SCUOLA INTERNAZIONALE
SUPERIORE di STUDI AVANZATI
International School
for Advanced Studies

Acoustic cues for body size: how size-related features are used and perceived

by

Linda Sebastianutto

Supervisor

Professor Evan Balaban

Doctoral thesis submitted in partial fulfillment of the requirements for a PhD degree

Academic Year 2007-2008

Acknowledgements

Many people helped me during my years in SISSA and afterwards, and contributed to opening my mind about Science with great discussions, help, and shared experiences.

First of all I am very grateful to my supportive supervisor, Professor Evan Balaban, for his great help, kindness and constant presence - in spite of the great distance - during my PhD project.

During these years I have had the privilege of meeting several brilliant people, with whom I shared thoughts, experiences, friendship. My lab colleagues and friends Valeria Caruso, Katya Vinnik, Mahan Azadpour, and fellow SISSA inhabitants Cristiano Crescentini, Valentina Daelli, Gioia Negri, Liuba Papeo, Claudia Civai, Silvia Benavides, Antonio Vallesi, Roma Siugzdaite, Alessio Isaja. Thanks to Professor Raffaella Rumiati who supervised me during my first year, and Professor Tobias Riede, who kindly provided me with stimuli for the experiment on dog growls.

I am infinitely grateful to Damir Kovačić and Ana Bedalov for making the two months spent in Zagreb one of the best periods of my life. I would also like to mention Duska Peric and Natasha Claric, thanks to whom I could test Croatian children in several schools, even though I could not speak one word of Croatian.

Special thanks to all of the teachers in the different kindergartens I worked with, and to Caterina Spiezio and the Parco Natura Viva staff, which welcomed me and helped during my macaque experiments.

Finally, I'd like to thank my family and *my* Fabio, for his love, infinite patience, and support during the difficult moments, and laughs during the many happy moments.

Table of Contents

1 INTRODUCTION

We live in a noisy world. There is no place on the Earth where it is possible to have the experience of complete silence, not even the deepest place in the ocean. Billions of living and nonliving objects around us produce sounds, which are extremely different in their physical structure. Some of these sounds are noisy, some are harmonic, some are continuous, others are impulsive, soft, loud; the sound environment contains an infinite combination of all these characteristics and more. Evolving in such an environment has resulted in a human auditory system that is able to extract useful information from sounds. We are able to say whether a sound source is still or moving (and in the latter case, the direction of movement), what kind of object produced the sound, and the meaning of the message if the perceived sound is an intentional communicative signal. When we hear someone's voice, for instance, we are able to extract useful information about talker identity apart from the meaning of the heard words. This thesis focuses on a particular kind of information that can be extracted from an acoustic signal: the apparent s*ize* of the sound-producing object.

1.1 SOUND PRODUCTION

1.1.1 MECHANISMS OF PHONATION

In humans, mammals in general, and other terrestrial animal species, sound is produced by exciting resonant cavities in the body with a stream of acoustic pulses.

In mammals sound is produced during an expiratory action that forces air to pass through the larynx. Sounds are produced by a repeating sequence of events. During *adduction*, the vocal cords are brought together, temporarily blocking the flow of air from the lungs and leading to increased subglottal pressure. When the subglottal pressure becomes greater than the resistance offered by the vocal folds, they open again. The folds then close rapidly due to a combination of factors, including their length and elasticity, and laryngeal muscle tension.

If the process is maintained by a steady supply of air, the vocal cords will continue to open and close in a quasi-periodic fashion. As they open and close, puffs of air flow through the glottal opening, producing a harmonic sound with a fundamental frequency (F0) that depends on the frequency of this cyclic movement (hereafter called glottal-pulse rate or GPR). The GPR contributes to the pitch of the sound.

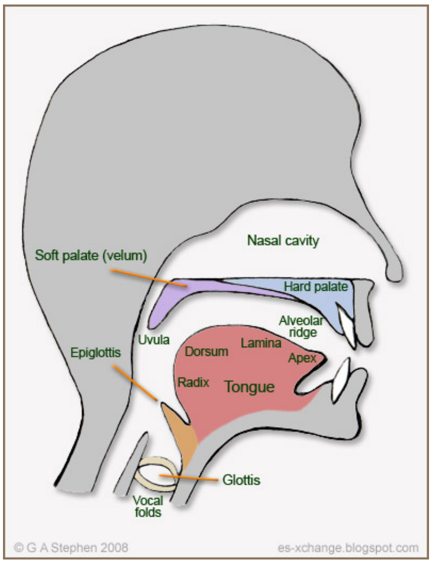

Fig 1 Structure of the human vocal apparatus

Several structures are involved in phonation (Fig 1): the harmonic sound produced at the level of the larynx passes through the supralaryngeal vocal tract (comprising the tracheal, the oral and the nasal part), which, as a time-varying acoustic filter, suppresses the passage of energy at certain frequencies while allowing energy at other frequencies to pass or even to be amplified. As a consequence, the output sound has some of the source frequencies attenuated while others are not, resulting in bands of resonant energy that in the case of speech are called *formants*. Formants are those frequencies at which local energy maxima are sustained by the supralaryngeal vocal tract and are determined, in part, by the overall shape, length and volume of the vocal tract. Changing the size and the shape of the vocal tract length, which is constrained by both bony and soft structures, determines many of the phonological properties of produced sounds (Fig 2). The flexibility of the human vocal tract, in which the articulators can easily be adjusted to form a variety of shapes, results in the potential to produce a wide range of sounds.

Several other sounds, such as those produced by musical instruments, are generated in a similar way (i.e. air passing through a vibrating structure, which produces harmonic energy peaks that are filtered through a series of tubes), and their spectra show resonances analogous to mammalian vocalizations.

The most commonly-accepted general model for speech production is the "source-filter theory of speech production" (Peterson and Barney 1952; Fant 1960) and stems from the experiments of Johannes Müller (1801-1858), who tested the mechanisms of phonation by blowing air through larynges excised from human cadavers. He obtained a vowel-like sound only when he placed a tube over the vibrating cords whose length was roughly equal to the length of the airways that normally intervene between the larynx and a person’s lips (Lieberman 1984).

In this model the source of acoustic energy is at the larynx, while the supralaryngeal vocal tract serves as a variable acoustic filter whose shape determines the phonetic quality of the sound.

In normal speech, fundamental frequency changes constantly, providing linguistic information, and information about a speaker’s emotional content, such as differences in speaker mood; in the case of tonal languages such as Mandarin, semantic information, is also provided by fundamental frequency changes.

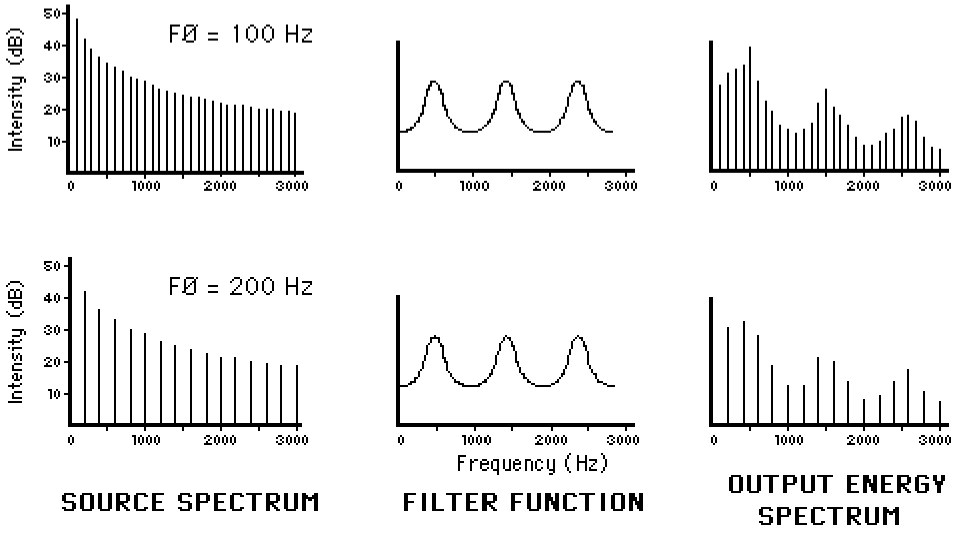

Fig 2 The source-filter model of speech production. The source spectrum represents the spectrum of typical glottal air flow with a fundamental frequency of 100 Hz. The filter, or transfer function is for an idealized neutral vowel, with formant frequencies at approximately 500 Hz, 1500 Hz and 2500 Hz. the output energy spectrum shows the spectrum that would result if the filter function shown here was excitedby the source spectrum shown at the left. Source: www.haskins.yale.edu

1.1.2 RELATION BETWEEN PRODUCED SOUND AND BODY SIZE

1.1.2.1 Humans and speech

The morphology of the vocal tract is extremely important in phonation: while the length and dimensions of the vocal folds are related to the fundamental frequency F0 of the produced sound, the length and all the possible shapes of the vocal tract are responsible for the timbre of each voiced sound.

Each of these physical structures changes as an individual grows: the vocal cords lengthen and become thicker, while the vocal tract lengthens as well. As a consequence, the frequencies associated with voiced sounds also change: both F0 and formant values decrease as the speaker grows larger.

The development of vocal tract morphology has been described in detail by a cross-sectional study on 129 people of both sexes ranging from 2 to 25 years of age (Fitch and Giedd 1999). This study confirmed a positive correlation between body size and the physical dimensions of the vocal tract. Sex differences in vocal tract length are not apparent until puberty, when in males, the descent of the larynx lengthens the vocal tract, especially in the pharyngeal region. The human larynx grows independently from the rest of the skull.

The acoustical implications of these morphological changes are that children's voices have a higher fundamental frequency and formant frequencies compared to adult voices, and that adult women have a higher F0 and formants compared to men (Titze 1989). In a cross-sectional study on 436 children between 5 and 18 years and 56 adults, Lee and colleagues (1999) confirmed these implications from anatomical data. F0 differences between sexes arise only after age 12. In males only, the mean F0 during puberty drops by 78%, with no significant population changes occurring after 15 years of age. In females, a drop in F0 was observed only between ages 7 and 12. Moreover, a constant and general decrease in formant values for both males and females reached adult levels around the age of 15 for males and 14 for females.

Therefore, sounds like speech convey important information related to the physical characteristics of the speaker, which are referred to as indexical cues (Ghazanfar et al. 2007).

1.1.3 THE IMPORTANCE OF SIZE-RELATED ACOUSTIC FEATURES

Several studies on animal behavior show that body size is an extremely important feature during fitness-related activities such as mate choice or agonistic interactions, where giving the impression of a larger body size can have a significant influence for a positive outcome (Bradbury and Vehrencamp 1998; Reby et al. 2005; Charlton et al. 2007).

In humans, size-related features are modulated to change the attractiveness of the talker (Feinberg et al. 2005), and / or to emphasize femaleness or maleness (Hasek et al. 1980). More controversially, it has been claimed that size-related features are manipulated to implicitly signal fertility (Bryant and Haselton 2008).

Understanding how indexical cues are extracted by the auditory system before semantic processing could also shed light on the phenomenon of speech normalization, in which an utterance pronounced by different speakers is perceived as the same by listeners, despite great acoustical differences in the perceived signals.

Several studies on animal species show how, with increasing body size, vocalizations become more low-pitched. Among primates, allometry of calls was described in rhesus monkeys (*Macaca mulatta*) and pigtail macaques, (*Macaca nemestrina*) (Hasek et al. 1980; Gouzoules and Gouzoules 1989, 1990; Hammerschmidt and Fischer 1998; Feinberg et al. 2005; Rendall et al. 2005; Bryant and Haselton 2008), but these structural changes in vocalization have not been very precisely quantified, with the exception of the “coo” call of rhesus monkey. There will be a more detailed review of the literature about size-related acoustic features in mammalian sounds at the beginning of Chapter 2 of this thesis.

1.1.4 ARTICULATION AND THE DESCENDED LARYNX

Historically, fundamental frequency F0 was considered the most reliable acoustic cue for body size. While this correlation has been extensively confirmed in several fish and amphibian species, it remains controversial in mammals (Reby and McComb 2003; Gonzales 2004; Rendall et al. 2005).

The ambiguity of the mammalian relation between fundamental frequency and body size could be due to a relative independence of body size and larynx size, the latter determining vocal fold length and hence lowest fundamental frequency (Fitch 1997). In human adults, the larynx at rest is in a relatively lower (more basal) position in the vocal tract when compared to its resting position in other mammals. While non-human mammals habitually breathe with the larynx inserted into the nasal cavity, adult humans are unable to do this. Because of the presence of a descended larynx and the increased possibility this provides to actively change and control vocal tract shape, in addition to the methods that are common to all mammals (i.e. protruding lips), humans are able to produce an extremely wide range of sounds.

Human infants pass through a stage during which the larynx at rest can be inserted in the nasal passages to allow simultaneous suckling and breathing, but between three months and three years of age, the larynx recedes deep in the throat, not allowing simultaneous passage of food and air. Historically the descended larynx was considered to be uniquely human, and was thought to have evolved specifically for speech articulation and formation of different formant patterns (Fitch and Reby 2001) .

Before Fitch's initial studies, human and other mammalian larynges were studied only in their static resting position, where they are inserted between the nasal passages.. However, species such as dogs, goats, pigs, cotton-top tamarins, and deer can lower the larynx *during vocalization* in order to lengthen the vocal tract and use a wider range of possible vocal tract shapes (Fitch 2000; Fitch and Reby 2001).

The active control of vocal tract filtering, broadly analogous to human speech production, is therefore widespread in mammalian species. A study on Diana monkey alarm calls (Riede and Zuberbühler 2003) showed that active modulation of the vocal tract resulted in formant profiles conveying different semantic information.

Hypotheses about the origin of the descended larynx thus shifted from the idea of a feature evolved for speech production to the idea of a feature evolved through sexual selection for body size exaggeration. Moreover the correlation between size and vocal tract length is stronger in males than in females, which is hypothesized to be due to the larger role that body size exaggeration may play in displays related to male mating success (Fitch and Giedd 1999; Fitch and Reby 2001; Rendall et al. 2005; Charlton et al. 2007).

1.1.5 A NEW QUANTITY: FORMANT DISPERSION

In contrast to the larynx, which is independent from the rest of the skull, vocal tract length (VTL) is constrained by body size since it depends on both bony and soft structures. Thus, the distribution of resonant tube sizes that produces the formant pattern of human speech depends directly on vocal tract shape and dimensions, which could be a more reliable cue for body size than fundamental frequency. In fact, as an animal grows, the spacing between the resonant spectral energy concentrations in its voice decreases (Smith et al. 2005).

In 1997 Fitch proposed a new quantity, *formant dispersion*, as an acoustic cue that may be more reliable than F0 for body size in most mammals. This new measure was calculated on two particular types of aggressive noisy and broadband vocalizations given by rhesus macaques to an unknown observer: ‘‘pant-threats’’ and ‘‘pant-barks’’.

Formant dispersion (i.e. the average distance between formants) is defined by the expression:

$$D_f = \frac{\sum_{i=1}^{N-1} F_{i+1} - F_i}{N-1}$$

Where

D_f is the formant dispersion in Hz,

N is the total number of measured formants and

F_i is the frequency in Hz of the formant i.

In rhesus macaques (*Macaca mulatta*), there is a strong correlation between body size and D_f. Formant differences are spontaneously perceived and associated with their corresponding body sizes by rhesus macaques (Fitch 1997; Smith and Patterson 2005; Fitch and Fritz 2006; Ghazanfar et al. 2007). Similar strong correlations between body size and formant dispersion were also found in other mammalian species, including dogs (Riede and Fitch 1999), and deer (Reby and McComb 2003; Vannoni and McElligott 2008). However, one study on humans (Gonzales 2004) has found a weak relationship between formant dispersion and body size, suggesting that human VTL is only a weak predictor of body size in human adults.

1.2 SIZE PERCEPTION

Apart from speech and animal vocalizations, many other sounds contain information about the size of the object that produced them. For both animate and inanimate sound sources, bigger objects generally produce more low-pitched sounds. In a way that is similar to speech discrimination, it is possible to extract information about physical attributes of sound sources relying on acoustic cues, in particular on their spectral content (Carello et al. 1998; Houben et al. 2001; van Dinther and Patterson 2006). However, since most of the experiments about size perception use speech stimuli, studies on nonspeech sounds will be discussed in a later section of the thesis.

1.2.1 INFERRING A SPEAKER'S SIZE FROM SPEECH

Humans are able to make fine judgments about the size of speakers, even when speech stimuli are artificially manipulated so that the size information in them falls beyond the normal range. The two important cues for speaker identification appear to be both vocal pitch (F0) and vocal tract spectral shape information. Pitch, as described in the previous section, is a highly-salient cue for sex and age. It is low for adult men, high for young children and has middle values for adult women. Similarly, VTL information can be extracted from formant frequencies and their spacing: formants are low and closely-spaced for adult men, high and widely-spaced for children, while women lie in the middle. F0 and VTL cues appear to be combined together in order to identify the attributes of a speaker. Sex and age judgments have been shown to be affected by both glottal pulse rate (GPR, analogous to F0) and vocal tract length information (Smith et al. 2005; Smith and Patterson 2005; Smith et al. 2007). For unnatural combinations of GPR and VTL, VTL information appears to be dominant for speaker identification, especially for *size* evaluation; the relationship between size and GPR is highly salient *across* classes of speakers, but not *within* classes, while VTL information helps discrimination between perceived speakers even in the latter case (Smith et al. 2005; Smith and Patterson 2005; Smith et al. 2007).

It has been proposed that the auditory system can separate information about the size of the vocal tract from information about the shape of the vocal tract at an early level. The Mellin transform, developed by Cohen (1993), can segregate size information in a sound from size-invariant information. Irino and Patterson (2002) developed a two-dimensional version of the Mellin transform to simulate size information processing in speech, and they proposed that a physiological form of the Mellin transform exists in the brains of animals and people which normalizes and segregates body size information at an early stage of processing in the auditory system.

They believe this information becomes segregated at an early stage of processing because they observed that adults perform perceptual tasks using this information rather automatically (without the involvement of learning), even with stimuli whose apparent size goes well beyond the range of everyday experience (Smith et al. 2005). Using STRAIGHT, a vocoder developed by Kawahara and Irino (2004) that allows independent manipulation of GPR or VTL information, it is possible to create a large sample of utterances with all possible combinations of GPR and VTL information, falling within and beyond the normal range (Fig 3). Smith and colleagues (2005) modified GPR while maintaining VTL constant, creating a voice changing in pitch, or VTL while maintaining GPR constant, creating a voice changing in scale values and therefore in speaker size.

By asking subjects to discriminate between utterances with different combinations of these acoustic cues, it was possible to show that human adults are able to make fine discrimination of speaker size relying only on acoustic cues.

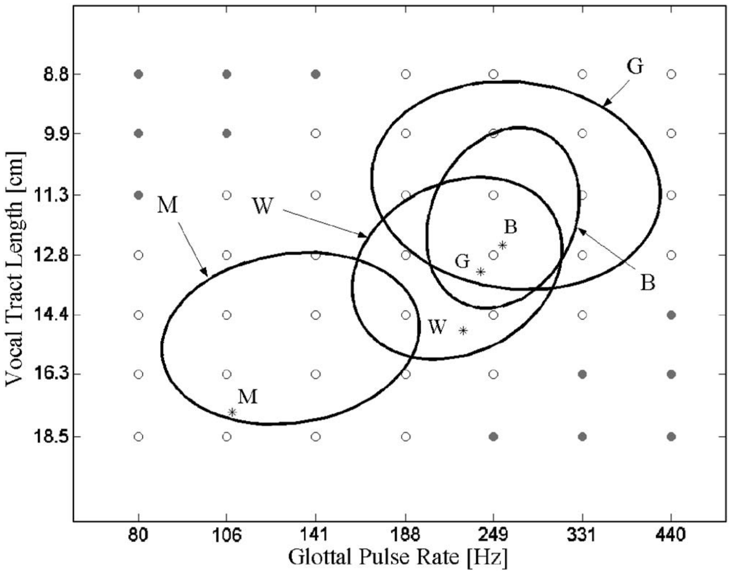

Fig 3. The VTL and GPR values used by Smith et al (2007) for size discrimination experiment. The four ellipses show the normal range of GPR and VTL values in speech for men (M), women (W), boys (B), and girls (G), derived from the data of Peterson and Barney (1952). Each ellipse contains 99% of the individuals from the respective category.

Subject performance in size assignment tasks depends upon the amount of acoustic information they receive. For instance, performance increases from syllables (4-6%) to isolated vowels (around 10%), suggesting that relatively dynamic and static cues may provide conflicting information (Ives et al. 2005). These studies have shown that both GPR and VTL are important features for size discrimination; however, the interaction between the two features in cases of discordant pitch and formant values may result in an impairment of the ability to correctly discriminate between two speaker sizes (Rendall et al. 2007).

1.2.2 NEURAL CORRELATES OF ACOUSTIC SIZE PROCESSING

What are the neural correlates of size processing? In order to answer this question, it is necessary to identify brain regions where the physical aspects of these cues are processed and may contribute to the subjective perception of size .

In terms of pitch processing, the time and frequency domain properties are thought to be processed in the primary auditory cortex (Griffiths et al. 2001), while a neural correlate of the percept has also been claimed in the lateral Heschl's Gyrus (lateral HG) in non-primary auditory cortex (Patterson et al. 2002). However, all these studies used a particular type of stimulus, Iterated Rippled Noise; two studies with MEG and fMRI (Hall and Plack 2007; Hall and Plack 2009) casted doubts on this localization, since using different types of pitch-evoking stimuli failed to produce activations in Heschel's gyrus. Rather, authors found pitch-related activity in a region behind HG: in Planum Temporale (PT). More recently Garcia, Hall and Plack (2010) found that this region (lateral PT), immediately behind lateral HG was activated by different types of pitch-evoking stimuli (i.e. a complex tone with unresolved harmonics and Huggins Picth).

