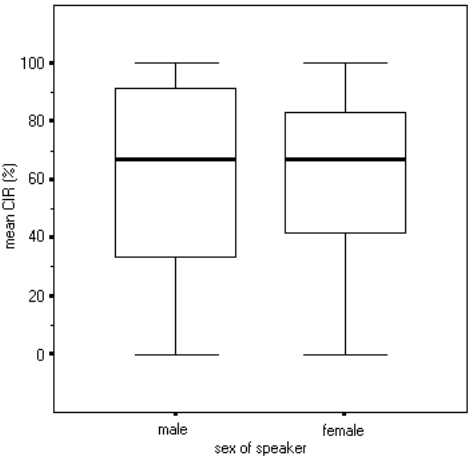

Figure 4.2: Correct identification rate (CIR) of boys and girls.

the older children being judged correctly in 68% of the cases, the younger ones settling at 55%. This is hypothesis compliant, in that higher correct identification rates were expected for the older childen.

If the groups of children are further subdivided into young and old males and young and old females, a pattern emerges that can be retraced in figure 4.4. Whereas the correct identification rate for girls at age five with 64% is only 2% higher than that at age 3 to 4, the correct identification of boys rises approximately 25%. The corresponding percentages are 46% versus 71% for the two male groups and 62% versus 64% for the girls. The boxplot shows, that no stimulus of older males and younger females was ever unanimously misjudged by all listeners. However, one of the older boys, SEA, had 8 of his 34 stimuli constantly misjudged by ten of the listeners[1].

There were slightly more male speech samples amongst the stimuli (170) than female ones (166). In percent, this amounts to 50.6% and 49.4% respectively. Table 4.4 shows the tendency of listeners to respond *boy* or *girl*. As it is

[1] See Appendix C of the original thesis for a complete listing.

Speaker	CIR	Sex	Age
ALE	84.80	male	5
JOS	79.90	female	5
JAC	79.66	male	5
LAB	71.88	female	3–4
RIC	64.39	female	3–4
ERI	53.19	male	3–4
MAR	49.48	female	3–4
ALS	48.10	female	5
SEA	47.55	male	5
PAU	38.73	male	3–4

Table 4.3: Mean correct identification rate (CIR %), descending from highest to lowest score, as well as sex and age, according to speaker.

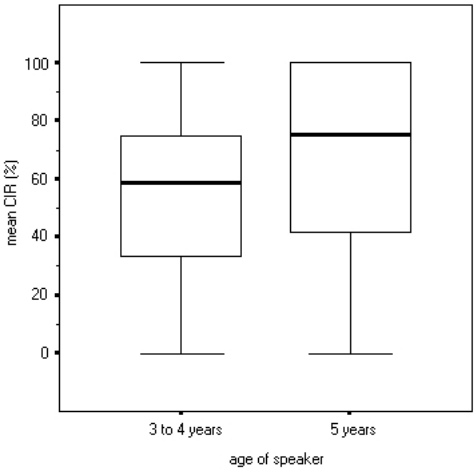

Figure 4.3: Correct identification rate (CIR) of 3- to 4- versus 5-year olds.

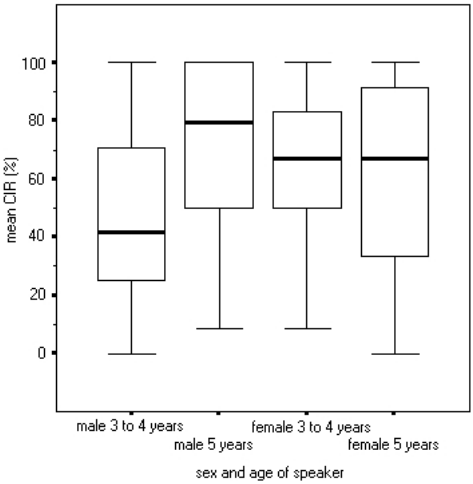

Figure 4.4: Comparison of correct identification rate (CIR) of younger and older boys and younger and older girls.

revealed, seven listeners tended to respond *girl* more often, five listeners favoured the *boy*-response. The fact that girls are identified with more accuracy than boys (63 versus 61%) can therefore hardly be attributed to an overall listener tendency towards the female sex. However, amongst male listeners, there is a stronger tendency to respond female (four out of six listeners tended towards this preference), whereas female listeners showed an even distribution of preference — three were more partial to answer girl, three more often opted for the answer boy.

4.4.3 Type of stimulus

The results can be analysed via a third perspective. So far, listener and speaker parameters have been presented. The type of stimuli and, with regard to the formant measurement and their results presented in Chapter 3, especially the patterns emerging for the four vowels are of interest in this and the following section.

Listener	Sex of Listener	boy response	girl response
MK	male	47.3	52.7
RK	male	51.8	48.2
SB	male	47.3	52.7
SG	male	47.6	52.4
JT	male	53.9	46.1
TB	male	48.5	51.5
UP	female	53.6	46.4
GS	female	45.5	54.5
SJ	female	52.1	47.9
CB	female	47.9	52.1
AF	female	51.8	48.2
MS	female	42.9	57.1

Table 4.4: Listener tendency to respond *boy* or *girl*.

As table 4.1 already showed, mean correct identification rate is lowest for vowels with 60%, followed by 62% for words and 75% for spontaneous speech. A difference of 2% between vowels and words is marginal as expected, but it confirms the hypothesis that the longer the utterances, the higher the correct identification rate. Figure 4.5 reveals that no spontaneous speech samples were unanimously misjudged.