For scale processing, the analysis of acoustic features (i.e. resonances) has been claimed to begin in the Medial Geniculate Body (MGB) of the thalamus (von Kriegstein et al. 2006). The activation of the MGB was detected for VTL values that were also outside the normal range, supporting the idea that the initial stage of size processing does not depend on learned associations, but rather on an early and generic mechanism for processing the scale of a sound. MGB activity was also left lateralized for speech perception. Other types of resonant cues (human voices, animal calls, and musical instruments) show bilateral activation of Superior Temporal Gyrus (STG) for spectral-envelope changes in general; responding to changes of stimulus category as well as acoustic scale. Activity in the left posterior STG was specific to acoustic scale in human voices and not responsive to acoustic scale in other resonant sources. Changes in acoustic scale across categories resulted in activation in the anterior temporal lobe and intraparietal sulcus. The authors of these studies suggest that the human voice requires special processing of acoustic scale, whereas the anterior temporal lobe and intraparietal sulcus process auditory size information independent of source category (von Kriegstein et al. 2007).

In summary, while pitch processing results in activity only at the cortical level, within the auditory cortices, either primary – Heschl's Gyrus – or secondary – Planum Temporale-, scale processing is much more "widespread" in the brain, involving activity at the subcortical level – MGB in the thalamus – and cortical areas such as the STS comprising both primary and secondary auditory cortices, and extra-auditory areas such as the anterior temporal lobe and the intraparietal sulcus.

1.3 SIZE PERCEPTION IN CHILDREN

Very few studies have addressed the development of size attribution of sounds by children, and these studies provide contradictory information.

Do children need training to extract information about the size of a source only from acoustic cues, or is this something children do naturally from an early age? Mondloch and Maurer (2004) found that 3-year-olds (30-36 months) reliably match high-pitched sounds with smaller objects (and low-pitched sounds with bigger objects); however previous work by Marks et al. (1987) did not find this ability in 9 year-olds.

1.4 AIMS OF THIS THESIS

This thesis will consider the relation between sound and size from two different points of view: the *production* and the *perception* of sound.

A naturalistic description of the relationship between sound production and size will be presented using a cross-sectional study on a colony of Barbary macaques. Their vocal repertoire will be examined to see if there is any context-dependent modulation of size-related acoustic features.

The relationship between sound and size perception will be examined by experiments that try to understand when children begin to attribute size to sound-producing objects in a manner that is similar to adults. If experience matters in developing this skill, then different categories of sounds may be associated with size judgments that have different levels of accuracy. For example, high accuracy may be shown for speech, which children have a lot of experience with, while lower accuracy levels may be shown for stimuli produced by dogs, and little accuracy may be shown for sounds produced by unfamiliar, non-biological objects. On the other hand, if the segregation of size information is an automatic process, and there is a single method used to process size information for all types of objects, there should not be major differences among the accuracy of size judgments given to different categories of stimuli. It may also be the case that size processing is automatic, but that this processing is done in a different way for different kinds of objects; the development of size percepts may therefore follow a different developmental trajectory for different kinds of sounds that is independent of the amount of experience with each of them.

2 DEVELOPMENTAL AND BEHAVIORAL FEATURES OF BARBARY MACAQUES VOCALIZATIONS

2.1 INTRODUCTION

2.1.1 PRIMATE VOCALIZATIONS: OVERVIEW OF VOCAL REPERTOIRES

Primates live in societies with complex social relationships. Like other social species, primates also use a rich repertoire of vocalizations that change according to behavioral context. Some species such as lemurs (Gamba and Giacoma 2007) need communicative signals that function over long-distances, and use a discrete signaling system, while others like macaques use structurally different vocal sounds that intergrade with each other (Peters and Ploog 1973). Primate infants learn very early during development how to respond properly to important sounds such as maternal calls (Fischer 2004).

The vocal repertoires among different species of primates have a wide frequency range and include both soft and loud calls. These sounds are typically classified by the pattern of their energy distribution across frequencies into noisy, tonal and harmonic sound types (Fitch and Hauser 1995). The mechanisms used to produce these vocalizations range from simple vibrating sources and vocal tract filters to more complex mechanisms such as biphonation (the simultaneous appearance of two different frequencies) and deterministic chaos (identified in the spectrogram of the sound as a broadband, noisy segment, with other features such as a sudden onset of the noise, preceding subharmonics or harmonic windows within the noise (Riede et al. 2004; Tokuda et al. 2002; Zollinger et al. 2008).

Several studies describe nonhuman primate vocal production from the standpoint of calls produced in particular behavioral contexts and their function, rather than the complete vocal repertoire -i.e. copulation calls in Barbary macaques, food-associated calls and recruitment screams (Gouzoules et al. 1984; Gouzoules and Gouzoules 1995; Oda and Masataka 1995; Semple 1998; Gros-Louis 2004). As one of the loudest vocalizations produced within a vocal repertoire, the “scream” has been a particular object of interest for many studies (Gouzoules et al.1984, 1998; Gouzoules and Gouzoules 1989, 1990, 1995, 2000; Owren and Rendall 2003; Siebert and Parr 2003; Slocombe and Zuberbuhler 2005, 2007; Tomaszycki et al. 2005; Fugate et al. 2008).

Primates rarely produce single screams. Screams often occur in sequences called bouts, and within each bout different sound structures can be found (Ghazanfar and Hauser 2001).

Vocalizations reflect the internal state of the caller and provide useful information to recipients. Macaques’ screams are interesting because macaques live in groups with strong matrilineal kinship ties and clear agonistic dominance hierarchies. In several macaque species combatants solicit support from allies that can even be out of sight using different “recruitment screams” (Gouzoules et al. 1984). The structure of these screams changes according to the severity of the attack and the relative rank of the opponent. Playback experiments show that recipients respond differentially to these vocalizations and for this reason it has been argued that screams have external referents that allow receivers to choose appropriate intervention tactics (Gouzoules and Gouzoules 1989, 2000; Gouzoules et al. 1984, 1998).

2.1.2 SIZE RELATED FEATURES IN PRIMATE VOCALIZATIONS

As in all mammalian species, the mechanism of primate vocal production uses a vibrating larynx that produces a harmonically complex sound and that is filtered by the upper vocal tract; however, sometimes other structures such as extra-laryngeal air sacs are involved (such as in great apes or siamang *Symphalagus syndactylus*), and often their role in the final output sound is not clear (Hewitt et al. 2002). In general, it is assumed that the larynx and the vocal tract work independently, but some source-tract interactions cannot be excluded (Fitch and Hauser 1995). It is difficult to get functional anatomical data on the precise mechanisms non-human primates use to produce sounds, especially when vocalizations of free-ranging monkeys are studied. Based on spectral vocalization features, some mechanistic models of sound production have been proposed.

Lieberman (1968, 1969) proposed a first vocal tract model resembling a uniform or flared tube (Riede and Zuberbühler 2003; Riede et al. 2005), but more recent studies showed that vocalizations were more likely the product of vocal tracts modeled as a connected series of short tubes which vary in diameter (Fitch and Hauser 1995, Owren et al. 1997;).

As in humans, both the larynx and the vocal tract change with size, hence vocalization acoustics too changes with age in a way that reflects increasing body size. Among different species body size is an important predictor of reproductive success and fighting ability.

In this dissertation we will focus only on the glottal and the laryngeal sources.

Since it is often difficult to obtain reliable direct measures of body size, this information is frequently inferred from available data on sex and age of the animal (Ey et al. 2007). The relationship between body size and vocalization (in terms of pitch and formant dispersion) is not straightforward, and many differences occur between species and, within a certain species, within/between different ages and sexes.

The general rule according to which the bigger the body size, the lower the F0 and formants shows many exceptions among primate species.

F0 and pitch are considered unreliable indicators of body size. However in some species these features significantly decrease with increasing body size (as in "coo" contact calls of rhesus monkeys *Macaca mulatta* (Hammerschmidt et al. 2000), in Hamadryas baboons *Papio hamadryas* (Pfefferle and Fischer 2006), and in the Japanese macaque *Macaca fuscata* (Inoue 1988). This relationship was not found in humans (Gonzalez 2004), or within sexes in Hamadryas baboons *Papio hamadryas ursinus* (Rendall et al. 2005).

Formant frequencies depend on the size and shape of the vocal tract, which is constrained by bony structures, and is therefore considered a better predictor for body size. Decreasing formant frequencies with increasing body size have been described in rhesus monkeys *Macaca mulatta*, pigtail macaques, *Macaca nemestrina*, (Gouzoules and Gouzoules 1989, 1990; Fitch 1997; Hammerschmidt and Fisher 1998; Hammerschmidt et al. 2000;), and in Hamadryas baboons *Papio hamadryas* (Pfefferle and Fischer 2006).

2.1.3 ACTIVE ARTICULATION

"In intentional communication, the behavior of the sender must involve a goal and some flexibility in the means for attaining it. Intentional communication specifically involves, for example, an individual using the same signal flexibly in different communicative contexts, or using different signals in the same context but slightly different situations" (Tomasello and Call 1997).

Primates in general cannot produce an infinite number of sounds, and for long time have been considered to have a rather inflexible system of sound production (Zuberbühler 2005). However, the seminal work by Owren and Bernacki (1988) on vervet monkeys (*Cercopithecus aetiops*) has shown that a certain degree of voluntary articulation change is quite common. This species produces semantic alarm calls after detection of various predators encountered in their natural environment. Two of these calls (snake and eagle alarms,) differ in spectral properties and elicit different and predator-appropriate behavioral responses in receivers.

Diana monkeys (*Cercopithecus diana*) also produce an initial formant downward transition that distinguishes between eagle and leopard alarm vocalizations. As in vervet monkeys, this formant modulation results from active control of vocal filtering used by the monkeys to encode semantic information, in a way that is analogous to human speech production (Riede and Zuberbühler 2003; Riede et al. 2005, 2006; Lieberman 2006).

Active articulation by lip protrusion, lip separation, teeth separation and mandibular position is used by non-human primates to produce different call types or to modulate the physical features of vocalizations. For example varying mandibular position changes the dominant frequency of "coo" calls in *Macaca mulatta*; lip configuration is also related to changes in fundamental frequencies of vocalizations in chimpanzees (*Pan troglodytes*) (Hauser et al.1993).

2.1.4 THE STUDY SPECIES: BARBARY MACAQUES (MACACA SYLVANUS)

Barbary macacques (*Macaca sylvanus*) are the phylogenetically most ancient species in the genus Macaca, and the only one living in Africa.

Barbary macaques live in female-bonded, multi-male, multi-female social groups. Their mating season occurs annually in autumn, and infants are born during spring. They are mainly terrestrial during the day, while during nights and resting periods they climb on trees and sleeping cliffs, and sleep in clusters.

Their societies present an extreme form of alloparental behavior: all the infants are subject to care from other members of the colony within the first days of life (Small 1990). The role of the infants is also important during male-male interactions since a baby presentation reduces the likelihood of aggression (this behavior is called “agonistic buffering”, Deag and Crook 1971).

Barbary macaques use an extremely rich vocal repertoire consisting of both tonal and noisy calls (screams, geckers, shrill barks, grunts and low-frequency panting) in a variety of behavioral contexts. Continuity occurs between different sound structures: for this reason this repertoire has been defined as a *graded signaling system* (Hammersmith and Fischer 1998; Fischer and Hammerschmidt 2002). Hammerschmidt and Fisher (1998) used a cluster analysis based on sound acoustic features to identify 7 different call types, going from noisy low-pitched barks or grunts (groups A-D) to tonal high pitched screams (types E-F-G). Animals of all ages and both sexes are potentially able to produce the whole array of call types, however there are clear age-related differences in the usage of different call types. Young animals frequently use scream-type calls and higher-pitched grunt-like calls, while adults use more low-frequency grunts. Vocal activity tends to decrease with age (Hammerschmidt et al. 1994). There is no unidirectional relationship between the type of vocalization and the context in which it is uttered, although call usage is more specific within some contexts. For example, Barbary macaques tend to produce noisy vocalizations during agonistic contests (Fisher and Hammerschmidt 2002). Barbary macaques tend to produce bouts (series) of several consecutive calls. Within each bout different call types can be found. Bout duration depends on the context of usage and can last for hours. Adults produce sequences with rather similar calls, while infants and juveniles tend to produce extremely variable vocal patterns within one sequence. Despite a clear sex dimorphism in body size (males are, on average, one third larger than females), no significant sex differences in the acoustic structure of calls in adults has been found to date (Hammerschmidt et al.1994); but, as in humans and other primates, there are age- and sex-related differences in their vocalizations (as animals grow, pitch and modulation decrease). Moreover, a peculiar "commenting behavior" has been described only in this species: macaques utter low-amplitude calls when observing close contact interactions of other members of the group (Fisher and Hammerschmidt 2002; Brumm et al. 2005).

The aims of this study were to:

- Quantify the developmental changes of Barbary macaques' vocalizations and check for sex-related differences.
- Investigate whether some modulation of F0/formant pattern occurs according to the behavioral context of sound production

2.2 MATERIALS AND METHODS

Subjects: The study colony of Barbary macaques was living in an outdoor enclosure in the Natura Viva Park in Bussolengo (Verona). At the time of the study, the colony contained 24 individuals (14 males and 10 females), ranging in age from 6 months to 16 years. Data on age, sex and kinship were provided by the park staff (Table 1).

Father → Mother ↓	***MAROCCO***	***IAGO***		***LUCKY 10/5/99***	***MAROCCO***
RAFFA +		RAY 9/6/03	REMI' 2/6/04		
MOIRA +		MILLY 11/6/03	MORFEO 10/6/04		
BROOKE 1/1/90	BABE 16/5/2	BARTOLOMEO 26/7/03	BENEDICT 18/5/04	BRUNO 14/4/05	Male1 11/5/06
FLO +		FRIDA 28/7/03			
LADY 31/5/00			LUCREZIA 22/5/04	LAILA 1/6/05	Male2 15/5/06
FRANCY 21/5/00			TEODOR 19/5/04	FUNNY 18/5/05	
BELINDA 8/6/01				BENNY 3/7/05	Male3 2006
BRENDA 31/5/00				BRYAN 6/6/05	

Table 1 Name, genealogy and available dates of birth for study animals. [+ indicates the animal died before the study started]

Experimental apparatus: A Dan Gibson EPM-650 Electronic Parabolic Microphone and Marantz PMD 671 solid state recorder were used to record vocalizations. This allowed high-quality recordings to be made at a variety of distances (from 1 – 60 meters) from the animals.

Procedure: The study was carried out in 3 phases:

i. One month for learning to individually identify the subjects, and for acclimating the colony to the presence of the researcher.

ii. Recording of sounds *ad libitum* using *focal sampling* and *all occurrence sampling*. For each sound the following data were scored: emitter's identity, receiver (if present) and the context of sound production.

iii. From 16 October 2006 to 16 November 2006 behavioural observation was carried out using the scan-sampling technique, in which an observer censuses all of the animals in the colony and records their behavior at the instant they are observed. Because scan sampling must be done relatively quickly, it was restricted to broad categories of behavior reported in Table 2. In each 2-hour session, 24-scans for each member of the colony (1 every 5 minutes) were recorded. Sessions were run either during the morning (from 9 to 12), during the afternoon (from 14 to 17) or both. This observation allowed the experimenter to understand the hierarchies of the colony. A total of 25 2-hour sessions (*24 scans) = 600 scans of the colony was carried out during the one month observation period. The data set consists of a total of 600 scans x 24 macaques = 14400 behavioral observations. In each session the experimenter choose one of the two possible positions for a broad view of the colony. Position was counterbalanced across sessions.

The behavioral categories used for analysis are described in Table 2, in accordance with previous studies on the same colony (Dell'Aira 1995). Behavioral contests (i.e. Social and Agonistic) will be taken into account along with the cases when an animal was ALONE, to investigate how sound production was related to behavior.

Context	*Behavioural Category*	*Behavioural unit*	*Description*
Social Context	GR/GRr	GR/GRr	Grooming / receive grooming
	IC infante care	AL	Breast feeding
		ALZ	To raise an infant, hugging it, sometimes during walking
		ISP	To inspect an infant
		RA	To ask for breast feeding
		RU	To recover an infant
		S	To watch over an infant
		TR	To carry an infant on the back
		TV	To carry an infant ventrally
		BEU	Being kept at the hips by an adult/subadult; sometimes the adult/subadult raises the back of the subject.
	SI r (receive social interaction)	Af Th.C r	To receive teeth chattering for affiliative reasons
		E-ABB	Being hugged
		E-AN	Being sniffed
		E-HT	Being touched
		GI r	Genital inspection
	SI (social interaction)	AA	Two animals facing each other grip one arm reciprocally while doing teeth chattering
		AB	To hug reciprocally (ventrally or not)
		ABB	To hug another animal
		AF	To hold someone else with a hand
		Af TH.C	Teeth chattering
		AP	An animal intentionally approaches another animal. This can happen with a mother-infant couple, but also with adult/subadult couples. The approaching animal sits next to the couple, raises the infant and does teeth chattering on its back.
		BP	Baby presentation: to present an infant to another animal

		FO	Follow
		FS	Feed-social: to feed all together
		GI	Genital inspection
		GT	The animal presents its back to another animal, bending itself and putting its head between two stretched legs, often doing teeth chattering
		HT	To touch another animal
		HU	Huddling in group
		IS	invite someone (usually an infant) to go up on the back
		RR GR	An animal asks someone else to be groomed showing its back
		SC	Sitting in contact without doing anything else
	SP (social play)	BI pl	To bite for play
		CH pl	To pursue for play
		PU pl	To push for play
		SCR pl	To scramble
		SF	To hold an infant’s hips doing teeth chattering
Agonistic Context	AA (dominant without contact)	AA	Alarm: to look excited when there is danger, to vocalize in order to call someone for help.
	AgC (aggressive with contact)	AT	Attack
		PU	Push:
		SPOSTA	Move an animal out of your own way by pushing it with the hand
	AgNC (Aggressive without contact)	STR	To jerk
		CH	To chase
	DoC (dominant with contact)	BI-BITE	To bite
		FN/FNr	pretended mount, receive a pretended mount
		RE	to get an animal out of the way pushing or pulling it

	DoNC (dominant without contact)	BS	branch shaking:; hitting the ground with the legs
		FISSA	To stare at another animal with a gape, or threaten (this behavior was never observed in the dominant male)
		M	Threatening facial expression (usually to get an animal out of the way): lips protruded forming an “O”, eyebrows raised
		REC	To ask for support during a fight
		RINC	To pursue another animal
		SB	To yawn in order to show teeth
		SO	To check what another animal (usually an infant) is doing.
		TH	To threat
	RD (dominant with contact)	RD	Redirection: A attacks B, then B attack another lower ranking C
	SoC (sumbissive with contact)	CL s	Being held at the hips
		E-RE	Being pushed by another animal
		E-SP	Being moved from your own position
		E-STR	Being jerked
		E-TR	Being kept by another animal
	SoNC (sub missive without contact)	Ag TH.C/Ag TH.C r	Teeth chattering/receive teeth chattering
		AV	Avoid: to avoid someone by walking around them (hierarchical index)
		DI	Displacement. To move in order to avoid someone who is approaching. Hierarchical Index.). When a certain animal approaches, the subject reacts going away or looking somewhere else
		E-M	Being threatened
		E-RIN	Being pursued
		F	The animal flees from an attacker
		PR	Presentation: to present the rump to a dominant animal.

	SU (support)	SU	Support an animal A who is fighting with another animal B, going first to A and then going together against B
ALTRO	Altro	FR	To steal an infant to his mother
		EAT	The animal eats
		LOC	The animal wanders
	ALONE	ALONE	The animal is alone (doing something or not), it does not interact with anybody.
	OS	OS	The animal is out of sight
	Sex	MO	Mount with ejaculation

Table 2 List of behaviors observed in the colony. The first column describes the different behavioral contexts (dark grey: Social; light grey: Aggressive; white: Other).The second column contains the list of behavioral categories scored during scan-sampling. The third column contains the abbreviations for different behavioral units belonging to each category; the fourth column contains a brief description of each behavior.

Sound analysis: Sounds (N = 4625) were analyzed with SIGNAL software (Engineering Design, Berkeley, CA, USA). Sound spectrograms utilized a Hanning window, with an FFT size of 256 points. Classification was based on visual inspection, and the types of sounds within each classification category are illustrated in Fig 4. Acoustic analyses examined the distribution of spectral energy (formant profiles) when energy was simultaneously present at more than one frequency. F0 and Formant dispersion were calculated with custom-written SIGNAL scripts.

F0 was calculated according to the structure of the vocalization, (as determined by visual inspection of the waveform, the spectrum and the spectrogram of each recorded sound) as follows: F0 was calculated with the autocorrelation method if the sound was clearly harmonic, calculating the peak frequency (frequency at which there was the most energy) if the sound was tonal, calculating the maximum value of the power spectrum of the difference between the original time signal and a highly smoothed version of it (Smoothing width 10Hz) in case of noisy sounds.

Formants were calculated from the peaks of the average spectrum of the entire time signal. Formant dispersion was calculated as the average distance between these peaks, according to Fitch (1997).

The relation between these measures and features of the opponent (such as status and size) were explicitly examined.

Presentation of the results of this study will be divided into two parts: a cross-sectional study of the vocal repertoire, and an examination of the relationship between sound structure and behavior.

The cross sectional study of vocal repertoire was undertaken to provide a fuller description of the spectral properties of the vocalizations initially described by Hammerschmidt and Fischer (1998) and to investigate the developmental and sex-related changes of the vocalizations. The context of sound production and the identity of the caller were also recorded for all vocalizations. Since it was not possible to obtain precise body size data, this quantity was inferred from the age and the sex of the emitter. According to the year of birth, a value representing the “Age Class” was assigned, starting from 1 for animals born in 2006, to 2 for animals born in 2005, and so on, until value 6, indicating adults (i.e. females born until 2001 and the dominant male, born in 1999). For this reason, an examination of the relationship between acoustic features and size was possible only between age and sex classes.

To investigate the behavioral context of sound production and context-related modulation of size-related features more in detail, this study focused on modulation of formant dispersion and F0. This modulation was examined on a single-subject level in relation to the rank of the emitter relative to the rank of the recipient/opponent, and the relative body size of the emitter. For this reason, it was only possible to analyze sounds produced by single animals when at least 5 sounds per condition were uttered. To be as conservative as possible, non-parametric analyses for independent samples were run (Kruskall-Wallis or Mann-Whitney U-tests). While this approach drastically reduces the number of sounds that can be utilized, it avoids attributing effects that may be due to body size or sex differences with those due to contextual modulation of vocalizations within individual animals.

2.3 RESULTS

Part 1 - Cross sectional study of the vocal repertoire

According to Hammerschmidt et al. (1994), Hammerschmidt and Fisher (1998), and Fischer and Hammerschmidt (2002), we identified both tonal and noisy screams emitted by all animals in different behavioral contexts. In these papers authors identified 7 different clusters of sounds according to a number of acoustic parameters; however, they did not provide a precise quantitative description of them that allowed the clustering of the sounds recorded in this study. Therefore we visually identified 10 different sound structures (Fig 4) for subsequent analysis.

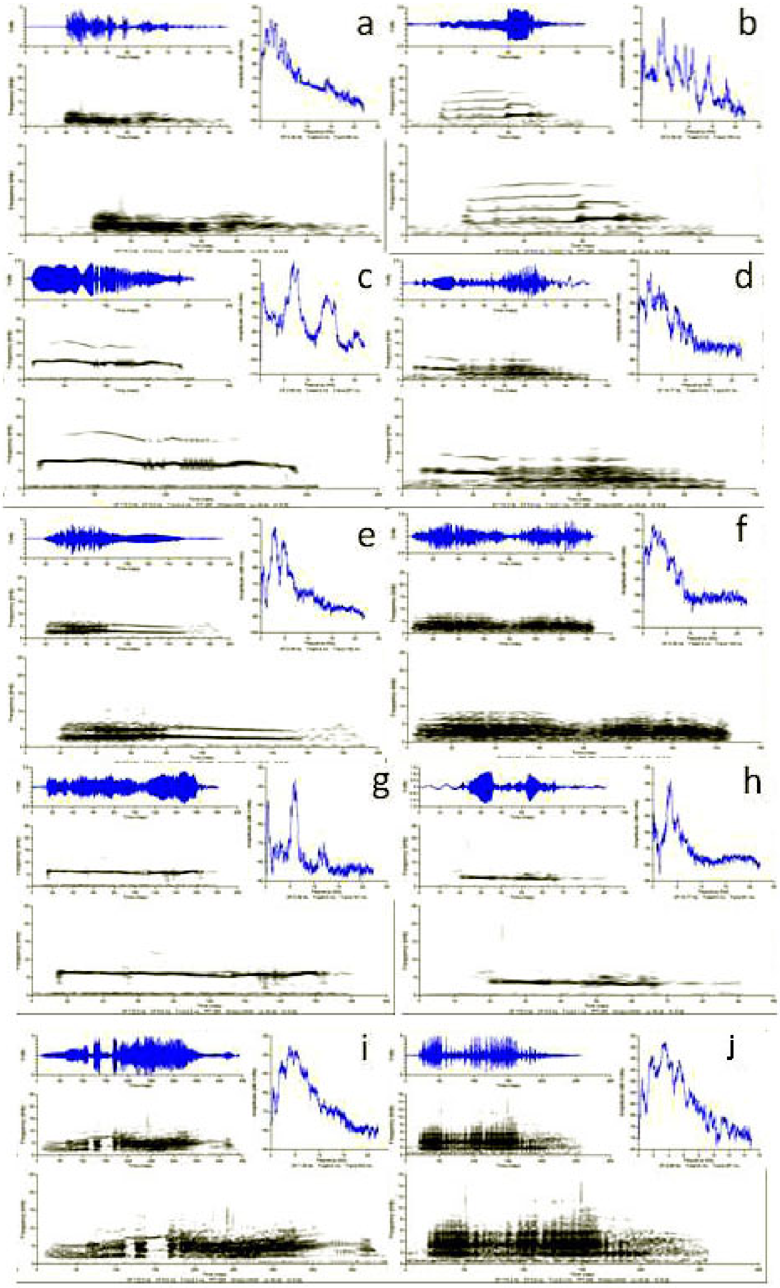

Fig 4 Classification of sounds after visual inspection. Waveform, spectrogram and power spectrum of the sound structures (see text).

The following structures were identified (the proportion of total vocalizations recorded that each category represents is given in parentheses):

a. **Short noisy sound SN**: sound shorter than 100ms, without harmonic/tonal structure (44,17% of total)

b. **Short harmonic sound SH**: sound shorter than 100ms, with harmonic structure (5,47%)

c. **Long harmonic sound LH**: sound longer than 100ms, with clear harmonic structure (4,71%)

d. **Half harmonic half noisy HHHN**: First harmonic half, noisy second half (3,81%)

e. **Half noisy half harmonic sound HNHH:** First noisy half, second harmonic half (0.91%)

f. **Long noisy sound LN**: sound longer than 100ms, without harmonic/tonal structure. It sounds very harsh (10.12%)

g. **Long tonal sound LT**: Tonal sound longer than 100 ms (0.48%)

h. **Short tonal sound ST**: Tonal sound shorter than 100 ms (1,08%)

i. **Tonal-Formant sound TF:** Sound alternating tonal parts with parts with visible formants. Tonal or harmonic parts can occur at the beginning, in the middle, or at the end of the sound (14,08%)

j. **Formant sound F:** Sound with formant structure(13,08%)

Additional rare utterances were placed into two more categories:

Composed sounds COMPOSED: sounds with several elements of the same structure (AC, F, NC, or TF, defined above) put into a sequence (not shown) (0.12%)

Complex sounds COMPLEX: diverse sounds with several elements of different basic structures put into a sequence (1,97%)

A. Relation between Sound and body size

A first analysis was run with Spearman rank correlations between F0 (fundamental frequency) or Fd (formant dispersion) and age class, considering the average value for each animal from *all* the sounds it emitted. In a second step the same analysis was run on each sex (Fig 5 and Fig 6).

F0 $\rho = -0.43$, $p = 0.03$** (N = 24)

Males $\rho = -0.77$, $p = 0.005$** (N = 14)

Females $\rho = -0.40$, $p = 0.23$ (N = 10)

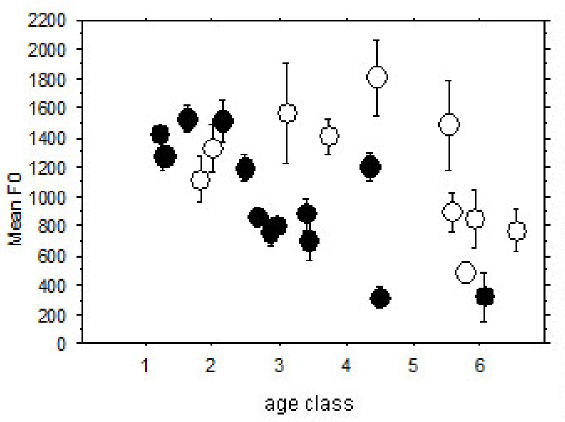

Fig 5 MeanF0 ± SEM for each animal of the colony in relation to age Class and Sex. Black circles represent males, white circles represent females.

Fd ρ = -0.45, p = 0.03** (N = 24)

Males ρ = -0.67, p = 0.01** (N = 14)

Females: ρ = -0.46, p = 0.16 (N = 10)

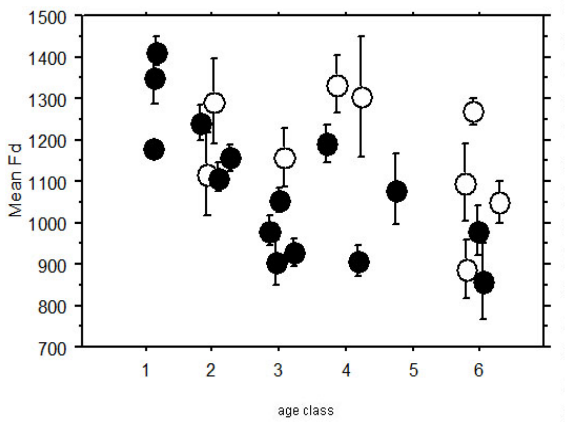

Fig 6 MeanFd ± SEM for each animal of the colony in relation to age Class and Sex. Black circles represent males, white circles represent females.

Analogous analyses were run for each type of sound structure (a total of 50 correlations). However, when corrected for these 50 comparisons, a correlation had to be significant at the p = 0.001 level in order to be considered significant, which none of these comparisons attained.

		F0		***Fd***	
Sound Structure	Subjects	ρ	p (nominal)	ρ	p (nominal)
Short Noisy Sound	All (N = 24)	-0.27	0.19	-0.40	0.05
	Males (N = 14)	-0.51	0.06	-0.49	0.07
	Females (N = 10)	0.25	0.46	-0.18	0.58
Short Harmonic Sound	All (N = 17)	0.003	0.99	-0.34	0.16
	Males (N = 12)	-0.47	0.12	-0.74	0.01
	Females (N = 5)	0.56	0.26	0.35	0.47
Long harmonic sound	All (N = 20)	-0.16	0.49	-0.39	0.09
	Males (N = 10)	-0.32	0.33	-0.62	0.08
	Females (N = 10)	-0.63	0.06	-0.42	0.20
Half harmonic half noisy sounds	All (N = 13)	0.05	0.87	-0.17	0.68
	Males (N = 8)	-0.14	0.71	-0.59	0.12
	Females (N = 5)	0.00	1.00	0.79	0.11
Half noisy half harmonic sound	All (N = 9)	-0.28	0.43	0.05	0.88
	Males (N = 8)	-0.57	0.13	-0.38	0.32
Long Noisy sound	All (N = 22)	-0.44	0.04	-0.60	0.006
	Males (N = 13)	-0.65	0.02	-0.67	0.02
	Females (N = 9)	-0.10	0.76	-0.42	0.23
Short tonal sound	All (N = 11)	-0.02	0.94		
	Males (N = 8)	0.08	0.84		
Tonal-formant sound	All (N = 23)	-0.48	0.02	-0.11	0.59
	Males (N = 13)	-0.74	0.01	-0.43	0.13

	Females (N = 10)	-0.51	0.12	-0.39	0.24
Formant sounds	All (N = 22)	-0.54	0.01	-0.37	0.08
	Males (N = 12)	-0.68	0.02	-0.75	0.01
	Females (N = 10)	-0.28	0.40	0.13	0.69

Table 3 Results of Spearman Rank Correlations between F0/Fd and body size for each sound structure

Examining just the nominal significance level of the general statistics for all vocalizations, an inverse correlation between both F0 and formant dispersion was found with body size (inferred from age and sex); these correlations held for males alone, but not for females alone. Looking more in detail at individual vocalization types, however, this relation did not reach true significance for any individual comparison, even if it attained nominal significance for some of the male comparisons.

Part 2- Focus on the relation between sound and behavior

B. Sound production in different contexts

Fig 7 shows the overall number of sounds recorded in the three main social contexts, while Fig 8 shows the distributions of the occurrences of each sound category recorded in social, agonistic, aggressive (during specific submissive or aggressive behaviors) contexts, or when alone, respectively, according to Table 2 in the Methods section (page 30).

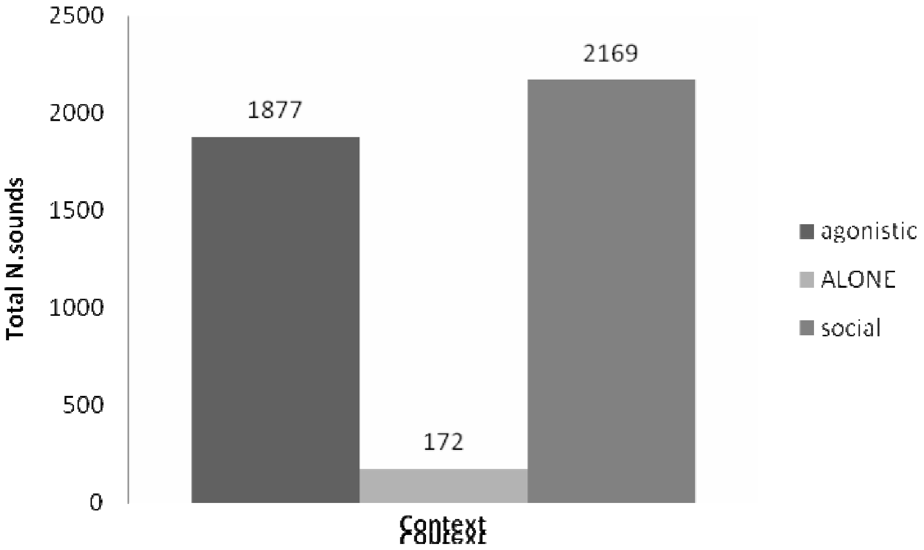

Fig 7 Total number of sounds emitted by animals in the three different main contexts (Social, Agonistic, Alone)

A G-test (Sokal and Rohlf 1995) to test the null hypothesis of an equal distribution of sounds in the 3 main contexts revealed a significant deviation from an equal distribution ($G_{(2)}$=1332.50, $p < 0.001$).

In the “alone” context, 7 sound categories were not produced: Complex, Formant, Long Tonal, Composed Short Harmonic, Composed Formant, Composed Short noisy, Composed Tonal-Formant.

In order to compare the distribution among the individual sound categories in the 3 main contexts, a G-test was run with the 8 sound categories present in all three main contexts (short harmonic, long harmonic, half harmonic half noisy, half noisy half harmonic, short noisy, long noisy, short tonal, tonal-formant). The distribution of call types differed among the three contexts ($G_{(14)}$=179.80, $p < 0.001$), indicating that sounds are unevenly used in these different contexts.

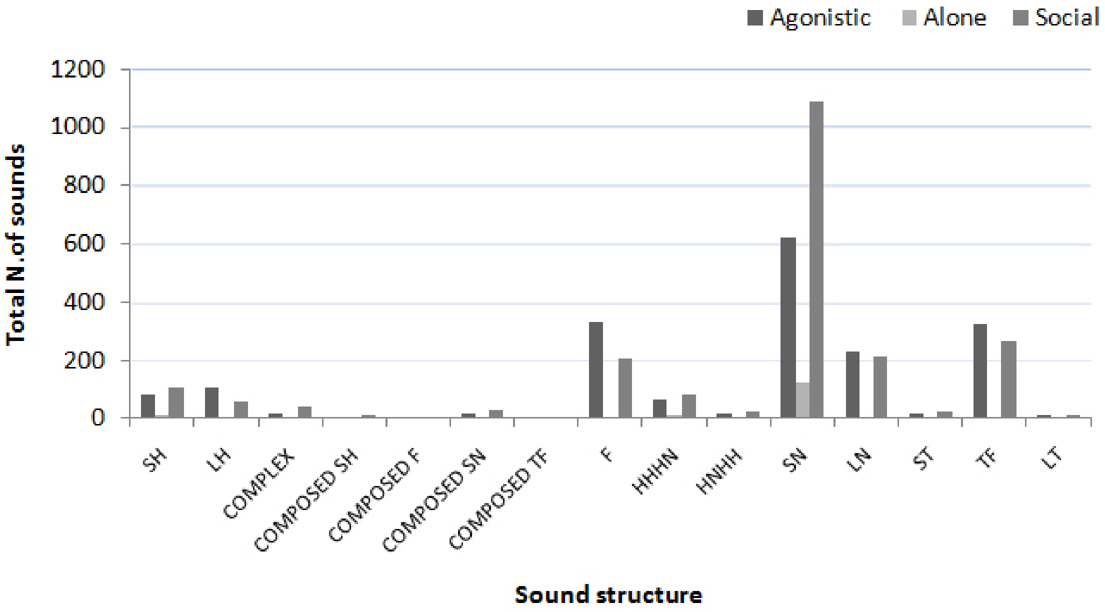

Fig 8 Distribution of different types of sounds in the three main behavioural contexts

Vocalizations uttered during agonistic interactions were more frequent in submissive contexts (Fig 9).

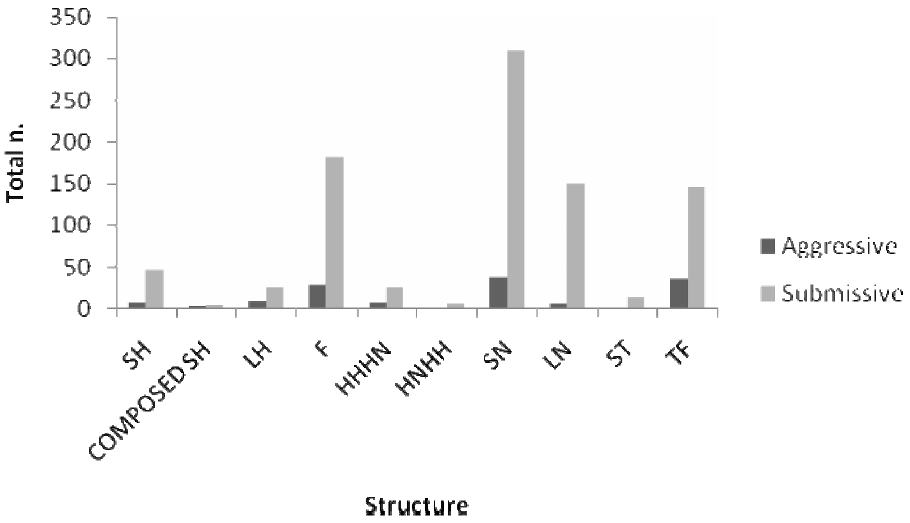

Fig 9 Sounds emitted in aggressive contexts (during submissive or aggressive behaviors)

In order to check whether context-dependent active articulation occurs, a non-parametric analysis was run on each single sound structure that had at least 5 vocalizations uttered by the same animal. Modulation of either F0 or Fd was examined in relation to:

- Relative Rank of the emitter respect to the recipient/opponent in agonistic and social context
- Relative Size of the emitter respect to the recipient/opponent in agonistic and social context

According to the Motivational-Structural Rules (MS) hypothesis (Morton 1977), the pitch of vocalizations should be modulated according to the motivational state of the emitter, with low, harsh sounds produced in aggressive contests, and high and tonal sounds uttered in fearful, social or submissive contests. This modulation should be more pronounced in situations involving a higher level of threat to the emitter of the sounds. In order to check for this possibility, we compared situations in which an animal was attacked with contact (high severity) or without contact (less severe attack). Active control of the vocal tract with protrusion or retraction of the lips and with raising or lowering the larynx could lengthen of shorten vocal tract length in a way that could change the perceived body size in the receiver. In this context, signalers should try to maximize the impression of their body size (i.e. have a lower Fd) in situations such as aggression, and they should try to simulate a smaller size during appeasement (i.e. increase Fd) (Fitch and Hauser 1995).

C. Modulation according to the Relative Rank of the emitter relative to the recipient/opponent

Emitter	***Type of sound***	$N_{higher\ rank}$	$N_{lower\ rank}$	***Mean*** $F0_{higer\ rank}-F0_{lower\ rank}$ ***(Hz)***	***p***	***Behaviour co-occurring with sound***
Milly	Formant	5	16	-190.89	0.02*	Aggressive (Submissive behaviors)
Milly	Tonal-Formant	9	25	-598,59	0.15	Aggressive (Submissive behaviors)
Brenda	Formant	7	7	-84,4	0.95	Aggressive (aggressive behaviors)
Brenda	Long noisy	26	9	146,83	0.04*	Aggressive (aggressive behaviors)
Laila	Short noisy	12	21	-185,73	0.95	Social

Table 4 Results of non–parametric Mann-Whitney U tests looking for modulation of F0 when a sound structure is uttered towards a higher or a lower-ranked animal during agonistic and or social contests. Significant p values are highlighted by asterisks.

With one exception (e.g Brenda – long noisy sound), F0 of a vocalization was higher when the emitter’s rank was lower than the opponent’s rank. This could be related to a higher level of fear that the emitter had when confronting a higher ranking animal, in accord with the MS rules hypothesis.

Emitter	***Type of sound***	***$N_{higher\ rank}$***	***$N_{lower\ rank}$***	***Mean $Fd_{higer\ rank}$-$Fd_{lower\ rank}$ (Hz)***	***P***	***Behaviour co-occurring with sound***
Milly	Formant	5	15	220.31	0.09	Aggressive (Submissive behaviors)
Milly	Tonal-Formant	9	25	-52.11	0.54	Aggressive (Submissive behaviors)
Brenda	Formant	7	7	-228,88	0.3	Aggressive (aggressive behaviors)
Brenda	Long noisy	25	8	-350.04	0.14	Aggressive (aggressive behaviors)
Laila	Short noisy	12	21	2.47	0.3	Social

Table 5 Results of non–parametric Mann-Whitney U tests looking for modulation of Fd when a sound structure is uttered towards a higher or a lower-ranked animal during agonistic and or social contests. Significant p values are highlighted by asterisks.

Analyses of Fd according to the rank of the emitter did not identify any clear pattern.

A. Modulation according to the Relative Size of the emitter relative to the recipient/opponent

Emitter	*Type of sound*	*$N_{bigger\ size}$*	*$N_{same\ size}$*	*$N_{smaller\ size}$*	*Mean $F0_{bigger}$-$F0_{same}$-$F0_{smaller}$ (Hz)*	*P*	*Behaviour co-occurring with sound*
Bryan	Short noisy	0	6	57	461.73	0.13	Aggressive (submissive behaviors)
Bruno	Long noisy	0	5	17	-421.24	0.84	Aggressive (aggressive behaviors)
Morfeo	Short noisy	0	14	23	-271.44	0.13	Aggressive (submissive behaviors)
Teodor	Short noisy	0	9	21	-675.22	0.13	Aggressive (submissive behaviors)
Teodor	Tonal-Formant	0	6	7	598.12	0.005**	Aggressive (submissive behaviors)
Milly	Formant	7	12	0	-136.02	0.12	Aggressive (submissive behaviors)
Frida	Formant	0	5	6	502.66	0.45	Aggressive (submissive behaviors)
Milly	Short noisy	24	35	10	-993.02	0.62	Aggressive (submissive behaviors)
Frida	Short noisy	0	6	11	1599.7	0.002**	Aggressive (submissive behaviors)
Frida	Tonal-Formant	0	8	7	1707.95	0.64	Aggressive (submissive behaviors)
Milly	**Tonal-Formant**	**11**	**18**	**5**	**-1573.46**	**0.01***	**Aggressive (submissive behaviors)**
Milly	Short harmonic	10	10	0	1731.27	0.01*	Aggressive (submissive behaviors)
Milly	Long noisy	0	18	5	365.08	0.6	Aggressive (submissive

							behaviors)
Bryan	Short noisy	0	17	18	362.33	0.11	Social
Benny	Short noisy	0	9	21	-402.64	0.11	Social
Morfeo	Short noisy	0	8	50	125.51	0.68	Social
Teodor	Short noisy	0	5	18	-103.35	0.79	Social

Table 6 Results of non–parametric analyses looking for modulation of F0 when emitter's size is either bigger, same or smaller than the size of the recipient/opponent. These were either Mann Whitney U tests or Kruskal-Wallis tests (in Bold case). In the latter cases, post-hoc Mann Whitney U tests with Bonferroni corrections were run and the results are reported in appendix NN. Significant p-values are highlighted by asterisks.