For all three conditions, there were several stimuli where speaker sex was correctly identified by all listeners, as the variance demonstrates. Indeed, of the fifty-five stimuli that were unanimously associated with the correct speaker sex, nineteen pertained to vowels (this amounts to 12% of all vowel stimuli), twenty-eight (18%) to words and six (30%) to passages of spontaneous speech. Nonetheless, on the other end of the scale, a different picture emerges. Whereas in the vowel and the word condition three vowels (5%) and nine words (14%) failed to get any correct gender judgement, the lowest correct identification rate for spontaneous speech was 33%. A passage of spontaneous speech was therefore never misidentified by all and also always correctly identified by four or more listeners. Whether the order in which stimuli were presented was instrumental in

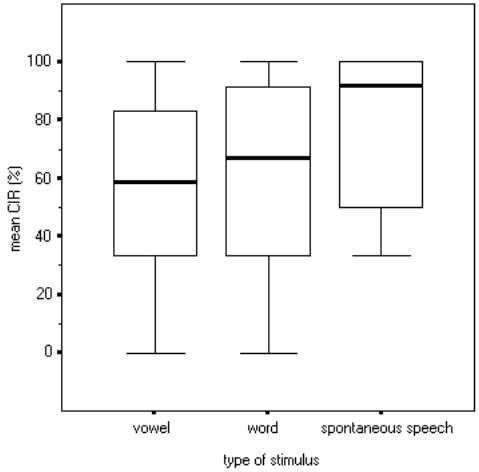

Figure 4.5: Comparison of correct identification rate (CIR) of vowels, words and spontaneous speech.

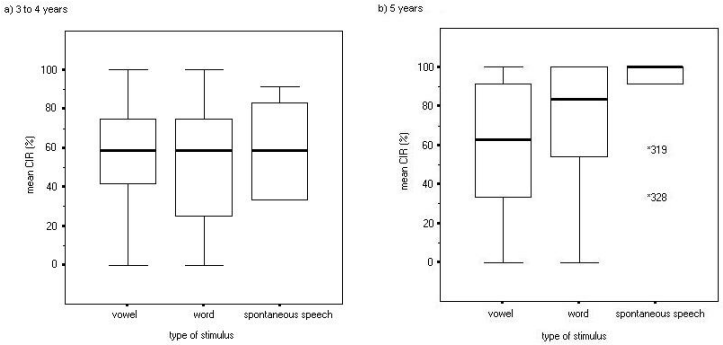

Figure 4.6: Comparison of correct identification rate (CIR) of vowels, words and spontaneous speech of 3–4 and 5-year-old children.

yielding this result, will be discussed in Chapter 5.

Since overall correct identification of older children was better than that of the younger ones, figures 4.6a and 4.6b differentiate correct identification of stimulus type with respect to the two age groups.

Here, it becomes evident that the differences in mean correct recognition rate with respect to stimulus type were probably not produced for the group of 3- to 4-year olds, as the values of CIR for all three types of stimuli are relatively close together. Additionally, lowest correct recognition was not achieved on vowels (58%), but on words (52%) for this age group. Spontaneous speech reached a CIR of 60%.

Instead, the results of the older group are apparently responsible for the overall picture, as clear differences in CIR between the three types of stimuli emerge. The smallest units received the worst correct identification of 61%. Words fared substantially better with 72%, topped by a correct identification rate for spontaneous speech of 88%. Two outliers were included in figure 4.6 which pertain to stimulus number 319 and 328, both produced by JAC[2].

4.4.4 Vowels

As it was predicted in the hypothesis, correct identification rate varied with vowel type (cf. figure 4.7). In the present study in general, speaker sex was most often correctly assigned for the vowel [o], followed by [i], [a] and [u]. The first two vowels reached 65% and 62% correct identification rate, whereas the other two both obtained 58%. This is contrary to what was expected from the literature, where it is reported, that the open vowel [a] is the one displaying the most distinctive gender features and should therefore be best identified.

[2]The corresponding information for these two can be found in Appendix C of the original thesis.

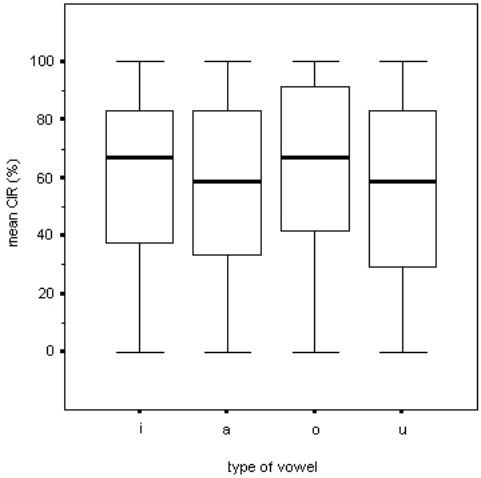

Figure 4.7: Mean correct identification rate (CIR) of vowel type for both sexes.

Vowel	both sexes	male	female
[i]	61.95	62.29	61.57
[a]	58.33	46.25	70.42
[o]	65.31	67.92	62.71
[u]	57.91	64.17	51.67

Table 4.5: Mean correct identification rate of vowels according to sex of speaker.

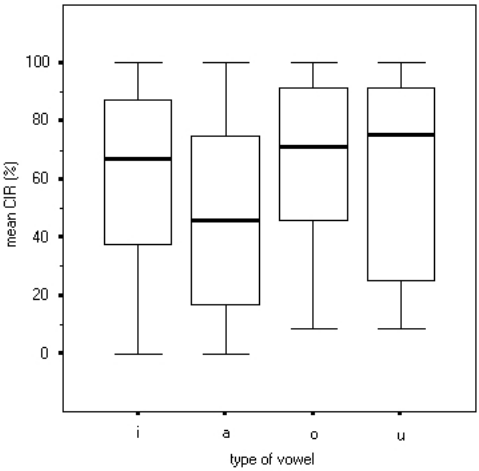

Figure 4.8: Mean correct identification rate (CIR) of vowel type for boys.

To reveal a possible difference in correct identification rate of sex with respect to vowel type, mean correct identification rate was calculated for each sex separately and is shown in table 4.5 and figures 4.8 and 4.9.

Here it becomes evident that while males achieved worst identification rates for [a], the pattern was exactly reversed for the girls, whose vowel [a] was best identified compared to the other three vowels. Boys, on the other hand, were more often correctly identified in the back vowel condition, yielding clearly higher results than the girls, especially for [u].