Emitter	***Type of sound***	$N_{bigger\ size}$	$N_{same\ size}$	$N_{smaller\ size}$	***Mean*** Fd_{bigger}-Fd_{same}-$Fd_{smaller}$ ***(Hz)***	***P***	***Behaviour co-occurring with sound***
Bryan	Short noisy	0	6	50	-223.91	0.21	Aggressive (submissive behaviors)
Bruno	Long noisy	0	5	16	457.81	0.01*	Aggressive (aggressive behaviors)
Morfeo	Short noisy	0	13	23	342.83	0.03*	Aggressive (submissive behaviors)
Teodor	Short noisy	0	9	21	79.3	0.72	Aggressive (submissive behaviors)
Teodor	Tonal-Formant	0	6	7	-466.61	0.21	Aggressive (submissive behaviors)
Milly	**Formant**	**7**	**12**	**50**	**209.75**	**0.51**	**Aggressive (submissive behaviors)**
Frida	Formant	0	5	5	130.7	0.25	Aggressive (submissive behaviors)
Milly	Short noisy	24	35	10	-958.7	0.34	Aggressive (submissive behaviors)
Frida	Short noisy	0	6	11	139.14	0.43	Aggressive (submissive behaviors)
Frida	Tonal-Formant	0	8	7	1491.43	0.02*	Aggressive (submissive behaviors)
Milly	Tonal-Formant	11	18	5	-1995.28	0.03*	Aggressive (submissive behaviors)

Milly	Short harmonic	8	10	0	1560.16	0.07	Aggressive (submissive behaviors)
Milly	Long noisy	0	16	5	-162.68	0.17	Aggressive (submissive behaviors)
Bryan	Short noisy	0	17	18	296.13	0.008**	Social
Emitter	***Type of sound***	$N_{bigger\ size}$	$N_{same\ size}$	$N_{smaller\ size}$	***Mean*** $Fd_{bigger}-Fd_{same}-Fd_{smaller}$ ***(Hz)***	***P***	***Behaviour co-occurring with sound***
Benny	Short noisy	0	9	20	29.88	0.37	Social
Morfeo	Short noisy	0	8	50	93.56	0.4	Social
Teodor	Short noisy	0	5	18	303.72	0.02*	Social

Table 7 Results of non–parametric analyses looking for modulation of Fd when emitter's size is either bigger, same or smaller than the size of the recipient/opponent. These were either Mann Whitney U tests or Kruskal Wallis tests (in Bold case). Significant p-values are highlighted by asterisks.

Even if it seems that some kind of modulation of either F0 or Fd occurs, it is not possible to identify any clear pattern as predicted by the MS rules hypothesis.

F0 changes and level of agonistic threat

If sounds are modulated by the level of agonistic threat as proposed by Morton (1977), sounds emitted by an animal should have a higher pitch when the attack is more severe, hence F0 should be higher when the sounds co-occur with a behavior involving physical contact. To test this hypothesis, U-tests were run on sounds emitted during agonistic behaviors, dividing these into 2 groups (i.e. aggressive or submissive context), and defining each sound as co-occurring with contact or non-contact.

Emitter	***type of sound***	***$N_{Contact}$***	***$N_{NoContact}$***	***Mean $F0_{contact}$- $F0_{NoContact}$ (Hz)***	***P***	***context***
Bruno	Formant	16	5	795	0.02*	Submissive
Bruno	Tonal-Formant	31	5	912.47	0.25	Submissive
Bryan	Short noisy	44	19	558.29	0.07	Submissive
Morfeo	Short noisy	20	25	132.32	0.18	Submissive
Morfeo	Tonal-Formant	9	8	110.29	0.7	Submissive
Milly	Short noisy	16	56	137.13	0.94	Submissive
Milly	Long noisy	8	17	360.41	0.95	Submissive
Milly	Tonal-Formant	11	23	603.19	0.73	Submissive
Brenda	Tonal-Formant	5	8	-23.62	0.31	Aggressive

Table 8 Results of Mann Whitney U tests looking for modulation of F0 when, during agonistic interaction, an emitter utters a certain type of sound when physical contact occurs or not. Significant p-values are highlighted by asterisks.

In this case a clear pattern is shown: even in non-significant results, the trend of the data indicates that F0 is higher when, in a submissive context, contact occurs: this could be due to a higher level of arousal of the emitter. Interestingly, the only case in which the trend was opposite was associated with an aggressive contest: in this case the animal (Brenda) screamed with a higher F0 when no contact occurred (i.e. when she was pursuing another animal) and therefore it can be assumed that in this case the arousal level was higher than in the case when the contact occurred, probably after the climax of the encounter was reached.

2.4 DISCUSSION

Relation between sound features and body size. A weak but significant inverse correlation between formant dispersion and body size (inferred from age and sex) was found when considering the entire vocal repertoire, and this appeared to be due to correlations in males but not in females; yet there was not a strong pattern when considering individual sound structures. The lack of correlation between body size and acoustic features in females has already been documented in humans and Hamadryas baboons (Rendall et al. 2005) and Japanese macaques (*M. fuscata*) but it is in contrast with the findings in other species such as in rhesus macaques, where a negative correlation between formant dispersion and body size exists for both sexes (Fitch 1997). This pattern could be explained if we take into account the particular social organization of Barbary macaques. As many other species of the genus, Barbary macaques live in multi-male/ multi-female groups with a strong matrilineal kinship. Males habitually leave the group, and could join other groups that do not belong to the matrilineage (Mènard and Vallet 1993; Prud'Homme 1991). In contrast with other species such as Rhesus macaques, their social structure can be defined as "egalitarian" (Matsumura 1999), with quite high levels of reconciliation after agonistic interactions. In this framework, it would be useful only for males to provide acoustic cues about their strength or fitness (i.e. body size), in order to ritualize fighting and reduce the likelihood of severe injuries during intrasexual interactions. For females, where matrilineal kinship defines the social status within the group, the ability to provide information about body size with vocalization may not be as important.

Focus on the relation between sound and behavior. As Fisher and Hammerschimdt (2002) described, Barbary macaques produce a wide variety of sounds in different behavioral contexts, especially when an interaction, either social or agonistic, is involved: the sound production was very low when the animal was alone, while during social or agonistic interactions vocal activity strongly increased. The different sound structures were unevenly distributed in the contexts and, within agonistic interactions, the sound production was higher when the emitter was in a submissive condition. This could suggest that the function of vocalizations was to recruit support and maybe broadcast useful information to the recipients. We also found that, within a single sound structure, acoustic properties of the sound (F0 or Fd) were modulated by the same animal according to the behavioral context. However, the pattern of results did not show a clear trend and we cannot infer the function of this modulation.

In submissive contexts, but not only under aggressive circumstances, mean F0 was higher when the attack was more severe and involved physical contact. The trend was the same for all of the cases except one, even when the results were not significant. The only case in which the trend was opposite was associated with an aggressive contest: in this case the animal (Brenda) screamed with a higher F0 when no contact occurred (i.e. when she was pursuing another animal).In this case, it is likely that the arousal level was higher than in the case when the climax of the encounter was reached and contact occurred We can conclude that F0 was modulated according to the arousal level of the emitter, according with the MS rules hypothesis (Morton 1977).

3 PERCEPTION OF SIZE-RELATED FEATURES OF NATURAL SOUNDS IN NORMAL-HEARING CHILDREN

3.1 INTRODUCTION

Many natural sounds contain information about the size of the object that produced them. In the case of both animate and inanimate objects, bigger objects generally produce more low-pitched sounds. For example, as a mammal grows, the lengthening of its vocal tract produces a decrease in resonant frequencies (formants), and, at the same time, age and sex-related growth of its vocal folds result in lower fundamental frequencies (Fitch and Giedd 1999; Riede and Fitch 1999; Smith et al. 2005). Adult humans are able to make fine judgments about the size of a speaker or of an object relying only on acoustic cues (Houben et al. 2001; Smith et al. 2005).

Patterson and colleagues have proposed the idea that the ability of the brain to extract size information from sounds is an inborn trait, and that the information extraction occurs at an early stage of the auditory processing, even with stimuli whose apparent size goes well beyond the range of everyday experience (Smith et al. 2005). Subjects' performance in size assignment depends upon the amount of acoustic information they receive – for instance, they do better with syllables instead of isolated vowels (Ives et al. 2005).

Very few studies have addressed the development of size attribution to sounds by children, and these studies give contradictory information. Does a child need much training to extract information about the size of a source only from acoustic cues, or is this something children do naturally from an early age? Mondloch and Maurer (2004) have found that 3-year-olds (30-36 months) reliably match high-pitched sounds with smaller objects (and low-pitched sounds with bigger objects); previous work by Marks and colleagues (1987) did not find this ability in 9 year-olds.

The aim of the present chapter is to understand when children begin to attribute size to sound-producing objects in a manner that is similar to adults. Moreover, if experience matters in developing this skill, then different categories of sounds may be associated with size judgments that have different levels of accuracy. For example, high accuracy may be shown for human-produced sounds, which children have a lot of experience with, while lower levels may be shown to stimuli produced by dogs, and little accuracy may be shown to sounds produced by unfamiliar objects. On the other hand, if the segregation of size information is an automatic process that is the same for all objects, there should not be differences among the accuracy of size judgments to different categories of stimuli. If it is an automatic process that is done differently for different objects, then it may exhibit developmental differences for different objects, but these would not necessarily correlate with the amount of experience with the objects.

As a first step, we wanted to test the accuracy of size judgments of natural sounds belonging to different categories but sharing the same basic physical structure (i.e. pulse-resonant sound). Assuming that the adult control group has the highest accuracy in size judgments for all the categories, we wanted to test at what age the children’s performance becomes similar to adult performance.

3.2 MATERIALS AND METHODS

3.2.1 ACOUSTIC STIMULI

Speech and dog sounds were recorded with a Marantz PMD671 digital recorder and high-quality microphones, stored as WAV files (16-bit; sampling rate 44.1 KHz). Sounds of interest were isolated for use as stimuli. Musical instrument stimuli were built from the Real World Computing database of musical instruments. From the strings database, 4 notes were chosen for each instrument and 4-notes melodies were built with SIGNAL software (Engineering Design, Berkeley, CA, USA). The root-mean-square of each sound was normalized across all stimuli.

· 6 SPEECH stimuli were selected from the CHILDES[1] (ANTELMI) database in order to have words with low ages of acquisition.

A question (Ti piace il riso? – Do you like rice?) and a sentence (Non si può aprire- You mustn’t open it) were recorded by a man, a woman and a child (all native speakers). The physical height of the 3 speakers was measured in order to have an estimation of the vocal tract length according to Fitch and Giedd (1999)

· 6 dog barks were selected from a small sample of 6 dogs of 3 different sizes (small/medium/large).

Measures of dog body length were also collected (from the tip of the nose until the beginning of the tail, and the skull length) in order to estimate the VTL following Riede and Fitch (1999).

· 2 4-note melodies played by a violin, viola, and cello were built using the Real World Computing database of musical instruments (Goto et al. 2003).

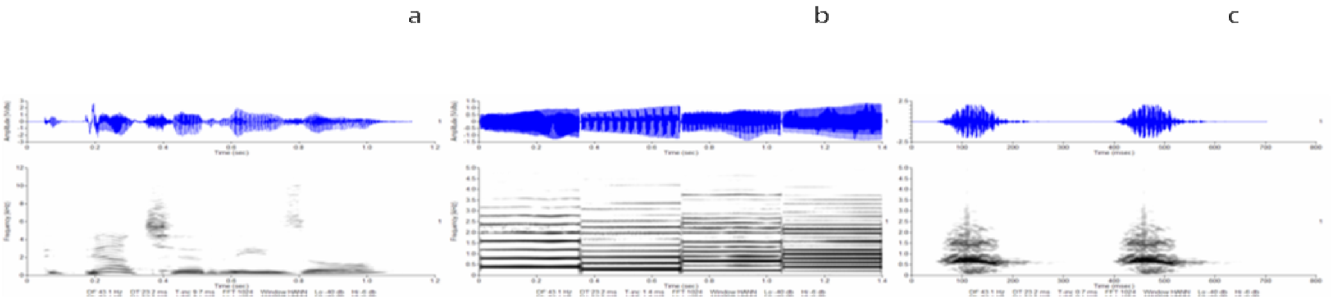

Fig 10 Representative spectrograms and waveforms of the three categories of stimuli: a) speech-question pronounced by a woman; b) short melody played by violin; c) two consecutive barks. Hanning window, FFT size 1024

---

[1] CHILDES (Child Language Data Exchange System) is a large database of samples of child communication in different languages. These sentences have been built with CLAN software using the more frequent Italian words found in the ANTELMI database (a corpus consisting of 7 recordings of one child from 2,.2 to 3,.4 years of age).

3.2.2 VISUAL STIMULI

Visual stimuli were images belonging to 3 categories:

· HUMANS representing the three categories described by Peterson and Barney (1952): man, woman, child.
· DOGS were selected according to Riede and Fitch (1999), using pictures of breeds representing different body sizes (small/medium/large) and consequently different VTLs: Yorkshire terrier, Irish setter, Rottweiler.
· OBJECTS: violin, viola, cello.

3.2.3 EXPERIMENTAL PROCEDURE

The experiment was carried out with an *acoustic stimulus-to-picture* matching technique and was implemented for computer presentation using E-prime software (Psychology Software Tools), a touch screen (Elo Entuitive TouchMonitor) and headphones.

A first *training session* (with feedback and verbal reward) was done to be sure that the task was completely understood. 3 pictures of different categories appeared on the touch screen and a sound (e.g. a bird, a musical instrument –drum- and a car) was presented on headphones at the same time. The child was required touch the correct sound source. 6 different sounds were played back (2 for each sound source). The child was part of the final sample if they performed 5 out of 6 correct trials.

In the *test phase*, the child was presented with a slide of three pictures for each category and listened to a corresponding sound. The child was required choose which of the represented characters is “making” the sound by touching the corresponding picture on the screen.

The order of presentation of different categories was randomized across subjects. Within each category, the order of acoustic stimuli was also randomized across subjects.

3.2.4 *DATA ANALYSIS*

Data on accuracy for each subject was analyzed with nonparametic tests (Kruskal Wallis/ Friedman Anova). Post-hoc tests corrected for multiple comparisons were carried out in cases with significant results. Data on reaction time was also tested with nonparametric tests.

3.2.5 *SUBJECTS*

A total of 54 children of both sexes in kindergartens in Trieste, Italy were used, ranging from 27 to 72 months of age when tested. Seven of them did not pass the training phase (3 3-year olds; 2 4-year olds and 2 5-year olds). A control group of 20 adults was tested. The final sample was composed as follows:

- 2-year-olds: 28.6 ± 1.4 months - tested 5 (2 females)
- 3-year-olds: 35.2± 4.2 months - tested 11 (2 females)
- 4-year-olds: 48.0 ± 2.9 months - tested 10 (5 females)
- 5-year-olds: 60.0 ± 3.7 months - tested 12 (6 females)
- 6-year-olds: 69.1 ± 1.5 months - tested 8 (4 females)
- Adults: tested 20 (10 females)

All the subjects were tested after informed consent by their parents. None of the children were reported as having hearing disorders.

3.3 RESULTS

3.3.1 *EFFECT OF AGE ON ACCURACY*

Accuracy increased with age (Fig 11). A positive significant correlation was found between age group and total accuracy (Spearman Rank correlation: $N = 66$; $\rho = 0.75$; $p < 0.001$), as well as individually for the dog category ($\rho = 0.36$; $p = 0.003$), speech category ($\rho = 0.67$; $p < 0.001$), and musical instruments category ($\rho = 0.59$; $p < 0.001$).

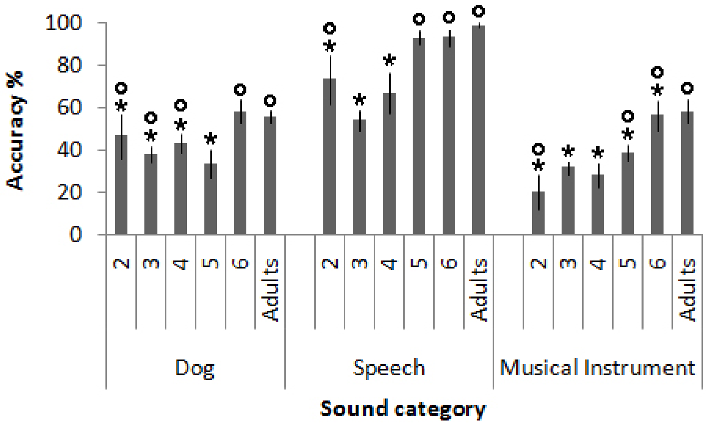

Fig 11 Mean accuracy (±SE) in each age group for each sound category. Bars with different symbols on the top are significantly different from each other, according to *post-hoc* multiple comparisons.

In the speech category, accuracy changed among age groups (Kruskal Wallis test: H (df = 4, N = 61) =35.57 p < 0.001). Post hoc multiple comparisons revealed significant differences between adults and 3-year-olds (p < 0.001); between adults and 4-year-olds (p = 0.013); between 5-year-olds and 3-year-olds (p = 0.008), and between 6-year-olds and 3-year-olds (p = 0.013).

In the Dog category accuracy also changed among age groups (Kruskal Wallis test: H (df = 4, N = 61) =17.58 p = 0.002). Post hoc multiple comparisons revealed significant differences between Adults and 5-year-olds (p = 0.015), and between 5-year-olds and 6-year-olds (p = 0.048).

In the Musical Instrument category, accuracy changed among age groups (Kruskal Wallis test: H (df = 4, N = 61) = 20.82 p < 0.001). Post hoc multiple comparison revealed significant differences between adults and 3-year-olds (p = 0.016), and adults and 4-years-olds (p = 0.012).

3.3.2 *DIFFERENCES IN ACCURACY AMONG CATEGORIES*

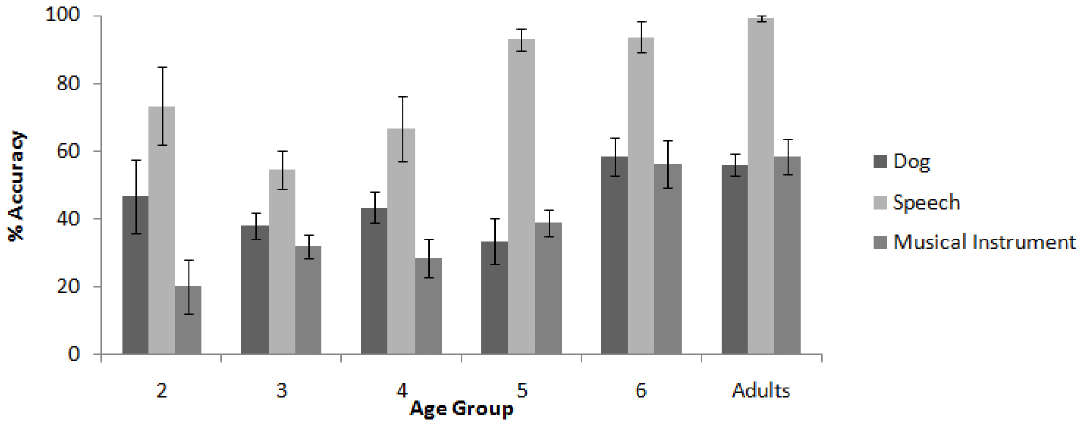

Fig 12 Mean accuracy (±SE) in the three sound categories in the different age groups

The level of accuracy for each category of sound in each age group was tested with Friedman Anova. In cases where $p < 0.05$, post-hoc tests corrected for multiple comparisons (Siegel and Castellan 1988) were made in order to investigate pair-wise differences. The test was significant within each age group (Adults: χ^2 (N = 20, df = 2) = 30.84 $p < 0.001$; 2 year-olds: χ^2 (N = 5, df = 2) = 8.40 p = 0.015; 3 year-olds: χ^2 (N = 11, df = 2) = 8.42 p = 0.015; 4 year-olds: χ^2 (N = 10, df = 2) = 6.53 p = 0.038; 5 year-olds: χ^2 (N = 12, df = 2) = 18.32 $p < 0.001$; 6 year-olds: χ^2 (N = 8, df = 2) = 10.78 p = 0.004). Multiple comparisons between the accuracy with Speech sounds and Musical instruments were significant for Adults, 2 year-olds, 3 year-olds, 5 year-olds, and 6 year-olds. The comparison between Speech sounds and Dogs was significant only for Adults and 5 year-olds ($p < 0.05$). The mean accuracy for speech sounds was always higher than the accuracy for other categories in all groups (Fig 12).

3.3.3 Reaction time (RT) analysis

For each category of stimuli, a Friedman ANOVA was run to analyze the distribution of mean RTs to see how relatively difficult it was to judge the size of a source, based on the hypothesis that longer RTs mean more difficult size judgments.

There were no significant differences in reaction times among Dog or Musical Instrument stimuli, but the Friedman ANOVA was significant for speech stimuli (χ^2 (N = 66, df = 2) = 16.12 $p < 0.001$). Post-hoc multiple comparisons revealed significant differences between the medium size and the small size (N = 66, $p < 0.05$); the medium size was the most difficult to judge (Fig 13).