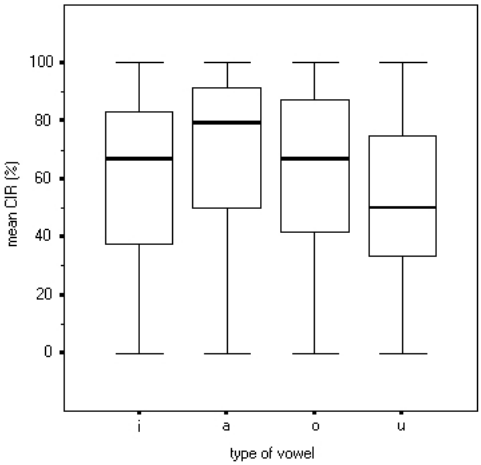

Figure 4.9: Mean correct identification rate (CIR) of vowel type for girls.

5 Discussion

In the following, the results of the phonetic measurements carried out on the material collected in the kindergarten, as well as the perceptual evaluation of the data performed by twelve adult listeners will be discussed with reference to the findings of earlier studies which have been described in Chapter 2. First, for a better overview, the production and the perception component will be dealt with separately. In the last subsections of this chapter, a merger of the two will be attempted.

5.1 Production Experiment

Several researchers have stated that before puberty children of similar age do not differ significantly with respect to the anatomy of their vocal apparatus. Sometimes, this was inferred from outer bodily measurements such as height and weight (Günzburger et al., 1987); others, like Fitch and Giedd (1999) have measured the vocal tracts of children via magnetic resonance imaging techniques and did not find significant sex differences in same-aged pre-pubertal children.

5.1.1 Height and Weight

The present study aimed at establishing whether these statements were true for a small sample of ten German children between the ages of three and five. Measurements similar to those of Günzburger et al. (1987) (they additionally measured circumference of the throat) were carried out and the results, though not statist-

ically significant, imply that the difference between the height and weight of 3- to 5-year-old girls in comparison to same aged boys is indeed small. Boys were on average 6 cm taller and 3 kg heavier than the girls. When the group was split according to age, these differences, however, became larger. The children in the younger age group of 3–4 years were on average 12 cm smaller and 7 kg lighter than those in the older group.

Notwithstanding these results, the individual height and weight of each child has to be stressed to put the aforementioned findings into perspective. The tallest boy in the present sample (ALE) e.g. was reported to be one of the tallest children of the kindergarten in general. To counterbalance this, a tall 5-year-old girl (JOS) was selected, but the discrepancy towards the younger ones was probably intensified by this choice. A better balance in such a small sample would have been desirable, but, as the children chosen produced the best speech output, sacrifices were made in other areas.

However, it is possible to compare the ten children analysed here with a much larger sample of data (cf. Zabransky 2003). From 1994 to 2002, 18,000 children and young adults between 2 and 20 years from the Saarland, Germany, were measured and weighed for the Saarländische Wachstumsstudie SWS (Saarland Growth Study). Height and weight were given in percentiles. When the ten children of the present study are compared to the referential values of the other study, it appears that most of them are in the upper or middle regions for height and weight. This holds especially true for the 5-year-olds, where height and weight are all in the upper regions, meaning that between 70 and 95% of their same aged peers are smaller than these five children. Two of the boys (ALE and JAC) displayed the height of 6-year-olds.

For the younger children, the picture is more varied. Four of the children are tall (ERI and MAR) or of average height (PAU and LAB), one girl (RIC) has to be considered small, with 75% of the SWS-girls of the same age being taller. The weight of these children is usually either one percentile-level higher or lower than

the height, e.g. LAB is of average height but weighs less than 75% of the girls of the same age. ERI is the only child who weighs considerably less compared to his height, which is at the upper end of the scale for his age group.

Still, the tallest child (JAC) was about 25% taller than the smallest one (RIC), and almost twice as heavy (26.5 kg) as the lightest child, MAR, with 14.7 kg.

5.1.2 Fundamental frequency

Fundamental frequency in other studies was often extracted from sustained isolated vowels. As many of the children in the present study were younger than the ones who participated elsewhere, they could or would not produce these vowels "at a comfortable pitch and level"(Nairn, 1997). Since authors usually do not specify how exactly they elicited this isolated material, another means of extracting F0 had to be chosen. The single words proved to be unsuited as often there appeared to be too much pitch variation on them; possibly due to the fact that with one or two syllables, they were quite short. F0 was therefore extracted from the spontaneous speech material elicited via the *Gruffalo*-story. It would have been desirable to extract several values of fundamental frequency from each child to be able to conduct inferential statistics. However, as the spontaneous passages of the children were quite varied and punctuated by a lot of vocal but non-verbal material, it was best deemed viable to attain pitch by an overall measurement of the complete passage.

Comparing ten values of mean fundamental frequency in a sample of children has to be done with caution. Boys and girls differed by 10 Hz, the younger children from the older by about 20 Hz. This might show a tendency observed with other measurements as well, namely that age in the present study is a stronger divide than sex. However, the inter-individual differences are usually quite strong

and could easily reveal a picture opposite of the tendencies observed when calculating a mean.

As mentioned before, the child with the highest F0 was a 5-year-old boy (SEA). He was the most active and agitated child, who would e.g. stand up during the recounting of the story and try to act out parts of it. The reason for SEA's higher F0 might be that due to his being more active, his breathing was stronger. This induced a higher subglottal pressure, which in turn caused the vocal folds to oscillate faster. If his F0 measurement is exempted from calculating the mean, the difference in fundamental frequency between the two age groups increases to 33 Hz, with the older children displaying a mean fundamental frequency of 266 Hz.

To interpret mean F0 results correctly, it would therefore be indispensible to conduct controlled intensity measurements. At least for the population of the present study, it would have been rather difficult to gather e.g. loud, modal and soft utterances, and it is doubted whether this is at all possible for 3- to 4-year-olds. The children were indeed sometimes asked to repeat a stimulus in a softer fashion, if their first attempt had been to loud and agitated, which happened several times. When asked to do so, the older children were able to speak softer, but they would not understand this instruction as to generally lower their volume for the rest of the task. It is assumed that they had no understanding of how loud they actually spoke and could therefore probably also not control their volume.