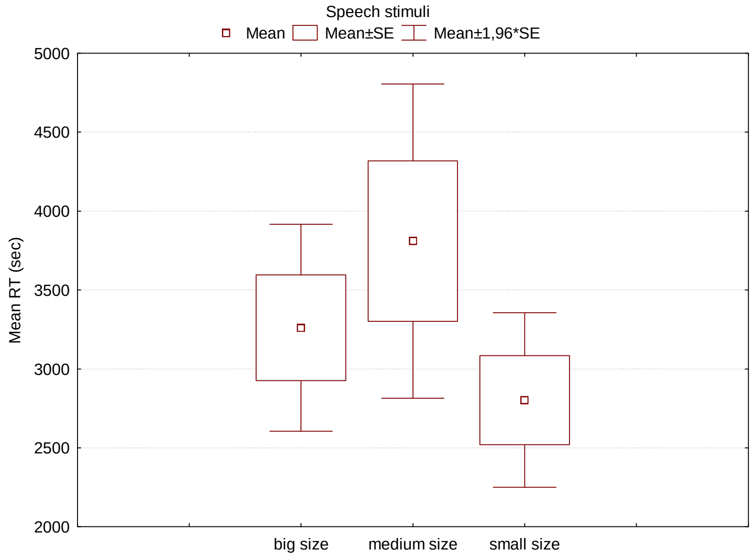

Fig 13 Mean reaction time (±SE) according to size judgment in speech stimuli

3.4 Discussion

According to Smith and colleagues (2005), the ability of the brain to extract size information from sounds is an inborn trait, and this information is extracted at an early stage of auditory processing. The results obtained here clearly show that performance in size attribution is different for different types of stimuli, even in the adult control group. This could mean either that the brain needs experience with particular types of objects, and practice with them, to extract size information from sound stimuli, or that the process of extraction uses different mechanisms for different sounds, and these develop according to a different schedule. However, since adults should presumably have these abilities in a fully-developed form, we believe the most likely explanation is that the process of size extraction is experience-dependent.

Consistent with experience-dependence, size extraction accuracy gradually increased with age. Probably because of their greater exposure to speech sounds as compared with dog growls and string instruments, the accuracies in the speech category were above chance in all the age groups, even though the 2-, 3- and 4- year-old children were less precise in associating a voice with the correct-sized speaker, while from the age of 5 performance did not differ from adult performance.

With the exception of the Musical Instrument category for 2 year-olds, children always performed above chance with the other two sound categories. These results confirm in part what Mondloch and Maurer found in their 2004 study: children as young as 2 reliably match high-pitched sounds with smaller objects (and low-pitched sounds with bigger objects), and their performance became similar to adults from the age of 5.

Reaction time analysis showed that medium size was the most difficult for subjects to judge.

4 PERCEPTION OF SIZE-RELATED FEATURES IN EXPERIMENTALLY MODIFIED VOICES IN NORMAL-HEARING CHILDREN

4.1 INTRODUCTION

Everyday perceived sounds give us a wealth of information beyond the general identity of the object that produced them. In the case of speech sounds, this information also goes beyond the semantic meaning of the words being produced. For example, when we chat with someone by phone, we can infer whether we are talking with a man, a woman, or a child; when we listen to music, we are able to say whether the same note is played by a violin or a contrabass. Many natural sounds contain information about the size of the object that produced them. For both animate and inanimate objects, bigger objects generally produce sounds with relatively lower pitches.

For mammals that produce sounds with their larynges, two acoustic features are linked to the size of the sound-producing source: the fundamental frequency F0 (or glottal pulse rate GPR), and the timbre or pattern of resonances (the frequency distribution of the energy, where resonances in the vocal tract attenuate some frequencies more than others).

Size attribution experiments show that adults are able to make fine judgments about the size of a speaker relying only on these two cues. Both GPR and VTL contribute equally to the final judgment, if these features fall in the "normal" range of sound characteristics for human vocal tract sizes, while VTL information is more important when abnormal combinations of low GPRs and short VTLs are tested that exceed the normal range of variation for human vocal tracts (Smith and Patterson 2005). Size attribution was claimed by these authors to be an inborn, automatic process in which scale-invariant vocal tract information is segregated as an "auditory image". It was also hypothesized that this process occurs at an early stage of the auditory pathway, at the level of the brainstem or of the thalamus (Irino and Patterson 2002).

However, there have been no prior examinations of the way in which acoustic size attribution develops in humans. Are these supposedly “automatic, early” processes fully functional from very early stages of postnatal life? Do GPR and VTL play similar roles in acoustic size judgment carried out by children as they do for adults?

At the physiological level, peripheral auditory function matures by the end of the first few postnatal months, while the development of the central auditory system appears to continue for at least several years (Moore 2002). For example myelination of the thalamic fibers innervating the auditory cortex progresses until the fourth year. Conduction pathways may not function in a fully mature fashion for a protracted period of early postnatal life, and this could affect auditory temporal and spectral processing. Indeed, performance on highly-demanding listening tasks remains at reduced levels compared to adult performance for many years after birth (Moore 2002), even though it is unclear how much of this is due to differences in auditory processing rather than other abilities that may affect these tasks. Despite the fact that their auditory systems may function differently than those of adults, newborns as old as one month can reliably recognize their mother’s voice (Mehler et al. 1978), suggesting that many auditory processing abilities function at a fairly sophisticated level.

As previously stated in chapter 3, very few studies have addressed the development of size attribution to sounds by children. Does a child need much training to extract information about the size of a source only from acoustic cues, or is this something children do naturally from an early age? Which acoustic features are important for this task?

The aim of the following experiment was to see whether children rely on F0 and timbre in a similar way to adults when they carry out an acoustic size judgment task. In order to understand in a more detailed way how preschool children use F0 and timbre to associate a voice with sex and size characteristics, a natural voice from three different speakers (a man, a woman and a child) was modified in VTL or F0 using the STRAIGHT vocoder (Kawahara et al. 1999; Smith et al. 2005). Each stimulus was modified keeping either the original F0 or the original VTL to test how speaker identification changed according to each acoustic parameter.

4.2 MATERIALS AND METHODS

4.2.1 ACOUSTIC STIMULI

Original speakers (one man, one child and one woman) recorded two sentences with varying pitch contours: a statement: (non si puo’ aprire- You cannot open it-) and a question (ti piace il riso? -Do you like rice?-). Mean F0s of each original sentence were calculated using SIGNAL software (Engineering Design, Berkeley, CA, USA). Vocal tract length was calculated from height according to the regression equation suggested by Fitch and Giedd (1999):

$VTL = 2.7 + 0.68\ Ht$

Where:

VTL is vocal tract length in cm

Ht is the height in cm

Each sentence was modified using the STRAIGHT vocoder (Kawahara et al. 1999) in order to have two experimental conditions:

Original VTL, Varying F0: F0s scaled to have mean values of 40, 80, 160, 320, 640 Hz.

Original F0, Varying VTL: VTL factor scaled to have simulated vocal tract lengths of 6.52, 8.81, 11.91, 15.97, 21.69 cm.

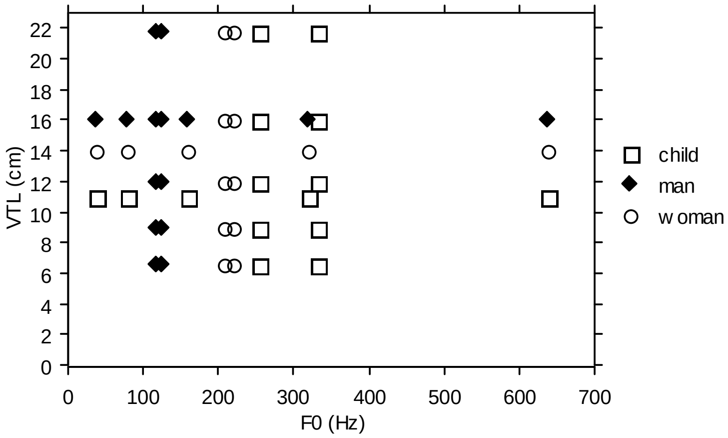

Fig 14 Combinations of F0 and VTL for all the stimuli used in Experiment 2a, 2b and 2c for each original speaker

After VTL manipulation, the first three formants of each stimulus were measured with PRAAT software; and the formant dispersion (Fd) was also calculated. These measures were compared in order to check whether the scale information from the three original voices differed after manipulation. Kruskal-Wallis tests were run for each manipulated sentence. F2, F3 and Fd of both the statement and the question were not always significantly different from each other ($p > 0.05$). The test was significant among the three manipulated voices for the F1 of the question ($H(df = 2, N = 15) = 6.02$, $p = 0.049$); multiple comparisons revealed that the F1 of the manipulated child voice was significantly lower than the manipulated man's voice, but that the other comparisons were not significantly different.

4.2.2 VISUAL STIMULI

For all experiments, visual stimuli were pictures of one adult male, one adult female and one child chosen randomly from pictures of 5 different women, 5 different men and 5 different children. A picture for each category was always presented in the same position on the touch screen of a computer monitor: the child on the left, the woman in the centre and the man on the right. Pictures appeared simultaneously along with each acoustic stimulus.

4.2.3 EXPERIMENTAL PROCEDURE

Training Phase: In the first *training phase* (with feedback and verbal reward) 3 pictures of different objects appeared on the touch screen, and a sound (e.g. a bird, a musical instrument [drum], or a car) was presented on headphones at the same time. The child was required to touch the correct picture indicating the sound source. Six different sounds were played back (2 for each sound source). This phase ended when a child answered 5 out of 6 trials correctly. If a subject did not pass the first training session, a second one with 6 trials was given. On average, subjects answered correctly to 6.03 ± 0.08 (SEM) trials to reach the test phase.

Testing Procedure: A 3-AFC paradigm was used to test children with 20 items (2 sentences with 5 values of F0 and 5 VTL). The procedure was the same as the one used in the training phase, except that the pictures of a man, a woman and a child were used. Stimuli in the two conditions (“Original F0 – Varying VTL”, and “Original VTL – Varying F0”) were presented in random order, and subjects attributed a speaker to each auditory stimulus them by touching the appropriate choice from among the 3 pictures simultaneously presented on the touch screen.

4.2.4 DATA ANALYSIS

A score was assigned to each chosen speaker: 1 to the child, 2 to the woman, 3 to the man. In this way each subject’s score for each condition ranged from 2 to 6: the higher the score, the bigger the size of the speaker associated with a certain voice. The F0 and VTL data were analyzed using the Friedman ANOVA, followed by post-hoc tests corrected for multiple comparisons (Siegel and Castellan 1988), with a criterion of $p < 0.05$ (two-tailed).

The partial Spearman rank correlation coefficients between F0. VTL and Perceptual Performance were calculated for each subject and for each condition, and then squared to give the proportion of the variance in perceptual performance that was explained by F0 and VTL (partial $\rho^2_{(VTL/F0)}$).

4.2.5 SUBJECTS

All the subjects came from different kindergartens in Trieste, Italy. They were tested in a quiet room at their schools after obtaining their parents' informed consent. None of the subjects was reported as having hearing disorders. In each experiment a control group of 10 adults was also tested.

4.2.5.1 Experiment 2a: Adult male voice

54 children of both sexes, ranging from 30 to 76 months in age were tested. Two 3-year-olds did not pass the training phase. The final sample was composed of:

- 3 year-olds: 34.57 ± 3.94months - tested 13 (5 females)
- 4 year-olds: 48.77 ± 3.47 months-tested 17 (7 females)
- 5 year-olds: 59.39 ± 3.10 months -tested 13 (8 females)
- 6 year-olds: 71.00 ± 2.40 months -tested 9 (3 females)
- Adults: tested 10 (6 females)

4.2.5.2 Experiment 2b: Adult female voice

55 children of both sexes, ranging from 36 to 70 months in age were tested. Two 3-year-olds did not pass the training phase. The final sample was composed as follows:

- 3 year-olds: 38.50± 2.14 months - tested 14 (6 females)
- 4 year-olds: 48.44± 3.07 months-tested 19 (9 females)
- 5 year-olds: 60.33± 3.46 months -tested 15 (10 females)
- 6 year-olds: 69.00± 1.23 months -tested 5 (2 females)
- Adults: tested 10 (4 females)

4.2.5.3 Experiment 2c: child voice

50 children of both sexes, ranging from 35 to 64 months in age were tested. One 4-year-old did not pass the training phase. The final sample was composed as follows:

3 year-olds: 38.57 ± 1.83months - tested 14 (3 females)

4 year-olds: 49.11 ± 2.87 months-tested 19 (10 females)

5 year-olds: 58.50 ± 3.48 months -tested 16 (3 females)

Adults: tested 10 (6 females)

4.3 RESULTS

Features of the original speakers and of their sentences are summarized in

Table **9.**

Original Speaker	Height (cm)	F0 question	F0 statement	Estimated VTL (cm)
Adult man	174	119 Hz	127 Hz	14.53
Adult woman	164	208 Hz	222 Hz	13.85
Child	121	254 Hz	334 Hz	10.93

Table 9 Features of original speakers and their sentences

4.3.1.1 Experiment 2a: Adult male voice

Original F0, varying VTL Condition

Considering all the age groups together, perceived size changed as a function of VTL variation (χ^2 (N = 62, df = 4) = 117.39 p < 0.001,Fig 15). Post hoc multiple comparisons revealed that the score at 15.97 cm VTL and at 21.69 cm VTL were significantly higher than at 6.52 cm, 8.81 cm and at 11.91 cm ($p < 0.05$). Moreover, the score at 8.81 cm VTL was lower than at 11.91 cm VTL, indicating that longer VTLs were associated with a bigger speaker, while shorter VTLs were associated with a smaller speaker.

The same was true for each age group individually for tests of significant heterogeneity among the stimuli (3-year-olds: χ^2 (N = 13, df = 4) = 9.65 p = 0.047; 4-year-olds: χ^2 (N = 17, df = 4) = 38.00 $p < 0.001$; 5-year-olds: (N = 13, df = 4) = 34.48 $p < 0.001$; 6-year-olds: χ^2 (N = 9, df = 4) = 21,95 $p < 0.001$; ADULTS: χ^2 (N = 10, df = 4) = 32.45 $p < 0.001$). With the exception of the 3-year-old group, post-hoc multiple comparisons revealed that the score in each condition changed significantly with VTL length (Fig 15).

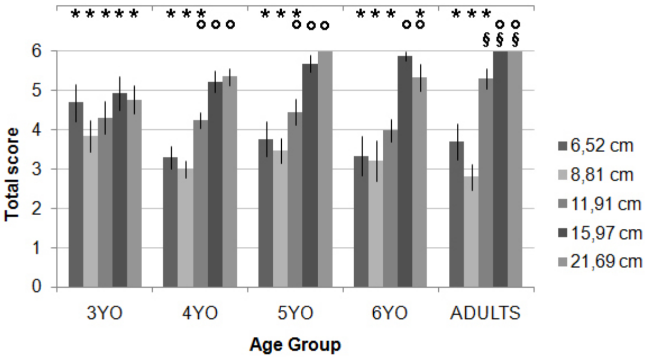

Fig 15 Total score (±SE) according to different perceived VTLs of the modified man's voice in the four age groups and in the adult control group. Bars with different symbols on the top are significantly different from each other (always $p < 0.05$), according to *post-hoc* multiple comparisons.

Original VTL, varying F0 Condition

For all age groups analyzed together, the score decreased with increasing F0 (X^2 (N = 62, df = 4) = 79.12 p = 0.001), and the scores at 40 Hz, 80 Hz, 160 Hz were significantly lower than at 320 Hz and at 640 Hz (Fig 16). This indicates that perceived size also changed as a function of F0 variation.

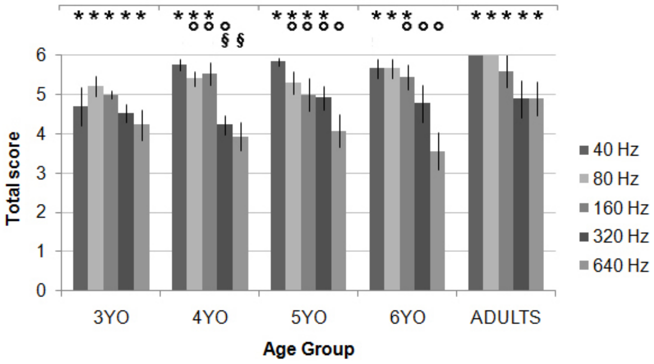

Fig 16 Total score (±SE) according to different F0 of the modified man's voice in the four age groups and in the adult control group. Bars with different symbols on the top are significantly different from each other (always $p < 0.05$), according to *post-hoc* multiple comparisons.

The same general pattern was repeated in each age group for tests of significant heterogeneity among the stimuli: 3-year-olds (χ^2 (N = 13, df = 4) = 9.74 p = 0.045); 4-year-olds (χ^2 (N = 17, df = 4) = 34.84 $p < 0.001$); 5-year-olds (χ^2 (N = 13, df = 4) = 13.71 p = 0.008); 6-year-olds (χ^2 (N = 9, df = 4) = 21.79 $p < 0.001$); adults (χ^2 (N = 10, df = 4) = 15.21 p = 0.004). Post-hoc multiple comparisons did not show significant differences according to F0 in 3-year-olds and adults, while the scores changed significantly with condition in the 4-year-old group, the 5-year-old group and the 6-year-old group (Fig 16).

4.3.1.2 Partial correlation coefficients

Fig 17 shows that partial ρ^2 significantly differed among age groups (for partial $\rho^2_{(VTL)}$: Kruskal-Wallis test: H (4, N = 62) =15.18 p = 0.004; for partial $\rho^2_{(F0)}$: Kruskal -Wallis test: H (4, N = 52) = 13.05 p = 0.01)

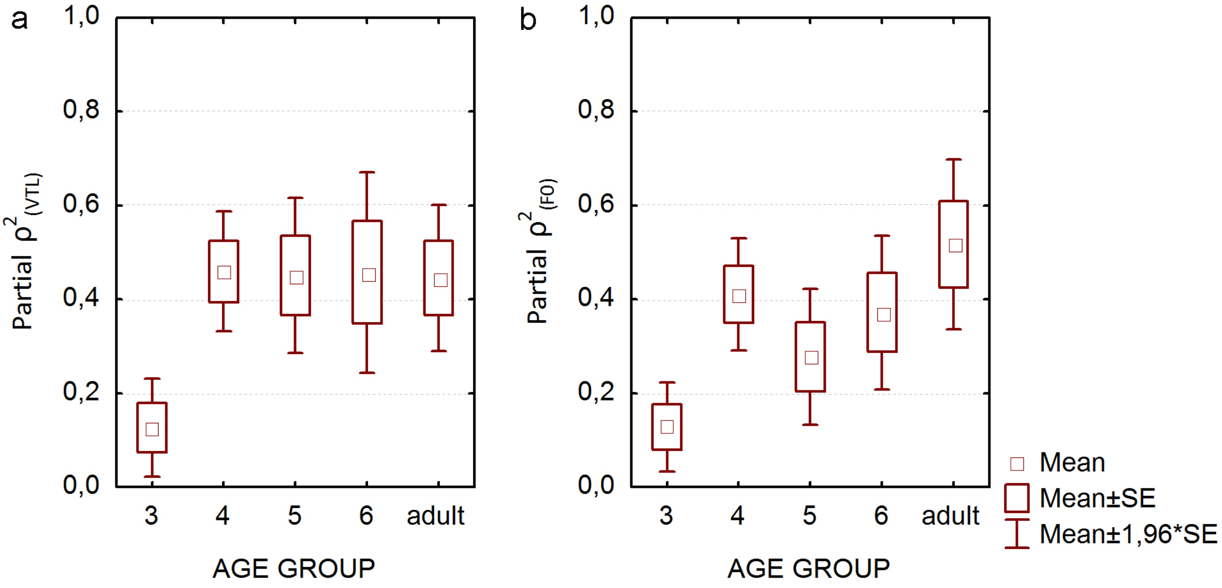

Fig 17 Boxplot showing a) partial $\rho^2_{(VTL)}$ and b) partial $\rho^2_{(F0)}$ in the different age groups. Multiple comparisons revealed that in both cases the values for 3 year-old subjects were lower than the partial ρ^2 of older groups.

Multiple comparisons revealed that for the 3 YO group, partial ρ^2(VTL) was significantly lower than for 4 YOs ($p = 0.008$), 5 YOs ($p = 0.03$) and adults ($p = 0.03$), and that the partial $\rho^2_{(F0)}$ for the 3YOs was lower than for the 4 YOs ($p = 0.03$) and adults ($p = 0.03$). In both cases, the perceptual performance of the 3 YOs explained an average of less than 20% of the variance in perceptual performance accounted for by either F0 or VTL.

For each age group, we compared partial $\rho^2_{(F0)}$ and partial $\rho^2_{(VTL)}$ using Wilcoxon Tests in order to check whether size perception was more influenced by one of the two acoustic parameters. No significant difference was found for 3 YOs ($p = 0.94$), 4 YOs ($p = 0.18$), 6 YOs ($p =1.0$) and adults ($p = 0.50$), while for 5 YOs partial $\rho^2_{(VTL)}$ was significantly higher than partial $\rho^2_{(F0)}$ ($p = 0.04$).

Experiment 2b: Adult female voice

Original F0, varying VTL Condition

The scores significantly changed according to changing VTL (χ^2 (N = 63, df = 4) = 114.72 p < 0.001, Fig 18). Post-hoc multiple comparisons revealed that the scores at 15.97 cm VTL and at 21.69 cm VTL were significantly higher than at 6.52 cm, 8.81 cm and at 11.91 cm (p < 0.05).

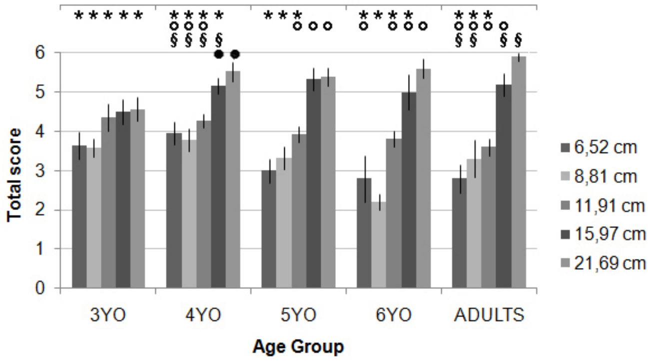

Fig 18 Total score (±SE) according to different perceived VTLs of the modified woman's voice in the four age groups and in the adult control group. Bars with different symbols on the top are significantly different from each other (always p < 0.05), according to *post-hoc* multiple comparisons.

The same was true in each age group for tests of significant heterogeneity among the stimuli (3-year-olds: χ^2 (N = 14, df = 4) = 10.45 p = 0.033; 4-year-olds: χ^2 (N = 19, df = 4) = 34.72 p < 0.001; 5-year-olds: χ^2 (N = 15, df = 4) = 32.58 p < 0.001; 6-year-olds: χ^2 (N = 5, df = 4) = 16.62 p = 0.002; ADULTS: χ^2 (N = 10, df = 4) = 32.81 p < 0.001). Multiple comparisons were not significant in 3-year-olds, while in the other age groups the score changed significantly with simulated VTL length (Fig 18).

Original VTL, varying F0 Condition

Considering all the age groups combined, the score decreased with increasing F0 (χ^2 (N = 63, df = 4) = 113.79 $p < 0.001$): the scores at 40, 80 and 160 Hz were significantly higher than the scores at 320 and at 640 Hz ($p < 0.05$, Fig 19).

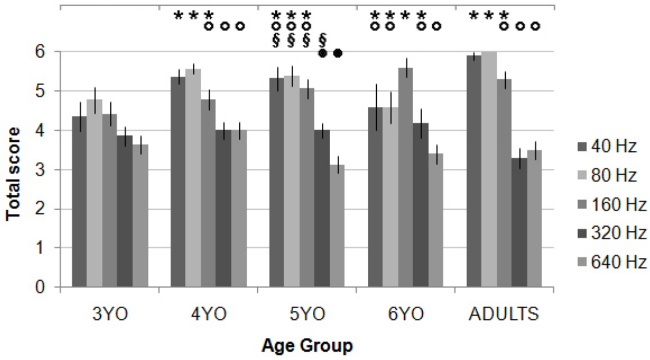

Fig 19 Total score (±SE) according to different F0 of the modified woman's voice in the four age groups and in the adult control group. Bars with different symbols on the top are significantly different from each other (always $p < 0.05$), according to *post-hoc* multiple comparisons.

However, there was not significant heterogeneity among the scores for different stimuli in the 3-year-olds, while there was in all the other groups (4-year-olds: χ^2 (N = 19, df = 4) = 40.31 $p < 0.001$; 5-year-olds: χ^2 (N = 15, df = 4) = 37.00 $p < 0.001$; 6-year-olds: χ^2 (N = 5, df = 4) = 10.09 $p = 0.039$; ADULTS: χ^2 (N = 10, df = 4) = 35.14 $p < 0.001$). Multiple comparisons showed that the score changed significantly with F0 in the 4-, 5-, 6-year-old group and in the adult control group (Fig 19): in general lower F0s were associated with higher scores (hence larger speaker sizes).

Partial correlation coefficients

Fig 20 Boxplot showing a) partial ρ2(VTL) and b) partial ρ2(F0) in the different age groups. Multiple comparisons revealed that in both cases the values for 3 years-old subjects were lower than the partial ρ2 of older groups. (partial $\rho^2_{(VTL)}$: Kruskal-Wallis test: $H(4, N = 63) = 14.05$ $p = 0.007$; partial $\rho^2_{(F0)}$: Kruskal-Wallis test: $H(4, N = 63) = 18.78$ $p < 0.001$).

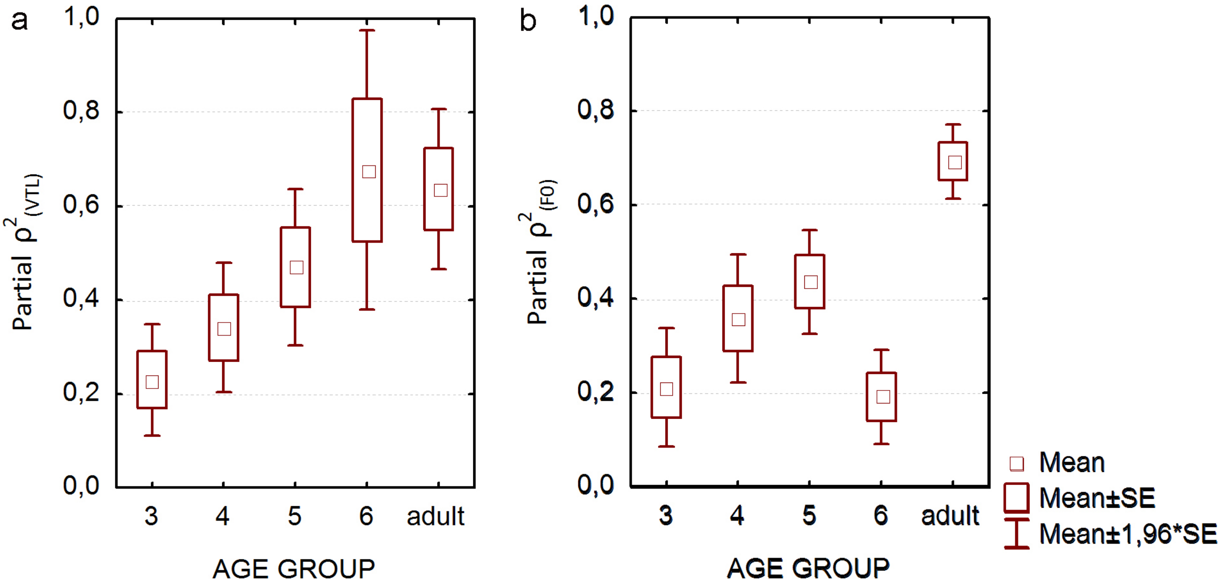

Fig 20 Boxplot showing a) partial $\rho^2_{(VTL)}$ and b) partial $\rho^2_{(F0)}$ in the different age groups. Multiple comparisons revealed that in both cases the values for 3 years-old subjects were lower than the partial ρ^2 of older groups.

Multiple comparisons revealed that for the 3 YO group, partial $\rho^2_{(VTL)}$ was significantly lower than for adults ($p = 0.02$), and that the partial $\rho^2_{(F0)}$ was higher for adults than for 3YOs ($p < 0.01$), 4 YOs ($p = 0.03$) and 6 YOs ($p = 0.03$). In this experiment, partial $\rho^2_{(VTL)}$ increased gradually as the subjects got older, while the partial $\rho^2_{(F0)}$ was low for child groups (around 30% of the variance) and increased to around 75% in the adult control group.

For each age group, we compared partial $\rho^2_{(F0)}$ and partial $\rho^2_{(VTL)}$ using Wilcoxon Tests in order to check whether size perception was most influenced by one of the two acoustic parameters or not. No significant differences were found in any of the age groups: 3 YOs ($p = 0.81$), 4 YOs ($p = 0.90$), 5 YOs ($p = 0.53$), 6 YOs ($p = 0.07$) and adults ($p = 0.72$).

Experiment 2c: Child voice

Original F0, varying VTL Condition

For all ages combined, the score increased with increasing perceived VTL (χ^2 (N = 59, df = 4) = 91.08, $p < 0.001$). Post-hoc multiple comparisons revealed that the scores at 15.97 cm VTL and at 21.69 cm VTL were significantly higher than at 6.52 cm, 8.81 cm and at 11.91 cm ($p < 0.05$, Fig 21).

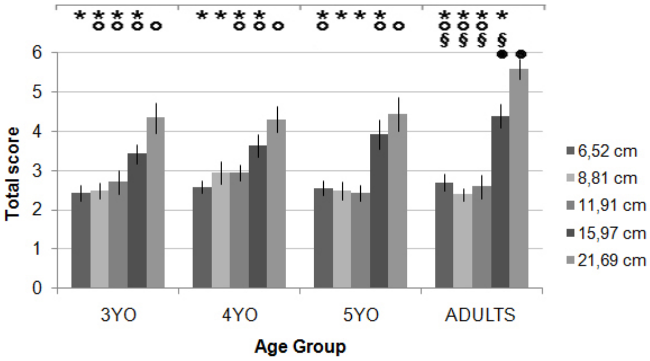

Fig 21 Total score (±SE) according to different perceived VTLs of the modified child's voice in the three age groups and in the adult control group. with different symbols on the top are significantly different from each other (always $p < 0.05$), according to *post-hoc* multiple comparisons.

The same was true in each age group for tests of significant heterogeneity among the stimuli (3-year-olds: χ^2 (N = 14, df = 4) = 19.03 $p < 0.001$; 4-year-olds: χ^2 (N = 19, df = 4) = 23.01 $p < 0.001$; 5-year-olds: χ^2 (N = 16, df = 4) = 23.37 $p < 0.001$; ADULTS: χ^2 (N = 10, df = 4) = 32,16 $p < 0.001$). According to multiple comparisons, in each age group and in the adult control the score changed significantly with simulated VTL length, with higher scores associated with longer VTLs (Fig 21).

Original VTL, varying F0 Condition

For all ages combined, the score decreased with increasing F0 (χ^2 (N = 59, df = 4) = 28.69 $p < 0.001$); post-hoc tests showed that the score at 40 Hz was significantly higher than the score at 320 and at 640 Hz ($p < 0.05$).

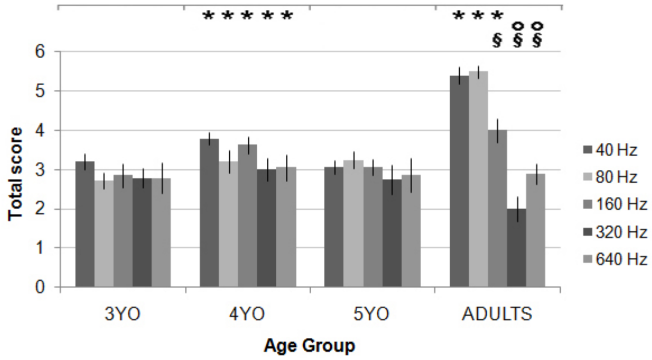

Fig 22 Total score (±SE) according to different F0 of the modified child's voice in the four age groups and in the adult control group.

Looking more in detail at each age group, only the 4-year-olds showed significant heterogeneity among the stimuli (χ^2 (N = 19, df = 4) = 12.93 p = 0.011), as did the adult controls (χ^2 (N = 10, df = 4) = 31.98 $p < 0.001$). However, post-hoc multiple comparisons within the 4-year-old group did not show any significant differences, while these existed in the adult control group (Fig 22).

4.3.1.3 Partial correlation coefficients

Fig 23 shows that partial ρ^2 significantly differed among age groups (for partial $\rho^2_{(VTL)}$: Kruskal-Wallis test: H (3, N = 59) = 9.12 p = 0.03; for partial $\rho^2_{(F0)}$: Kruskal -Wallis test: H (3, N = 59) = 12.72 p = 0.005).

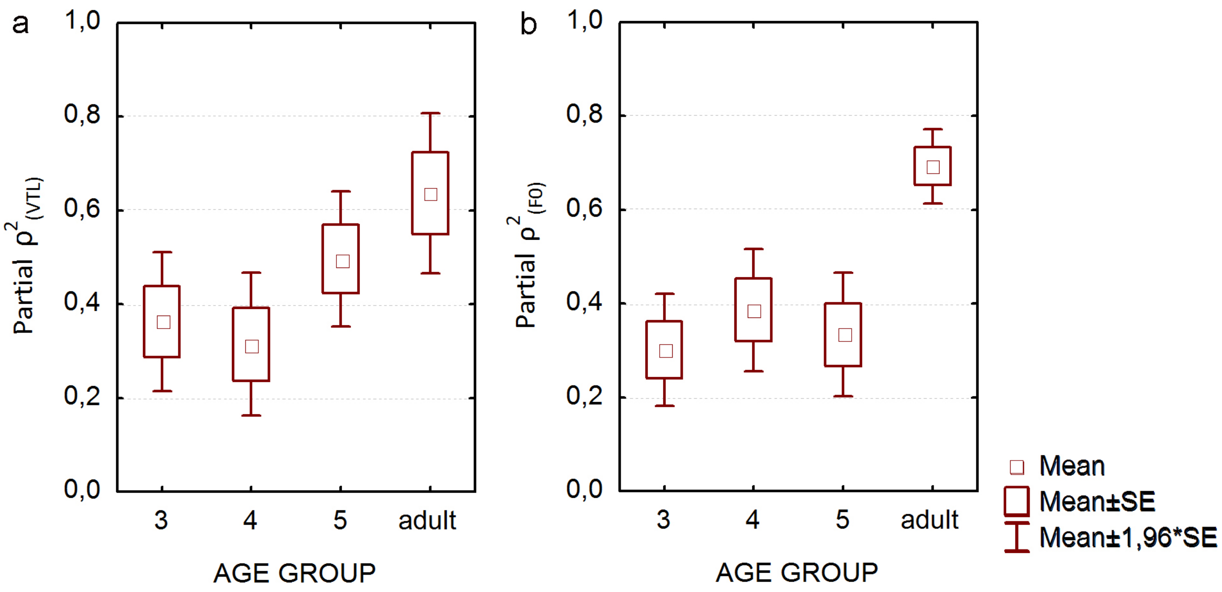

Fig 23 Boxplot showing a) partial ρ2(VTL) and b) partial ρ2(F0) in the different age groups.

Multiple comparisons revealed that for the adult group, partial $\rho^2_{(VTL)}$ was significantly higher than for 4 YOs ($p = 0.04$), and that the partial $\rho^2_{(F0)}$ was higher in adults than in 3 YOs ($p = 0.06$), 4 YOs ($p = 0.03$) and 5 YOs ($p = 0.01$).

For each age group, we compared partial $\rho^2_{(F0)}$ and partial $\rho^2_{(VTL)}$ using Wilcoxon Tests in order to check whether size perception was most influenced by one of the two acoustic parameters or not. No significant difference was found for 3 YOs ($p = 0.65$), 4 YOs ($p = 0.66$), 5 YOs ($p = 0.08$) or adults ($p = 0.72$).

4.4 DISCUSSION

In the preceding Chapter 3, children as young as 2 reliably recognized the size classes of speakers in an adult-like fashion using natural speech stimuli (where both F0 and VTL information are available at the same time, and this information is congruent). On the other hand, the results of the present experiments with modified voices suggest that when only *one* of these parameters changed, and there was a disparity in these two types of information, there was a difference in the way that adults and children performed.

The present results confirm the trend shown by previous research in adults: a longer perceived VTL and a lower F0 are associated with a bigger speaker, with size judgment changes according to both subjects' age and original speaker's voice (Smith et al. 2005).

The size judgment of the 3 year-old group was, in general, less influenced by changes in either VTL or F0; with the exception of the child's voice original F0-varying VTL condition, the score did not change significantly with changing acoustic parameters. When significant heterogeneity was found, post-hoc multiple comparisons were generally not significant. On the other hand, from the age of 4 onwards, changing either one of these parameters significantly influenced size judgments.

There was also an effect of the original speaker's voice. While changes in the perceived VTL influenced size judgments of all the age groups, this was not true for changes in F0 (the original VTL, varying F0 condition). All the age groups changed their judgments for the stimuli made with the adult male voice. Three year-olds did not change their judgments for the stimuli made with the adult female voice, while other age groups did; and none of the child groups reliably changed their judgments for the stimuli made with the child's voice (the 4 year-olds showed significant heterogeneity among the stimuli, but no post-hoc multiple comparisons were significant), while only adults did.

These results suggest that with sizes associated with a bigger formant dispersion, children rely less on F0 when making acoustic size judgments. VTL appears to be initially more heavily weighted by children when determining the size of a speaker from acoustic information, and the ability to use both F0 and VTL information equally for performing speaker size judgments appears to require additional maturation and/or experience. In particular, we noted that adults were able to rely on F0 for size judgments much more than children, especially when listening to high-pitched voices. It is unclear at present whether this might be due to some kind of immaturity of brain processing mechanisms for pitch perception (Moore 2002), or with the development of experience-dependent strategies for weighting different kinds of perceptual information.

5 PERCEPTION OF SIZE-RELATED FEATURES IN DOG GROWLS BY NORMAL-HEARING CHILDREN

5.1 INTRODUCTION

As already stated in the other chapters, formant dispersion can be used as a reliable cue for body size in many mammalian species included humans (Titze 1989; Fitch 1997; Riede and Fitch 1999; Reby and McComb 2003; Rendall et al. 2005). Mammals naturally attend to this acoustic parameter, and associate more spaced formants with a bigger animal and more clumped formants with a smaller animal (Smith et al. 2005; Fitch and Fritz 2006; Charlton et al. 2008). Fundamental frequency is a less reliable acoustic cue for body size, but it still appears to be used for distinguishing among sexes and among age groups (Fischer and Hammerschmidt 2002; Rendall et al. 2005; Pfefferle and Fischer 2006).

In all these studies, however, subjects (either animals or humans) were tested with vocalizations of their own species. The first study in which subjects were asked to infer body size from acoustic cues *inter-specifically* was by Taylor et al. (2008). Human adults were tested using dog growls. Dog vocalizations were chosen for several reasons: dogs live in close proximity with humans and communicate with them as much or more than with their conspecifics; dog breeds present a huge variation in terms of body size, and a close relationship between formant dispersion and body size has previously been shown in dogs (Riede and Fitch 1999).

In that study, two different groups of adult subjects were tested: one group using dog growls with unchanged F0 and manipulated formant dispersion, and a second group using growls with manipulated F0 and unchanged formant dispersion. Results showed that both the acoustic parameters influenced size perception: while formant dispersion provided a reliable and more salient cue for body size, a weak but significant effect of F0 was also found.

In the present experiment, we wanted to replicate the experiment of Taylor and colleagues, with some modifications:

1-we tested pre-school children because we wanted to check whether the ability to *inter-specifically* infer size from sound appears early during life, or develops much later.

2-we presented both stimulus conditions (varying VTL and varying F0) to the same group of subjects using a within-subjects design

3- in order to make the test suitable for children, we had to use a lower number of test items. We therefore changed both the original VTL and the original F0 by ±20%, hypothesizing that increasing the perceived VTL by 20% would be associated with a large-sized dog, while decreasing it by 20% would be associated with a small dog. We expected the opposite pattern for F0.

5.2 MATERIALS AND METHODS

5.2.1 ACOUSTIC STIMULI

Dog growls uttered by ten different dogs of different body sizes that were used in the study by Riede and Fitch (1999), and kindly provided by Dr. Tobias Riede, were used here as stimuli. Dog weight, Formant dispersion Fd and fundamental frequency F0 (calculated with the autocorrelation method) are reported in Table 10.

5.2.2 STIMULUS MANIPULATION

VTL change: VTL proportional changes of 1.2 or 0.8 were achieved using a Praat script originally created to resynthesize deer roars by Charlton et al. (2007), and originally written by C.J. Darwin. We modified each original growl in order to decrease or increase the apparent VTL by 20%. This PSOLA-based algorithm changed the apparent VTL by shifting the formants by a desired factor. In a first step, along with formants it also changed pitch and duration, but after resampling it brought these back to their original values. As a result, the formants are shifted while all the other parameters remain unchanged.

F0 change: In order to get manipulated growls that sounded as natural as possible, F0 manipulation was done in several steps. First SIGNAL software was used to increase or decrease the original F0 by 20% by increasing or decreasing the sampling rate by 20%, thus changing the frequencies of all components without changing their spacing. This altered sound was then resampled back to its original sampling rate (44.1 kHz) using cubic spline interpolation. It was then read into Praat, and the PSOLA algorithm was used to lengthen or shorten the manipulated sound back to its original duration.

All the stimuli were root-mean-square normalized in amplitude, then presented with headphones (Sennheiser HD515) at constant sound level.

5.2.3 VISUAL STIMULI

Visual stimuli were pictures of one small dog, one medium size dog and one large dog chosen randomly from pictures of 5 different small dogs (Dalmatian pup, Yorkshire Terrier, Pug, Pembroke Welsh Corgi, Chihuahua), 5 different medium size dogs (Cocker Spaniel, Fox Terrier, Dalmatian adult, Irish Setter, Poodle) and 5 different large dogs (Neapolitan Mastiff, Great Dane, Newfoundland, Rottweiler, German Shepherd). A picture for each category was always presented in the same position on the touch screen of a computer monitor (Elo 1526L 15” Medical Touchscreen Monitor, Resolution 1024 x 768): the small on the left, the medium in the centre and the big on the right. Pictures appeared simultaneously along with each acoustic stimulus.

5.2.4 EXPERIMENTAL PROCEDURE

Training Phase: In the first training phase (with feedback and verbal reward) three pictures of different categories appeared on the touch screen and a sound (e.g. a bird, a musical instrument –drum- and a car) was presented at the same time. The child was required touch the correct sound source. Six different sounds were played back (two for each sound source).

Children had to complete a training session with five out of six correct trials before they were tested.

Procedure: a 3-AFC paradigm was used to test subjects, with 50 items (10 original growls, 10 with 0.80 F0. 10 with 1.20 F0. 10 with 0.80 VTL and 10 with 1.20 VTL). Stimuli in the two conditions were presented in random order, and subjects attributed a source to them by touching the appropriate choice from among the 3 pictures simultaneously presented on the touch screen.

5.2.5 DATA ANALYSIS

A score was assigned to each chosen dog: 1 to the small dog, 2 to the medium-sized dog, and 3 to the big dog. In this way each subject's mean score for each condition ranged from 1 to 3: the higher the score, the bigger the size of the dog associated with a certain growl. Score variations according to F0 or VTL were tested with Friedman ANOVA followed by post-hoc test corrected for multiple comparisons (Siegel and Castellan 1988), using a criterion for significance of $p < 0.05$ (two-tailed).