Data about child fundamental frequency in older literature are usually a lot higher than the values found in the present study. Neppert and Pétursson (1986) claim that fundamental frequency in children lies between 350 and 500 Hz, although he admits that "genauere Angaben für Kinder bedürften einer Spezifizierung nach dem Lebensalter"[1]. Laver (1994) quotes Fant(1956), stating that "average values for fundamental frequency in conversational speech in European lan-

[1]This translates to:"more detailed accounts for children would involve a specification according to age"

guages are approximately [...] 330 Hz for children about ten years old". These values lead to the conclusion that F0 in 3- to 5-year-olds must be generally higher than 330 Hz. However, Nairn (1997) describes an average F0 for his eighty-nine 4- to 5-year-old boys and girls of 263 Hz, which is closer to those values measured in the present study. This discrepancy might stem from several sources. The fact that Fant collected his data 40 years before Nairn might explain the different values, as F0 in children might have been higher at that time. Pinto and Hollien (1982) measured F0 values for women in 1945, 1953 and 1982. The downward shift in F0 went from 224 Hz measured in 1945 to 200 Hz in 1953 where it basically stagnated until 1982 (204 Hz). It is once again necessary to exert some caution with respect to the cultural background of the subjects, as the first study involved Australian women, the two latter American ones. Data collected by other researchers prior to the 1950s, which is summarised in de Pinto and Hollien, indicates that this downward trend can be reported from the 1920 onwards – a finding if true for children as well, might explain the fairly low values in the present study.

Another reason for differing estimates of fundamental frequency in children might be that child F0 was often not actually measured, especially not separately for boys and girls, and often not controlled for age either, but simply inferred mathematically from vocal tract measurements or estimations. A third possible reason for strongly varying accounts of speaking fundamental frequency in children could be the nature of the elicitation task. Swerts et al. (1996) write that F0 in spontaneous speech is lower than in read speech. However, as the studies cited here use either a comparable age group or a comparable method to the present one, but not both, it is difficult to come to a definite conclusion. The closest approach to the present study is probably that of Nairn (1997), who extracted fundamental frequency from sentences which, though not read, were not spontaneous either. They were elicited via pictures and were identical for all children. He, however found slightly lower values around 260 Hz for boys and girls than

the present study with 284 and 294 Hz, respectively.

5.1.3 Formant frequencies

When examining the two scatterplots in figures 3.5 and 3.6, a difference between boys and girls with respect to F1 for [a] is immediately palpable. This difference was confirmed in a MANOVA, yielding a significant interaction between sex of speaker and vowel type. Although all vowels show slightly higher values for F1 in females, this difference is more pronounced for [a] than for the other three vowels. As F1 correlates with jaw opening, the females in this study obviously produce the open vowel closer towards its cardinal quality than the boys. A similar influence vowel type on F2 could not be found, as the influence of speaker sex and vowel type was not significant for the second formant. This finding mirrors those of Sachs et al. (1973) and Whiteside et al. (2002), whose girl subjects displayed overall higher formants than the boys. Unlike Sachs' subject, however, in this study the boys'lower formants were not accompanied by a higher F0 than that of the girls, who at 294 Hz had a mean F0 value which lay 10 Hz above that of the boys.

The pattern for the interaction of age of speaker and vowel type on F1 is that of an overall decrease in the first formant from age three-four to age five. The first formant of [a] again decreases to a larger extent than those of the other three vowels, but the overall tendency reflects the findings of Busby and Plant (1995) that formant values decrease with age. However, unlike Busby and Plant, where F2 also decreased, the development of F2 takes a different direction in the present study. Here, the values for the back vowels [o] and [u] rise, the ones for [a] remain stable and the ones for [i] fall with age (cf. figure 3.10).

This phenomenon may be interpreted in two ways. Either, it is an indication

of the general tendency of the 5-year-olds in this study to centralise their vowel qualities or it reflects the fact that this group contained more boys (ALE, JAC and SEA) than girls. A claim for the first explanation might be supported by the fact that one of the girls behaved more boy-like in displaying a very low F0 value and, as will be seen in 5.2, was frequently misidentified by listeners as a boy. The second explanation would assume that, like adult males (Byrd, 1992), boys at age five already display a tendency to reduce their vowels.

It might be proposed that the general tendency of a lower F1 in the group of 5-year-olds reflects changes in anatomy, whereas the tendency to reduce F2 towards values pertaining to more central vowel qualities reflects a social accommodation.

Again, a word of caution is necessary with respect to generalising formant frequency values. Although, for example, the first formant of vowel [a] displays the largest amount of difference between the two age groups, this difference is small compared to the variability of the vowel itself. When examining [a] solely for the group of 5-year-olds, values for F1 between 798 and 1449 Hz emerge. The other three vowels also exhibit great variability, which is systematic in that the vowels with lower first formant values show less discrepancy between lowest and highest value than those with higher values.

5.2 Perception Experiment

5.2.1 General Discussion

The fact that adult listeners are able to correctly identify children's sex above chance level from speech samples alone, a finding reported in several earlier studies for different languages, was replicated for a group of German listeners, evaluating the speech of German kindergarten children. In 62% of the cases, gender

judgement was correct, with men performing better overall. However, the pattern of average CIR is different for males and females. Whereas in the group of men, all performed close to the mean recognition rate, in the group of women, some performed extraordinarily well, while others barely exceeded the chance level. Interpersonal differences are certainly worth some consideration and will be discussed in more detail below.

As to other general trends, older children were generally better identified than younger ones, who with 55% overall correct identification are closer to chance level. This mirrors the remarks of several listeners, who reported that at times they were able to say that the voice was very young, but this in turn then made it difficult to decide which sex the speaker had. It was also reported that they felt inclined to respond girl because “the voice sounded so high”. Older boys were best identified, followed by older girls, younger girls and younger boys. Young age is therefore detrimental to the correct identification of male voices, or positively phrased, age in males strongly enhances correct identification. Since the young boys were frequently misjudged as girls a clear gender differentiation at this age is problematic. However, as the group of younger males contained only two subjects, conclusions remain tentative. At the age of 5 years, correct identification increases, implying that at this age, assigning correct gender becomes more and more manageable.