5.2.6 SUBJECTS

Children from a kindergarten in Trieste were tested in a quiet room at their school after obtaining their parents' informed consent. None of the subjects was reported as having hearing disorders. The final sample comprised 31 children (age: 56.27 ± 13.61 months) and 10 adults all older than 25 years old as a control group. Subjects were also asked their level of familiarity with dogs.

5.3 RESULTS

5.3.1 ACOUSTIC STIMULI

Name	*Breed*	*Weight (Kg)*	*F0 (Hz)*	*Fd (Hz)*
sub16	Mongrel	10.00	420.57	1510
sub17	Dachshund	10.50	360.68	1229
sub26	Staffordshire terrier	25.00	414.51	
sub28	Giant Schnauzer	31.00	473.39	
sub37	Mongrel	8.00	398.36	1546
sub38	German shepherd	15.00	444.12	1102
sub39	Boxer	22.00	380.87	705
sub40	Rottweiler	25.00	357.99	678
sub41	Rottweiler	28.00	434.03	679
sub43	Mongrel	33.50	300.12	734

Table 10 Characteristics of the dogs making up the final sample and of their vocalizations (from Riede and Fitch 1999)

In this final sample there was a significant negative correlation between Fd and dog weight (Spearman rank correlation: ρ = -0.83, N = 8, p = 0.02), but no significant correlation between F0 and weight (ρ = -0.03, N = 10. p = 0.92).

5.3.2 FAMILIARITY WITH DOGS

50 % of the adults were familiar with dogs, while 40% had no experience with them; 10% of adults had little experience with dogs. Among children, 84% had no familiarity with dogs while 16 % were familiar with them.

5.3.3 SIZE JUDGMENTS

For each of the two conditions, the performance of adults and children were compared with a repeated measures ANOVA with 1 within-subject factor (VTL or F0; 3 levels: 0.8 original, original, 1.2 original), and GROUP (2 levels: adults *vs* children) as the between-subject factor. Another analysis was done with different age groups (3, 4, 5, 6 and adults), in this case one subject was excluded from the analysis because his age was not known.

Original VTL, varying F0 Condition

The assumptions for normality, homogeneity of variances, and sphericity were all met (K-S test for normality: 0.8*original, d=0.14, p> .20; original, d=0.11, p > 0.20; 1.2*original, d=0.10. p > 0.20.; Levene's Test for Homogeneity of Variances: 0.8*original, F(1,39) = 0.39, p = 0.54; original, F(1,39) = 1.92, p = 0.17; 1.2*original, F(1,39) = 2.07, p = 0.16; Mauchley's Sphericity Test: W = 0.99; Chi-Sqr. = 0.44, df = 2, P = 0.80).

The ANOVA revealed non-significant main effects of GROUP (F(1,39) = 0.08, p = 0.77) and VTL (F(2, 78) = 0.37, p = 0.70), and a non-significant VTL x GROUP interaction (F(2, 78) = 0.28, p = 0.76). In the ANOVA with AGE GROUP as a between-subject factor, none of the effects were significant (AGE GROUP: F(4, 35) = 0.64, p = 0.65) and VTL (F(2, 70) = 0.21, p = 0.81), the VTL x AGE GROUP interaction was also not significant (F(8, 70) = 0.40. p = 0.91).

Original F0, varying VTL Condition

The assumptions for normality, homogeneity of variances, and sphericity were all met (K-S test for normality: 0.8*original, d=0.18, p > .15; original, d=0.11, p > 0.20; 1.2*original, d=0.11, p > 0.20.; Levene's Test for Homogeneity of Variances: 0.8*original, $F_{(1,39)} = 0.44$, p = 0.51; original, $F_{(1,39)} = 1.92$, p = 0.17; 1.2*original, $F_{(1,39)} = 3.73$, p = 0.06; Mauchley's Sphericity Test: W = 0.99; Chi-Sqr. = 0.29, df = 2, P = 0.86).

The effect of the factor GROUP was not significant ($F_{(1,39)} = 0.05$, p = 0.82). On the other hand, the factor VTL was significant ($F_{(2,\ 78)} = 16{,}32$, p < 0.001): the mean score increased with perceived VTL (Mean ± SEM: 0.8 original: 2.01 ± 0.07; original 2.22 ± 0.06; 1.2 original 2.37 ± 0.06). Also the interaction VTL x GROUP was significant ($F_{(2,\ 78)} = 6.06$, p = 0.004). Children were less sensitive than adults to variation in perceived VTL, and gave a higher score (2.14 ± 0.07) to the "small" VTL as compared with adults (1.88 ± 0.11), but the difference was not significant (Bonferroni-corrected p = 0.68). They also assigned a lower score (2.28 ± 0.06) as compared with adults (2.46 ± 0.11) to the "big" perceived VTL, although this difference was also not significant (p = 1.00). In adults, the score at 0.8*original was significantly lower than the score at the original VTL (p = 0.04) and at 1.2*original VTL (p < 0.001), while the score at original VTL was not significantly lower than at 1.2*original VTL (p = 0.49).

In children none of these contrasts was significant (0.8 vs original: p = 1.00; 0.8 vs 1.2 p = 0.39; original vs 1.2 p = 1.00) (Fig 24).

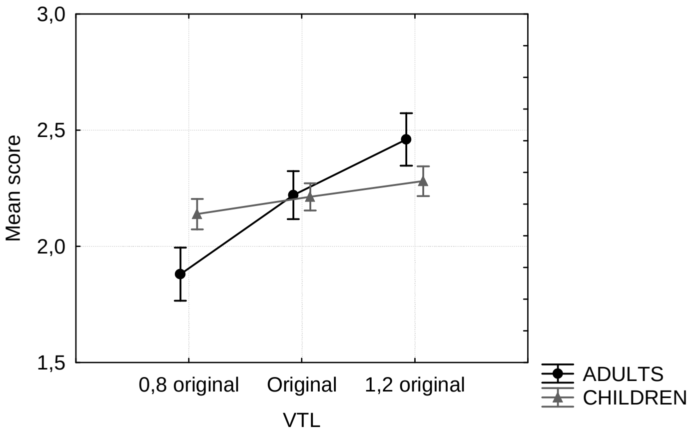

Fig 24 Mean score (±SE) according to different VTL values in children and adults

In the same analysis with AGE GROUP as a between-subject factor, the VTL effect was also significant (F(2, 70) = 8.79, $p < 0.001$), the AGE GROUP effect was not (F(4, 35) = 1.21, $p = 0.32$), and the interaction VTL x AGE GROUP was significant (F(8, 70) = 2.16, $p = 0.04$). According to Bonferroni-corrected *posthoc* tests, the mean score for 0.8 original was lower than the score at 1.2 original, only in the Adult group ($p < 0.001$).

Partial correlation coefficients

In this experiment the partial ρ^2 was calculated for both VTL and F0 in the two conditions.
In the Original F0, varying VTL Condition we calculated the partial $\rho^2_{(VTL)}$ and partial $\rho^2_{(originalF0)}$; in the Original VTL, Varying F0 condition we calculated the partial $\rho^2_{(F0)}$ and partial $\rho^2_{(originalVTL)}$.

Original VTL, varying F0 Condition

The value of partial $\rho^2_{(F0)}$ in this condition did not differ among age groups (Kruskal-Wallis test: H (4, N = 40) = 3.90 $p = 0.42$), while partial $\rho^2_{(originalVTL)}$ did (Kruskal-Wallis test: H (4, N = 40) =12.11 $p = 0.02$). Multiple comparisons revealed that this value for 5 YOs was lower than for 4 YOs. In general, however very little of the variance in perceptual performance was accounted for by F0 or VTL .

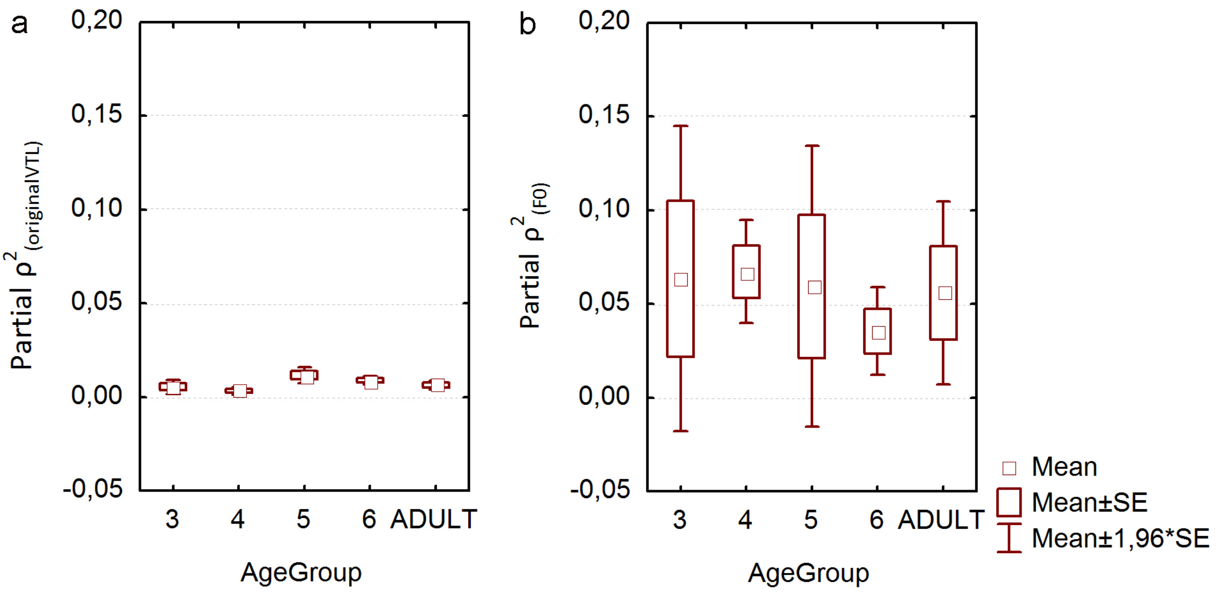

Fig 25 Box plot showing a) partial $\rho^2_{(originalVTL)}$ and b) partial $\rho^2_{(F0)}$ in the different age groups.

For each age group, we compared partial $\rho^2_{(F0)}$ and partial $\rho^2_{(originalVTL)}$ using Wilcoxon Tests. No significant difference was found for 3 YOs ($p = 0.91$), 5 YOs ($p = 0.07$), 6 YOs ($p = 0.05$) and adults ($p = 0.07$), while for 4 YOs partial $\rho^2_{(F0)}$ was significantly higher than partial $\rho^2_{(originalVTL)}$ ($p = 0.008$).

Original F0, varying VTL Condition

The value of partial $\rho^2_{(VTL)}$ in this condition did not differ among age groups (Kruskal-Wallis test: H (4, N = 40) =1.87 p = 0. 76, and neither did $\rho^2_{(originalF0)}$ (Kruskal-Wallis test: H (4, N = 40) =8.03 p = 0.09).

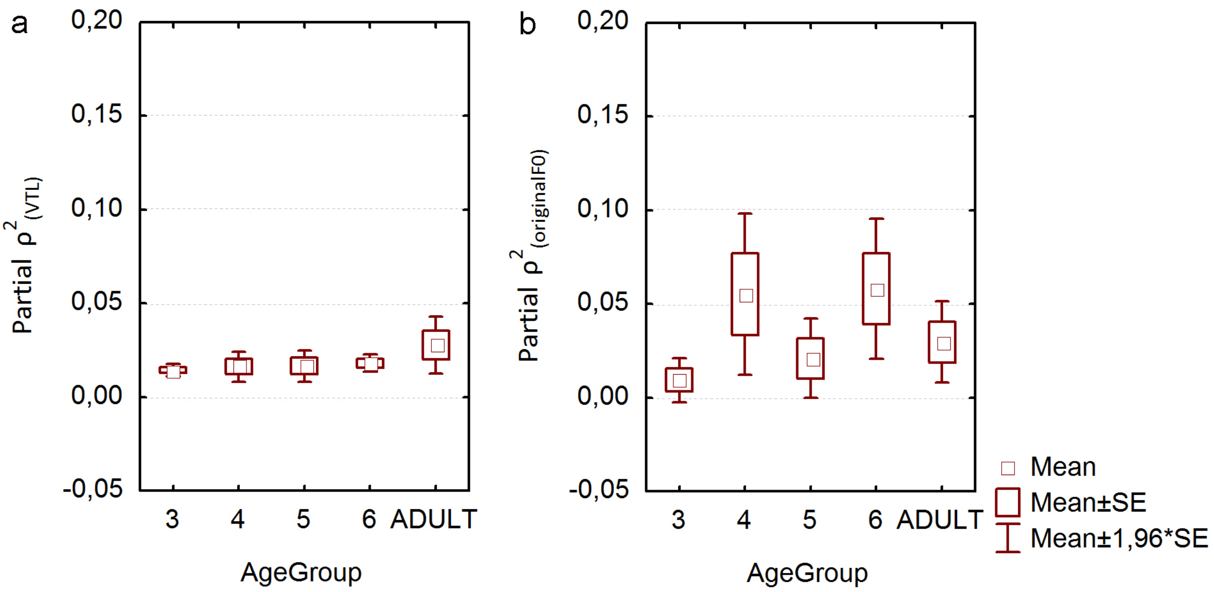

Fig 26 Box plot showing a) partial $\rho^2_{(VTL)}$ and b) partial $\rho^2_{(originalF0)}$ in the different age groups.

For each age group, we compared partial $\rho^2_{(F0)}$ and partial $\rho^2_{(originalVTL)}$ using Wilcoxon Tests. No significant difference was found for 3 YOs ($p = 0.34$), 4 YOs ($p = 0.13$), 5 YOs ($p = 0.34$), 6 YOs ($p = 0.06$) and adults ($p = 0.95$).

In both the conditions, very little of the variance in perceptual performance was accounted for by F0 or VTL values.

5.4 DISCUSSION

This experiment replicates in part the results of Taylor and colleagues (2008): human adults can successfully infer dog size from the formant dispersion of their perceived growls, but not from F0. In many of the former studies the relation between F0 and body size is weak or non-existent (McComb 1991; Rendall et al. 2005), hence this acoustic cue does not appear to be highly salient for size judgments.

Studies on different mammalian species (Smith et al. 2005; Fitch and Fritz 2006; Reby and McComb 2007) have shown that body size perception is influenced by formant dispersion: more spaced formants are associated with a bigger body size and *vice versa*. However, in all these studies subjects were tested using conspecific vocalizations. The study by Taylor et. al (2008) and the present results show that adult human body size judgment is influenced by changes in formant dispersion for both *intra-specific* and *inter-specific sound sources.* However, changes in growl formant dispersion did not as reliably change children's size perception of dogs, even if the trend was increasing for both adults and children. In fact, our results showed that children were less sensitive than adults to variation in perceived VTL, and gave a higher score than adults to the "small" sized stimuli, and a score lower than adults did to the "big" size. It appears that either the ability to infer another species' body size from acoustic cues appears later during life, or that children's ability to make judgments from stimuli with conflicting cues is not as well-developed as it is in adulthood.

When asked to judge dog body size using natural dog sounds (Chapter 3), children as young as 2 years performed well above chance, even if they performed with an accuracy lower than for natural speech items. In that experiment, formants *and* F0 information were not in conflict. When given conflicting human acoustic parameters, children's performance was also very different from adults, depending on their age group (Chapter 4). Three-year-old children showed little difference in the mean score between the two extremes of perceived VTL, while performance converged with that of adults as children aged. Moreover, children of all age groups seemed not to rely on F0 for size judgments, while adults did.

In the case of dog growls, although the acoustic cues were similar to the ones used in human speech, the amount of regular exposure to this type of sound was low in this study population (only 5 children out of 31 were dog owners), and this may have had an effect on performance. While ρ^2 for both the acoustic parameters was at least 20% with speech sounds (Chapter 4), both VTL and F0 explained less than 10% of the size attribution of dog sounds in all the age groups.

In summary, with this experiment we showed that in adult normal hearing subjects size perception was influenced by formant dispersion, but we did not find any hint that F0 was a cue for body size. Children, on the other hand, did not change their perceived size according to modified acoustic parameters, suggesting that the ability to infer body size from acoustic cues of sources other than human speech may appear later during development.

6 PERCEPTION OF SIZE-RELATED FEATURES IN COCHLEAR-IMPLANTED AND NORMAL HEARING CHILDREN

6.1 INTRODUCTION

A cochlear implant is a small neuro-prosthesis that helps to provide a sense of sound or restore hearing in profoundly deaf people, bypassing damaged portions of the ear and directly stimulating the auditory nerve. The implant, with an external portion that sits behind the ear and a second portion surgically placed under the skin, is made of four parts: a microphone, which picks up sound from the environment; a speech processor, which selects and arranges sounds picked up by the microphone; a transmitter and receiver/stimulator, which receives signals from the speech processor and converts them into electric impulses, and finally an electrode array, which is run through the temporal bone and inserted into the scala tympani of the cochlea, which collects the impulses from the stimulator and sends them to different regions of the auditory nerve. Signals generated by the implant are sent by way of the auditory nerve to the brain, which recognizes the signals as sound (Fig 27).

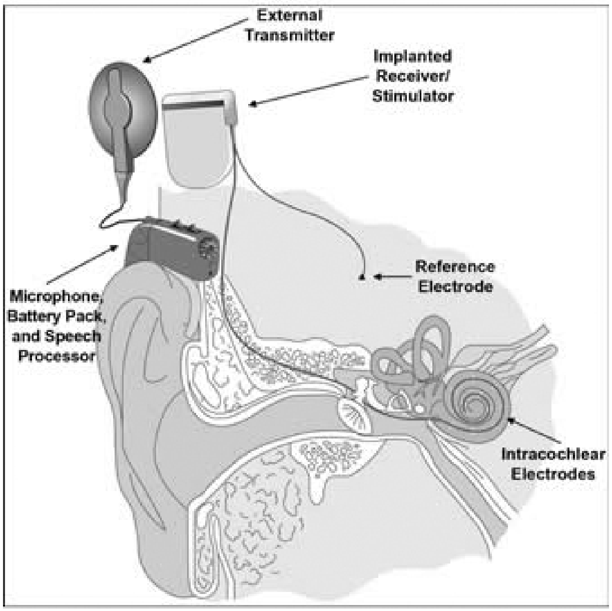

Fig 27 Cochlear implant diagram (from Wilson and Dorman 2008).

The earliest cochlear implants had only one single electrode for transmitting acoustic information to the auditory system, and were not effective in transmitting spectrotemporal information of the perceived sound, required for instance for spoken communication. In the last 20 years different models with multiple channels have been developed. These kinds of devices are capable of transmitting a considerably greater amount of spectral and temporal information on speech and other environmental sounds than single channel devices. Temporal cues are delivered as envelope modulations to one or more stimulation electrodes and depend on the stimulation rate of each electrode, while spectral cues depend on the spatial pattern of electrode array stimulation (McKay 1994, 1995; Laneau and Wouters 2004; Nie et al. 2006). Modern implants have a maximum of 22 electrodes which deliver electrical stimulation along the tonotopic array. Electrodes usually extend for not more than 2 turns into the cochlea, even if recently some attempts have been made to create longer arrays which reach the apex of the cochlea (Rosen et al. 1999; Clark 2004; Boyd 2011).

Natural speech and other sounds carry a great deal of spectral and temporal information to normal-hearing listeners. Considering the limited number of electrodes and the limited dynamic range of effective stimulation, both spectral and temporal cues are encoded poorly by

CI and limit everyday tasks such as understanding speech in noise, appreciating music, or the ability to localize sound sources accurately (Rosen 1992; Loizou 1998; Nie et al. 2006).

Hence, an implant does not restore normal hearing, but it can give a useful representation of sounds in the environment and help a deaf person to understand speech (even by telephone), recognize warning signals, and understand environmental sounds, ultimately improving the quality of life for the severely hearing-impaired. After surgery, extensive post-implantation therapy is required.

It has been shown that better results in speech comprehension and production have been achieved if implant surgery occurs very early during life. Children implanted before 2 years of age who had good language training can achieve age-appropriate speech perception and speech production (Holt and Kirk 2005; Nicholas and Geers 2007).

Several studies show that cochlear implanted subjects show a high intra-individual variability in different tasks such as speech perception (Blamey et al. 2001), as well as with more simple tasks such as gender identification (Kovačić and Balaban 2009, 2010). In these studies, subjects performed better when they proficiently used temporal information, while it was not clear that place cues were actually used to perform this task.

Previous studies have shown that size perception was influenced both by temporal (F0) and spectral information (i.e. formant dispersion), but relied more on the latter. In this experiment we tested a speaker identification task using a group of CI children who followed speech therapy at the Polyclinic SUVAG in Zagreb (Croatia). The aim of this experiment was to compare CI children with NH subjects in 1- the ability to identify the size of speakers from natural and synthetically modified voices, and 2- to see whether they used F0, formant information or both to infer a speaker's size.

In collaboration with Damir Kovačić, PhD, and speech therapists, we tested children using both naturalistic and synthetically-modified voices from 3 different speakers. The modified voices changed either VTL or F0, keeping the other acoustic parameter constant.

We also tested a group of age-matched NH children and a group of NH adults as controls.

6.2 MATERIALS AND METHODS

6.2.1 ACOUSTIC STIMULI

1 short sentence “Odvori vrata” (“Open the door”) pronounced by a man, a woman, a child, with a Marantz PMD671 digital recorder and high-quality microphones, stored as WAV files (16-bit; sampling rate 44.1 KHz).

This sentence was scaled using the STRAIGHT program (Kawahara et al. 1999) to alter either F0 (condition A) or VTL (condition B). STRAIGHT segregates the original sentence into two independent frames: one representing the glottal pulse rate GPR (F0), the other representing spectral envelope ratio SER (vocal tract length information). Each of these can be scaled by a desired factor independently from the other parameter, and the two frames are then recombined together and a new high-quality stimulus is obtained (Smith et al. 2005).

- Original VTL, varying F0 condition: F0 values of 40.80.160.320.640 Hz + 3 original speaker F0s (115, 209, 301 Hz)
- Original F0, varying VTL Condition: VTL values of 6.52, 8.81, 11.91, 15.97, 21.69 cm + 3 original speaker VTLs[2] (14.74, 13.92, 10.66 cm)

A total of 48 stimuli were presented to each subject in random order.