The present study was also able to show that correct identification rises with utterance length. This is in accordance with most of the literature, where correct identification rates for vowels or, as in the case of Moore (1995) short syllables, were low, compared to longer utterances. Only Nairn (1997) found that while correct identification was significantly different for vowels compared to the other two conditions, the difference was not significant between sentences and spontaneous passages. In the present study, the difference between vowels and words is rather small (60% versus 62%) and only the spontaneous passages are decidedly better identified. It is reasonable to assume that the amount of accessible

phonetic information is similar in both vowels and words, but larger in sentences, since differences in accentuation or rhythm presumably do not surface on single words. Hence, the more information on prosody and voice quality, the better the correct identification. Nairn achieved CIR of 76% on sentences and slightly less on spontaneous speech, values basically equal to the 75% CIR listeners achieved on spontaneous passages in the present study.

However, a possible training effect in the present study cannot be excluded. Listeners might have performed worse in the spontaneous speech condition, if it had been offered first instead of last. Nonetheless, this training effect can at least not be found for the two listeners RK and GS, who performed worse on the spontaneous speech samples than on vowels.

Howard (2002) notes that creaky, breathy or harsh voice qualities are often produced by young children in situations where they feel shy, often accompanying a reluctance to perform. When first examining the data, all three voice qualities were found in the speech samples of all children. Although no comprehensive instrumental analyses were conducted with respect to voice qualities, the auditory impression as well as obvious differences in the signal, such as irregular periodicity accompanied by F0 around 70 Hz for creaky voice, were noted and taken as sufficient indicators for making preliminary assumptions. Irregular periodicity, sometimes accompanied by differences in amplitude for each period was taken as an indication of harsh voice, if fundamental frequency was higher than would have been expected for creaky voice (cf. Laver 1980 for a comprehensive description). Creaky voice was for example found for the female RIC, who was nonetheless identified well, and breathy voice for PAU, one of the younger boys. The 5-year-old girl ALS displayed several instances of harsh voice. As harsh voice is associated with low fundamental frequency and breathy voice primarily with the female sex, this might be another possible explanation for the poor recognition rates for PAU and ALS. In personal statements after the perception test, several listeners reported thaving used voice quality as a cue for their decisions.

The fact that the vowels in the present study were not sustained isolated ones but rather syllables, as they contained transitions from the preceding consonant, did not raise correct identification rate in comparison to other studies. While vowels in the present study received 60% correct identification, Bennett and Weinberg (1979) e.g. report a CIR on vowels around 65%, similar to Sachs (1975) reporting on the results of their previous study with 66% (Sachs et al., 1973), whereas Günzburger et al. (1987) found a slightly lower value of 55% for their normally sighted adults. All three groups of researchers used sustained isolated vowels. Using vowels including transitions was therefore a good substitute.

The type of vowel did not appear to have an influence on the correct identification rate, since the rates were quite similar across vowels. The vowel [a] received a fairly low recognition rate, which is contrary to the findings in the literature, where it was suggested that as the open vowel [a] best reflects anatomical differences in the vocal tract that might exist between boys and girls prior to puberty, this vowel should cause less confusion as to what sex the speaker of the stimulus had. It was therefore deemed feasible to evaluate possible differences for type of vowel separately for boys and girls, even though vowel type did not seem to have an influence on mean correct recognition. It emerged that the correct identification rate for [i] was similar for both sexes, whereas [o] and [u] received slightly higher rates for boys. The vowel [a] displays the largest difference between the sexes. Whereas for boys, [a] is the only vowel that receives correct identification rates below chance level, this vowel obtains highest correct identification rates in girls. A tentative explanation might be that the boys in this study did display gender specific behaviour but rather that of the opposite gender. If e.g. their voices were rather breathy (the boys studied by Robb and Simmons (1990) had breathier voices than the girls due to less vocal fold contact) and breathiness is taken as a cue for femaleness, the poor results might be explicable. As no analyses for voice quality were conducted, these statements must remain speculation at this point.

5.2.2 Individual listener differences

Four listeners, two males (RK and TB) and two females (GS and SJ) performed best on the single vowel stimuli; two of them (RK and GS) exhibited very low overall correct identification rates. This is surprising, as the general trend stipulates that the longer the utterance, the better the recognition. Particularly for the present study this was assumed to be the case, as vowels and words are very short and do not provide a lot of the cues, such as intonation or pausing, which might be found in spontaneous passages or in the sentences other researchers elicited. These two listeners, however, might have been constrained by what Nairn (1997) called *additional cognitive load* (cf. Chapter 2). Both RK and GS reported the vowel task to be different from the other two in that it was unclear what stimulus to expect next in the word and spontaneous speech condition. Since the vowels were ordered according to quality, RK and GS (and possibly TB and SJ) might have had an advantage in this task, as they only needed to detect acoustic features; and they were possibly distracted or additionally strained in the other two conditions, where they had to process content information as well. This finding also suggests that listeners might use different strategies to come to gender judgements and that the four mentioned here might rely more heavily on acoustic information for reaching their conclusions, whereas others need longer material.

All listeners reported difficulties with one or the other subtask, but many also stated that in some cases they were absolutely certain to have ticked the correct sex. Some even indicated this on the answer sheet. Most listeners were certain to have performed better the longer the utterance was, while others stressed that more material did not give them more confidence in their choice – independent of how they actually performed.

Other authors have found that listeners perform better on recognising one of the sexes; usually, boys were better identified than girls. In the present study,

however, girls received marginally higher correct judgement rates than boys, and several listeners indeed tended to respond girl. As the tendency towards responding female was more pronounced in male listeners, a cross-sex preference would have to be assumed.

5.2.3 Individual speaker differences

As the results indicate, the group of children best identified (71%) were the five year old boys, followed by 3- to 4-year-old girls (64%). This is not surprising, as the younger voices were generally higher pitched and higher pitch is usually associated with female sex. The fact that one of the small boys, PAU, was overwhelmingly misidentified as a girl while one of the younger girls, MAR, was several times mistaken as a boy, corroborates this finding, since PAU displayed a higher fundamental frequency than MAR.

Similarly for the group of 5-year-olds, the boys with the lowest and the girl with the highest fundamental frequency (ALE, JAC and JOS) were best identified. ALS, the girl with the second lowest fundamental did not reach a correct identification rate above chance level, and was therefore more often mistaken for a boy than correctly identified as a girl. SEA, the boy with the highest fundamental frequency of all children, was indeed very often misidentified as a girl, namely in 52% of the cases.

When trimming SEA's data for the perception test, the amplitude had to be strongly reduced, as he had spoken very loudly into the microphone. As the literature reports, if a child is perceived as being loud, it is usually assumed that he must be a boy. Günzburger et al. (1987) used the three boys and girls best identified with respect to their actual sex and asked their listeners to rate them on bipolar scales such as dull — clear, loud — soft. It turned out that two of the

boys were consistently rated as loud, but none of the girls was. Günzburger et al.'s data has to be handled with care, as listeners might have been influenced by their preconceptions to not mark a voice that sounds like a boy's with the label *soft*, rather than by what they actually heard. In the case of SEA, amplitude reduction possibly did not completely mask the underlying strong vocal effort which might have been perceived as loudness by listeners.

This indicates that F0 does not seem to be the only cue to correct sex identification. PAU, although displaying a much lower fundamental frequency than SEA, was more often misidentified than the latter. If F0 were the only cue, SEA with an F0 of 335 Hz should have been identified worst, but with 48% correct identification rate, he is much closer to chance level than PAU with 39%.

A complete evaluation of the inter-individual differences between least and best identified boys and girls is beyond the scope of this study. However, after it emerged that interindividual differences were rather pronounced, the formant values of the least and best identified boys and girls were evaluated on a preliminary basis. Line diagrams were produced with F1 and F2 values plotted for each vowel. Different speaker groups (old and young males and old and young females, boys and girls) showed a picture similar to that in figure 5.1 for the formants of 5-year-old boys.

Two of the worst identified children, PAU and ALS conformed to this pattern, as well as the two best identified children. SEA, however, was the only one who displayed a different pattern which is displayed in figure 5.2. As these are very informal data, they simply serve to propose that while SEA was possibly frequently mistaken as a girl due to his high fundamental frequency, he also displayed a behaviour different from all other children with respect to formant frequencies.

A final remark is to be made on intensity. One can say that in preparing the data for the perception test, as the older children were louder, their samples generally needed to be tuned down, whereas the younger voices usually had to be

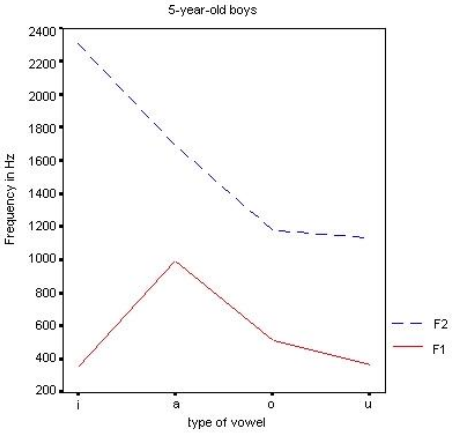

Figure 5.1: Diagram of first and second formant for the vowels [i], [a], [o] and [u] for 5-year-old boys.

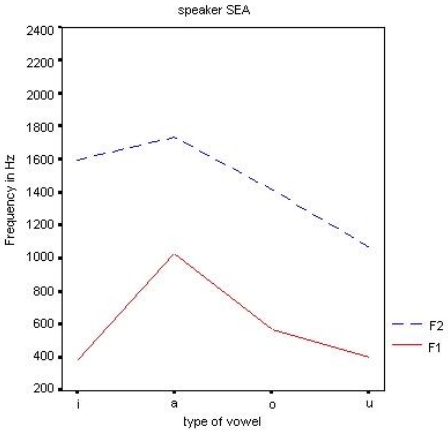

Figure 5.2: Diagram of first and second formant for the vowels [i], [a], [o] and [u] for speaker SEA.

boosted in amplitude to be comparable in loudness to the older voices. However, interindividual variability was high. Whilst ALS, a 5-year-old girl spoke very softly, the samples of 4-year old MAR had to be reduced in amplitude. Interestingly, ALS was not perceived well with respect to sex, although softness is a label listeners in the study of Günzburger et al. (1987) attributed to girls only. This might have been due to the fact that ALS also spoke in a very monotonous fashion, which could be seen with the pitch tracking function in PRAAT. Both Günzburger et al. and Bennett and Weinberg (1979) have reported on the effect of monotonicity. In the Dutch study, only the male speakers were consistently perceived as monotonous, whereas Bennett and Weinberg found that monotonicity is detrimental to the correct identification of girls' voices, as rates for it went down on sentences produced with a level pitch compared to those with normal pitch modulation.

6 Conclusion and Outlook

The aim of the present study was to evaluate whether pre-adolescent German children exhibit differences in certain acoustic parameters of their voices on the one hand side and irrespective of this being the case, whether adult listeners are able to assign correct gender judgements to these voices.

In adult voices, fundamental frequency (F0) is assumed to function as primary cue in the correct identification of gender. The difference in mean F0 of about 100 Hz between men and women is said to originate from anatomically different laryngeal dimensions which develop during puberty. For pre-pubertal children, it is therefore assumed, that the sexes do not differ with respect to larynx size and weight, assuming that this results in F0 differences which, if at all, are not significant. The first hypothesis to be tested was therefore, whether children between three and five do display sex dependent differences in height, weight and fundamental frequency. As there were too few values, no inferential statistics were conducted but descriptive statistics suggest that girls are only slightly smaller and lighter than boys. As to mean fundamental frequency, boys and girls differed by 10 Hz, a difference which in all likelihood is not significant, as the variance of F0 within each sex is quite large.

Since the kindergarten children studied here spanned an age range of more than two years, a regrouping according to age (3- to 4-year-olds versus 5-year-olds) was carried out and height, weight and F0 were again compared. The hypothesis assumed, that children would display differences in the aforementioned measurements, which could be confirmed. Older children were on average 7 kg heavier and 12 cm taller than younger ones. The fundamental frequency was as expected lower in older children, but the difference of 20 Hz is again marginal, considering the variance within each age group.

Furthermore, the frequencies of the first two formants were examined. It was expected that children display differences with respect to formant frequencies in their vowels depending on sex and age. This hypothesis could be confirmed for these two main effects for the first formant (F1), but not for F2. As both sexes displayed lower F1 values with age, this was taken as an indication of larger vocal tracts in the older group. The interaction between age of speaker and vowel type was found to be significant for both formants. However, whereas F1 decreased with age for all four vowels, F2 did not follow this general trend, but its values appear to centralise.

As the literature in this field generally reports that adult listeners are able to correctly identify the voices of children with respect to their sex, a listening experiment was administered with twelve adults judging isolated vowels, words and spontaneous passages spoken by ten children.

The hypothesis was confirmed that adult listeners perform above chance level on this task. Similarly, it was verified, that in this study, contrary to most findings in the literature, adult males performed marginally better than females. Correct identification rate (CIR) did not differ with respect to speaker sex, but a difference in mean CIR for age could be confirmed, in that 5-year-olds were generally better identified than 3- to 4-year-olds.

Mean CIR differed depending on the type of stimulus. The hypothesis, that the longer the utterance the higher the correct identification rate, was confirmed. In general, vowel type did not show any differences with respect to mean correct identification rate, as all vowels were correctly identified between 58% and 65% of the time. However, when each sex was analysed seperately, it became evident that the vowel [a] displayed strong variations in CIR. While this vowel was best identified in girls (70%), it was the one worst identified in boys (46%).

Although the results mirror those of previous studies to a certain extent, it has to be stressed that the data also exhibited a large amount of variance. While some

children were consistently identified correctly, other were consistently misidentified and it is not always certain, what parameters are responsible for these differences.

As this study had a very broad focus, it raised at least as many questions as it was able to answer. For further research, it would first be necessary to enlarge the group of children studied. With only two males in the group of 3- to 4-year olds and two girls in the older group, mere tendencies could be expressed. The material for seven more children around the age of four has already been collected and digitally stored. With a larger group of subjects, statistic tests can be conducted that could serve to support the tendencies found so far.

As a first step, the relationship between F0 and formants would have to be investigated to see whether the children in the present study exhibited patterns of formants or F0 that serve to explain differences found in the sexes and possibly why certain speakers were better identified than others. A direct comparison between F0 and formant frequencies was not possible in the present study, as F0 variation was too strong on the isolated words from which the formants were extracted. However, as authors such as Sachs et al. (1973) noted, it is not merely F0 or formants by themselves but the pattern of for example high F0 and low formants in boys and low F0 combined with high formants in girls, it would be interesting to see what pattern of fundamental and formant frequencies the boys and girls of the present study displayed.

In the present study, for each vowel mono- as well as bisyllabic words were elicited. It was beyond the scope of this work to evaluate the number of syllables as yet another possible factor of influence, but the data could be examined with respect to this parameter in the future.

The subject of expert versus lay judges was briefly touched on in this work. Since so called experts such as kindergarten teachers usually perform worse on the task of identifying children's voices with respect to gender, women, however,

usually do better than men, it would be interesting to gather a group of listeners who equally represent sex and expertise and thereby control for the different influence these parameters may have on correct identification.

Bibliography

Awan, S. N. and Mueller, P. B. (1992). Speaking fundamental frequency characteristics of centenarian females. *Clinical Linguistics and Phonetics*, 6(3):249–254.

Bennett, S. and Weinberg, B. (1979). Acoustic correlates of perceived sexual identity in preadolescent children's voices. *Journal of the Acoustical Society of America*, 66(4):989–1000.

Bezooijen, R. v. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech*, 38(3):253–265.

Boersma, P. and Weenik, D. (2004). Praat: doing phonetics by computer. `http://www.fon.hum.uva.nl/praat/`.

Borden, G. J., Harris, K. S., and Raphael, L. J. (1994). *Speech Science Primer – Physiology, Acoustics, and Perception of Speech*. Williams and Wilkins, 3rd edition.

Borjazin, S. (2003). Stimme Störungsbilder Therapieaufbau. `http://www.logopaedie-bochum.de/stimme_stoerungsbilder_therapieaufbau.html`.

Bürger, B. (2002a). Männliche Pubertät. `http://www.netdoktor.at/Sex_Partnerschaft/fakta/pubertaet_boys.shtml`.

Bürger, B. (2002b). Weibliche Pubertät. `http://www.netdoktor.at/Sex_Partnerschaft/fakta/pubertaet_girls.shtml`.

Busby, P. A. and Plant, G. L. (1995). Formant frequency values of vowels produced by preadolescent boys and girls. *Journal of the Acoustical Society of America*, 97(4):2603–2606.

Byrd, D. (1992). Preliminary results on speaker dependent variation in the TIMIT database. *Journal of the Acoustical Society of America*, 92(1):593–596.

CancerWEB Project (2004). On-Line Medical Dictionary. `http://cancerweb.ncl.ac.uk/omd/`.

Clark, J. and Yallop, C. (1995). *An Introduction to Phonetics and Phonology*. Blackwell, 2nd edition.

Coleman, R. O. (1976). A comparison of the contributions of two voice quality characteristics to the perception of maleness and femaleness in the voice. *Journal of Speech and Hearing Research*, 19:168–180.

Crowther, J. (1995). *Oxford Advanced Learner's Dictionary*. Oxford University Press.

Diverse Populations Collaborative Group (2004). Weight–height relationships

and body mass index. Some observations from the Diverse Populations Collaboration.
http://ogygia.stat.fsu.edu/papers/AJPAWtHtRevisedApr04.pdf.

Edwards, J. R. (1997). Social class differences and the identification of sex in children's speech. In Coupland, N. and Jaworski, A., editors, *Sociolinguistics – A reader and coursebook*, chapter 22, pages 284–290. Macmillan Press Ltd.

ESPRIT project 1541 (1987). Sampa - computer readable phonetic alphabet.
http://www.phon.ucl.ac.uk/home/sampa/home.htm.

Fitch, W. T. and Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *Journal of the AcousticalSociety of America*, 106(3):1511–1522.

Günzburger, D., Bresser, A., and ter Keurs, M. (1987). Voice identification of prepubertal boys and girls by normally sighted and visually handicapped subjects. *Language and Speech*, 30(1):47–58.

Hasek, C. S., Singh, S., and Murry, T. (1980). Acoustic attributes of preadolescent voices. *Journal of the Acoustical Society of America*, 68(5):1262–1265.

Haselager, G. J. T., Slis, I. H., and Rietveld, A. C. M. (1991). An alternative method of studying the development of speech rate. *Clinical Linguistics and Phonetics*, 5(1):53–63.

Henton, C. (1995). Cross-language variation in the vowels of female and male speakers. In *ICPhS 95 Stockholm*, pages 420–423.

Henton, C. (1999). Where is female synthetic speech? *Journal of the International Phonetic Association*, 29(1):51–61.

Henton, C. G. and Bladon, R. A. W. (1985). Breathiness in normal female speech: Inefficiency versus desirability. *Language and Communication*, 5:221–227.

Howard, D. M. (2002). Quantifying developmental singing voice changes in children.
http://www.med.rug.nl/pas/Conf_contrib/Howard/Howard_child_voice.pdf.

Karlsson, I. (1987). Sex differentiation cues in the voices of young children of different language background. *Journal of the Acoustical Society of America*, 81:68–69.

Kent, R. D. (1976). Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of Speech and Hearing Research*, 19:421–447.

Kiesler, K. (2004). Phoniatrie Graz – Stimmstörung und Heiserkeit.
http://www.kfunigraz.ac.at/phoniatrie/diagnostik%20und%20therapie/heiserkeit/2.7.htm.

Lass, N. J., Hughes, K. R., Bowyer, M. D., Waters, L. T., and Bourne, V. T. (1976). Speaker sex identification from voiced, whispered, and isolated vowels. *Journal of the Acoustical Society of America*, 59(3):675–678.

Laver, J. (1980). *The Phonetic Description of Voice Quality*. Cambridge University Press.

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

Lippincott, Williams, and Wilkins (2004). Stedman's Online Medical Dictionary.
http://www.stedmans.com/section.cfm/45.

Mattingly, I. (1966). Speaker variation and vocal-tract size. *Journal of the Acoustical Society of America*, 39(6):1219.

McRoberts, G. W. and Best, C. T. (1997). Accomodation in mean F0 during mother–infant and father–infant vocal interactions: a longitudinal case study. *Journal of Child Language*, 24:719–736.

Merriam-Webster (2004). Merriam-Webster Online.
http://www.m-w.com/.

Montenegro, M. C. B. (2003). Social class, sex stereotypes and children's speech in the Philippine setting. *Tanglaw*, 9:14–28.

Moore, K. (1995). Identification of male and female voice qualities in prepubescent children. *Proceedings of the International Conference of the Phonetic Sciences (ICPhS)*, pages 298–301.

Nairn, M. (1997). *Acoustic and perceptual gender differences in the speech of 4.5 to 5.5 year old children*. PhD thesis, Queen Margareth University College Edinburgh.

Neppert, J. and Pétursson, M. (1986). *Elemente einer akustischen Phonetik*. Helmut Buske Verlag, 2nd edition.

North American Menopause Society (2003). Menopause guidebook.
http://www.menopause.org/Misc/guidebook.htm.

Pinto, d. O. and Hollien, H. (1982). Speaking fundamental frequency characteristics of australian women: Then and now. *Journal of Phonetics*, 10:367–375.

Pompino-Marschall, B. (1995). *Einführung in die Phonetik*. de Gruyter.

Robb, M. P. and Simmons, J. O. (1990). Gender comparisons of children's vocal fold contact behavior. *Journal of the Acoustical Society of America*, 88(3):1318–1322.

Sachs, J. (1975). Cues to the identification of sex in children's speech. In Thorne, B. and Henley, N., editors, *Language and Sex: Difference and Dominance*, chapter 10. Newbury House Publishers: Massachusetts.

Sachs, J., Lieberman, P., and Erickson, D. (1973). Anatomical and cultural determinants of male and female speech. In Shuy, R. and Fasold, R., editors, *Language attitudes: Current trends and prospects*, pages 74–84. Georgetown University Press.

Schwartz, M. F. and Rine, H. E. (1968). Identification of speaker sex from isolated, whispered vowels. *Journal of the Acoustical Society of America*, 44(6):1736–1737.

Swerts, M., Strangert, E., and Heldner, M. (1996). F0 declination in read-aloud and spontaneous speech. speech, music and hearing. http://www.asel.udel.edu/icslp/cdrom/vol3/205/a205.pdf.

Tanner, J. M. (1962). *Wachstum und Reifung des Menschen*. Georg Thieme Verlag.

Titze, I. R. (1985). Physiologic and acoustic differences between male and female voices. *Journal of the Acoustical Society of America*, 4:1699–1707.

Traunmüller, H. and Eriksson, A. (1993). The frequency range of the voice fundamental in the speech of male and female adults. Manuscript.

Webnox Corporation (2004). Hyperdictionary – medical dictionary. http://www.hyperdictionary.com/medical.

Weinberg, B. and Bennett, S. (1971). Speaker sex recognition of 5- and 6-year old children's voices. *Journal of the Acoustical Society of America*, 50:1210–1213.

Whiteside, S. P. and Hodgson, C. (1998). The development of fundamental frequency in 6- to 10-year-old children: A brief study. *Journal of the International Phonetics Association (JIPA)*, 28(1&2):55–62.

Whiteside, S. P. and Hodgson, C. (1999). Acoustic characteristics in 6–10-year-old children's voices: some preliminary findings. *Logopedic Phonetic Vocology*, 24:6–13.

Whiteside, S. P. and Hodgson, C. (2000). Speech patterns of children and adults elicited via a picture naming task: An acoustic study. *Speech communication*, pages 267–285.

Whiteside, S. P., Hodgson, C., and Tapster, C. (2002). Vocal characteristics in pre-adolescent and adolescent children: a longitudinal study. *Logopedic Phonetic Vocology*, 27:12–20.

Yamazawa, H. and Hollien, H. (1992). Speaking fundamental frequency patterns of Japanese women. *Phonetica*, 49:128–140.

Zabransky, S. (2003). Saarländische Wachstumsstudie. http://www.wachstum-ipep.de/Literatur/SWS/SWS-Inhalt.html.