---

[2] calculated from speaker’s height according to Fitch and Giedd 1999

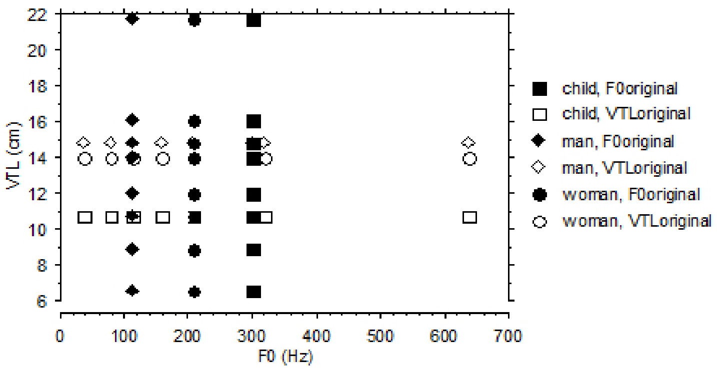

Fig 28 Values of VTL and F0 according to the original speaker voice and the condition

6.2.2 VISUAL STIMULI

Visual stimuli were pictures of one adult male, one adult female and one child chosen randomly from pictures of 5 different women, 5 different men and 5 different children downloaded from the web and already used in the experiments with natural sounds described in previous chapters. A picture for each category was always presented in the same position on the touch screen of a computer monitor: the child on the left, the woman in the centre and the man on the right. Pictures appeared simultaneously along with each acoustic stimulus

6.2.3 EXPERIMENTAL PROCEDURE

Training Phase: In the first *training phase* (with feedback and verbal reward), 3 pictures of different categories appeared on the touch screen and a sound (e.g. a bird, a musical instrument –drum- and a car) was presented at the same time. The child was required touch the correct sound source. Six different sounds were played back (2 for each sound source). Children had to complete a training session with five out of six correct trials before they were tested.

Procedure: a 3 AFC paradigm was used to test children with 48 items. Stimuli in the two conditions were presented in random order, and subjects attributed a source to them by touching the appropriate choice from among the 3 pictures simultaneously presented on the touch screen (Elo 1526L 15” Medical Touchscreen Monitor, Resolution 1024 x 768).

6.2.4 DATA ANALYSIS

A score was assigned to each chosen speaker: 1 to the child, 2 to the woman, 3 to the man. Score variations according to changes in VTL or F0 were tested with Friedman Anova, followed by post-hoc tests corrected for multiple comparisons (Siegel and Castellan 1988) with $p < 0.05$ (two-tailed). Different analyses were run to test how variation in F0 or perceived VTL changed size perception.

6 out of the 48 items were the original speaker’s sentences (hereafter “Original speech items”). It was only possible to calculate accuracy for this set of items. Accuracy scores ranged from 0 to 6 in total (Total Accuracy). Since there were two items from each speaker, the accuracy for each type of speaker (Accuracy_original speaker) ranged from 0 to 2.

6.2.5 PARTIAL CORRELATION COEFFICIENTS

In this experiment the partial ρ^2 was calculated for both VTL and F0 in the two conditions.
In the Original F0, varying VTL Condition we calculated the partial $\rho^2_{(VTL)}$ and partial $\rho^2_{(originalF0)}$; in the Original VTL, Varying F0 condition we calculated the partial $\rho^2_{(F0)}$ and partial $\rho^2_{(originalVTL)}$.

6.2.6 SUBJECTS

Subjects were: 10 normal-hearing adults; 71 normal-hearing children: 7 3YO (40.0 ± 2.0 months), 8 4YO (47.63 ± 3.38 months), 26 5YO (60.12 ± 3.86 months), 23 6YO (72.26 ± 3.91

months), 7 7YO (82.14 ± 1.57 months); and 18 Cochlear Implanted children from SUVAG Polyclinic (both from SUVAG kindergarten or coming weekly to the Polyclinic for individual speech therapy) whose features are summarized in the table below.

Subj n.	***Sex***	***Age at Implantation (months)***	***Age at Testing (months)***	***Implant Use(months)***	***Group***	***Age Group***
38	M	14	35	21	CI_polyclinic	3
41	F	17	37	20	CI_polyclinc	3
34	M	20	45	25	CI_polyclinic	4
35	M	15	53	38	CI_polyclinc	4
36	F	23	49	26	CI_polyclinic	4
44	F	15	47	32	CI_polyclinic	4
47	F	14	48	34	CI_polyclinic	4
28	F	46	72	26	CI_kindergarden	6
29	M	56	70	14	CI_kindergarden	6
42	M	19	75	56	CI_polyclinic	6
45	M	43	67	24	CI_kindergarden	6
57	M	42	67	25	CI_kindergarden	6
30	M	44	80	36	CI_kindergarden	7
31	M	59	84	25	CI_kindergarden	7
32	F	48	83	35	CI_kindergarden	7
43	M	22	83	61	CI_polyclinic	7
46	M	44	93	49	CI_polyclinic	7
89	F	21	92	71	CI_polyclinic	8

Table 11. Information for each CI subject.

6.3 RESULTS

6.3.1 *ACCURACY ON ORIGINAL SPEECH ITEMS*

6.3.1.1 Analyses of the 3 groups of subjects

Subject accuracy on original speech items was calculated for all 3 groups (Fig 29, ChildNH, AdultNH, CI).

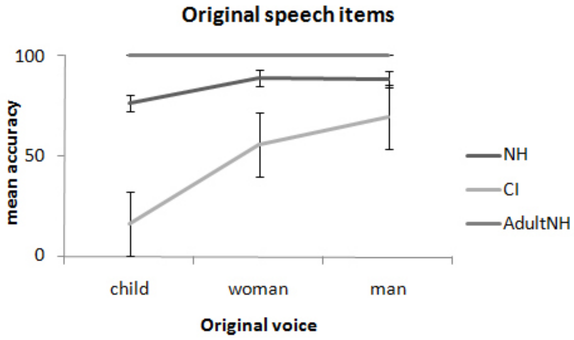

Fig 29 Mean accuracy (±SE) on original speaker's sentences in adult normal hearing (AdultNH; N = 10), normal hearing children (NH; N = 71), cochlear implanted children (CI; N = 18).

A Kruskal-Wallis test revealed significant differences in Total Accuracy among the groups of subjects (H(2, N = 99) = 34.65, $p < 0.001$). Post hoc multiple comparisons showed significant differences between the CI group and NH ($p < 0.001$), CI and AdultNH ($p < 0.001$) and a significant difference between NH and AdultNH ($p = 0.05$).

Looking in more detail, the same test was run for each original speaker:

Adult male voice: The Kruskal-Wallis test was significant (H(2, N = 99) = 8.02, $p < 0.01$), but *post-hoc* multiple comparisons did not show any significant difference between groups (CI *vs* NH $p = 0.34$; CI *vs* AdultNH $p = 0.12$; NH *vs* AdultNH $p = 0.71$).

Adult female voice: The Kruskal-Wallis test was significant (H(2, N = 99) = 13.70. p = 0.001). Multiple comparisons revealed significant differences between CI and NH (p = 0.05) and CI and AdultNH (p = 0.05), but not between NH and AdultNH (p = 1.00).

Child voice: The Kruskal-Wallis test was significant (H(2, N = 99) = 34.99, p < 0.001). As with the adult female voice, multiple comparisons were significant for the CI group and NH (p < 0.001), CI and AdultNH (p < 0.001) but not for NH and AdultNH (p = 0.28).

Friedman ANOVAs with post-hoc tests corrected for multiple comparisons (Siegel and Castellan 1988) were also run for each group of subjects to test differences in accuracy among the 3 types of voices (adult male – adult female – child). Since the adult accuracy was 100% for each type of voice, this test wasn't run for this group.

NH children: The Friedman ANOVA was significant (χ^2 (N = 71, df = 2) = 8.75 p = 0.013). Post-hoc tests did not reveal any significant differences in accuracy between different voices, even if accuracy on the child voice was lower than on the adult male and the adult female voices.

CI: The Friedman ANOVA was significant (χ^2 (N = 18, df = 2) = 11.92 p = 0.003). Post-hoc tests were significant comparing accuracy on the child *vs* the adult male voice (p < 0.05).

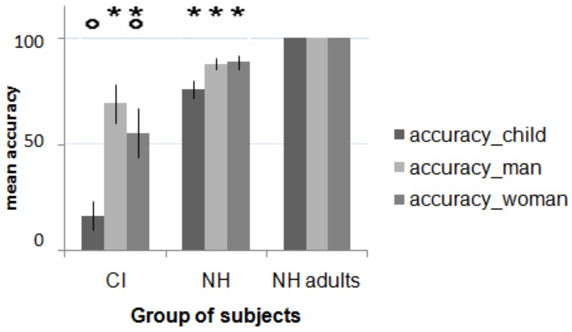

Fig 30 Mean accuracy (±SE) on original speech items in the three groups of subjects. Bars with different symbols above them are significantly different from each other (always p < 0.05), according to *post-hoc* tests corrected for multiple comparisons.

In general, the performance of CI children was lower on the high-pitched child's voice, while it improved with voices of lower pitch. Suvag speech therapists reported that some of the tested CI children had residual hearing in the low-frequency regions, while before the implant they were completely deaf to higher frequencies. This suggests the possibility that correct size attribution from acoustic cues may rely on auditory experience in particular frequency ranges. The better performance in judging low-pitched adult male voices may be due to a longer period of exposure to sounds in low frequency bands.

6.3.1.2 Analyses of normal hearing children

The same tests were run only on NH children to compare the performance of different age groups.

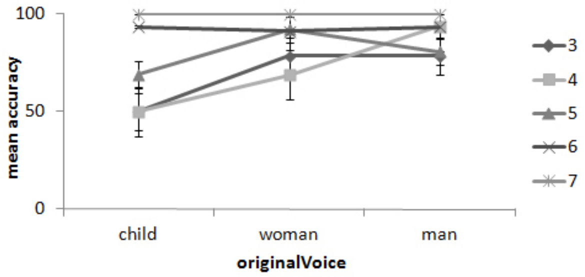

Fig 31 Mean accuracy (±SE) on original speech items in each age group of Normal Hearing: 3 YO N =7; 4 YO N =8; 5 YO N =26; 6 YO N =23; 7 YO N = 7. The accuracy in judging child voices is lower in 3-4-5 YO groups.

6.3.1.2.1 Differences among groups of subjects

There were significant differences in Total Accuracy among age groups (Kruskal -Wallis H(4, N = 71) = 26.22, p < 0.001). The post-hoc multiple comparisons were significant when comparing the two youngest to the two oldest ages (3 vs 6 p = 0.02; 3 vs 7 p = 0.01; 4 vs 6 p =

0.03; 4 vs 7 p = 0.02), while all other comparisons were not (3 vs 4 p = 1.00; 3 vs 5 p = 1.00; 4 vs 5 p = 1.00; 5 vs 6 p = 0.06; 5 vs 7 p = 0.05; 6 vs 7 p = 1.00).

Looking more in detail at the children's accuracy on each original voice, Scheirer-Ray-Hare 2-way non-parametric ANOVA (Sokal and Rohlf 1995) with Age (5 levels) and Original voice (3 levels) as factors revealed that both factors, as well as their interaction, significantly affected performance (Age: H = 10.34, df = 1, p = 0.001; Original voice H = 62.05, df = 2, p < 0.001; Age x Original voice: H = 155.12, df = 2, p < 0.001), showing that the original voices differently affected children's accuracy according to their age. In fact, differences among age groups were not significant for accuracy on the adult male voice (Kruskal-Wallis H(4, N = 71) = 8.54, p = 0.07) or on the adult female voice (Kruskal-Wallis H(4,n = 71) = 7.07, p = 0.13), but were significant for the child voice (Kruskal-Wallis H(4, N = 71) = 19.38, p = 0.0007). However, none of the individual post-hoc multiple comparisons reached significance (3 vs 4 p = 1.00; 3 vs 5 p = 1.00; 3 vs 6 p = 0.26; 3 vs 7 p = 0.36; 4 vs 5 p = 1.00; 4 vs 6 p = 0.06; 4 vs 7 p = 0.13; 5 vs 6 p = 0.20; 5 vs 7 p = 0.54; 6 vs 7 p = 1.00).

6.3.1.2.2 Differences in accuracy among types of voices

Friedman ANOVAs were run for each age group to test differences in accuracy among the 3 types of voices (adult male - adult female - child). Pairwise post-hoc analyses were run using Bonferroni-corrected Wilcoxon tests.

Since the accuracy with 7 YO was 100% for each type of voice, it was not possible to run these tests for this group.

Accuracy was not different among voices in any of the following age groups (3YO: (χ^2 (N = 7, df = 2) = 1.73 p = 0.42; 4 YO (χ^2 (N = 8, df = 2) = 5.33 p = 0.07; 6YO (χ^2 (N = 23, df = 2) = 0.33 p = 0.84). For 5 YOs the test was significant: (χ^2 (N = 26, df = 2) = 8.65 p = 0.01), but none of the post hoc multiple comparisons were significant.

6.3.1.3 Accuracy of CI children

In CI children (N = 18), the correlation between accuracy on original speech items and the attributes of the subjects listed in Table 11 were examined using Spearman Rank correlations. Total Accuracy, Accuracy_adult male; Accuracy_adult female and Accuracy_child were correlated with subject age at testing (expressed in months), Implant Use (i.e. how long they have been implanted – in months), and Age at surgery (i.e. how old they were when they underwent CI surgery).

Nominally-significant positive correlations (uncorrected for multiple comparisons) between accuracy with adult male voices and age at testing ($\rho = 0.45$, $p = 0.04$) and accuracy with adult male voices and Implant Use ($\rho = 0.53$, $p = 0.03$) were found, but these correlations would not survive a correction for multiple comparisons. No nominally-significant correlations were found for any other variables. However, NH children exhibited nominally-significant positive correlations between age and Total Accuracy ($\rho = 0.58$, $p < 0.001$), and between age and accuracy on child voice ($\rho = 0.49$, $p < 0.001$) that survive corrections. There was also a nominally-significant correlation between age and accuracy on the woman's voice ($\rho = 0.28$, $p = 0.02$) that would not survive correction; the correlation with accuracy on the man's voice was not nominally significant ($\rho = 0.23$, $p = 0.06$).

If, as already discussed on page 97, some CI subjects had residual hearing in low-frequency regions, these results would further suggest that exposure to acoustic information from particular frequency bands may be important for the development of size attribution for sounds.

6.3.2 *CHANGES IN SIZE PERCEPTION WITH CHANGES IN ACOUSTIC PARAMETERS*

Original F0, varying VTL Condition

In all the groups of subjects, perceived size changed significantly as a function of VTL variation. CI (χ^2 (N = 18, df = 7) = 21.81 p = 0.003); NH (χ^2 (N = 71, df = 7) = 281.11 p < 0.001); AdultNH: (χ^2 (N = 10, df = 7) = 59.71 p < 0.001) (Fig 32). The mean score increased with increasing simulated VTL in all the three groups of subjects; however, while the patterns of NH and AdultNH are similar and the mean score increased *gradually* with increasing VTL, the effect of changing VTL on size perception in CI users was radically different: there was a complex pattern of change that produced a main statistical effect of VTL, but post-hoc multiple comparisons did not reveal any significant difference among the different VTL values.

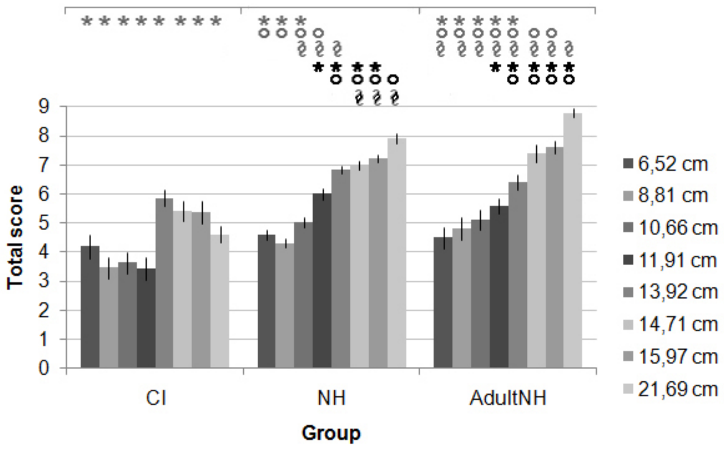

Fig 32 Total score (±SE) according to different perceived VTLs of the modified voices in the CI, NH and AdultNH group. Bars with different symbols on the top are significantly different from each other (always p < 0.05), according to *post-hoc* multiple comparisons.

Given these results, we ran another series of analyses,, considering separately the mean score on the first four, “small” VTLs (from 6.52 cm to 11.91 cm) and the second four “big” VTLs (from 13.92 cm to 21.69 cm). This breaks the stimulus continuum into two groups. “small” VTL and “big” VTL.

Mann-Whitney U tests for the three categories of subjects (Bonferroni corrected for 3 comparisons) showed that the mean score of “big” VTLs was significantly higher than for “small” VTLs: CI (U = 85.00, p = 0.01), AdultNH (U = 0.00, p < 0.001); NH (U = 293.5, p < 0.001).

If CI users have two categories (“small” and “big” VTLs) while normal hearing listeners have many, then Friedman ANOVAS should be significant for AdultNH and NH, but not for CI within these “small” and “big” groups. With the small VTLs the results were: AdultNH (χ^2 (N = 10, df = 3) = 9.20 p = 0.03); NH (χ^2 (N = 71, df = 3) = 54.94 p < 0.001); CI (χ^2 (N = 18, df = 3) = 2.13 p = 0.54).

With the big VTLs results were AdultNH (χ^2 (N = 10, df = 3) = 23.81 p < 0.001); NH (χ^2 (N = 71, df = 3) = 49.91 p < 0.001); CI (χ^2 (N = 18, df = 3) = 3.56 p = 0.31. These results confirm that, unlike AdultNH and NH, CI divide these stimuli into two categories: “small” and “big” VTLs.

6.3.2.1.1 Partial correlation coefficients

Partial $\rho^2_{(VTL)}$ differed among the three groups of subjects (Kruskal-Wallis test: H (2, N = 99) = 19.55, p < 0.001); this value was lower for CI than for other two groups (p < 0.001), indicating that perceptual performance was little influenced (less than 10%) by VTL information in this group, while NHchildren and adultNH did not differ from each other.

Partial $\rho^2_{(originalF0)}$ was also different among groups (Kruskal-Wallis test: H (2, N = 99) = 19.95, p < 0.001), but in this case CI and NHchildren did not differ from each other, while this quantity for adultNH was higher than both child groups (both ps < 0.001) (Fig 33).

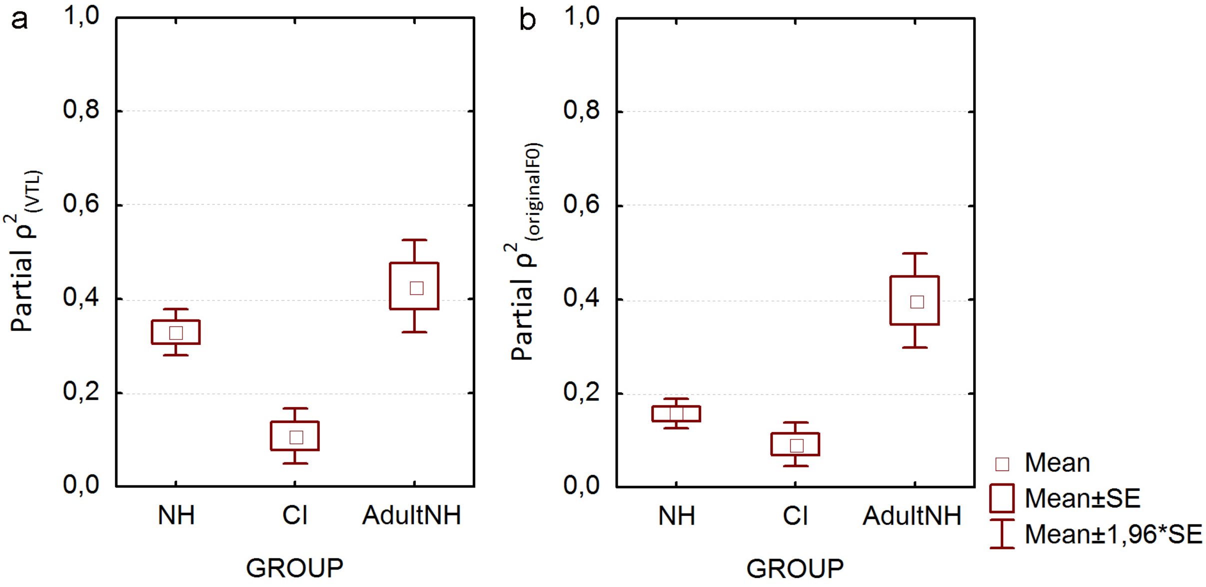

Fig 33 Box plot showing a) partial $\rho^2_{(VTL)}$ and b) partial $\rho^2_{(originalF0)}$ in the groups of subjects.

Partial $\rho^2_{(VTL)}$ was not different from partial $\rho^2_{(originalF0)}$ in CI (Wilcoxon test: $p = 0.98$) nor in AdultNH ($p = 0.80$), while in NH children partial $\rho^2_{(VTL)}$ was higher than partial $\rho^2_{(originalF0)}$ ($p < 0.001$), showing that VTL influenced perceptual performance more than F0 in this group.

6.3.2.1.2 NH children by age

An analysis of age variation in NH children revealed that, in each age group, perceived size changed significantly as a function of VTL variation: 3YO: χ^2 (N = 7, df = 7) = 19.43, p = 0.007; 4 YO: χ^2 (N = 8, df = 7) = 29.59, p < 0.001; 5 YO: χ^2 (N = 26, df = 7) = 91.62, p < 0.001; 6YO: χ^2 (N = 23, df = 7) = 111.42, p < 0.001; 7YO: χ^2 (N = 7, df = 7) = 39.96, p < 0.001 (Fig 34).

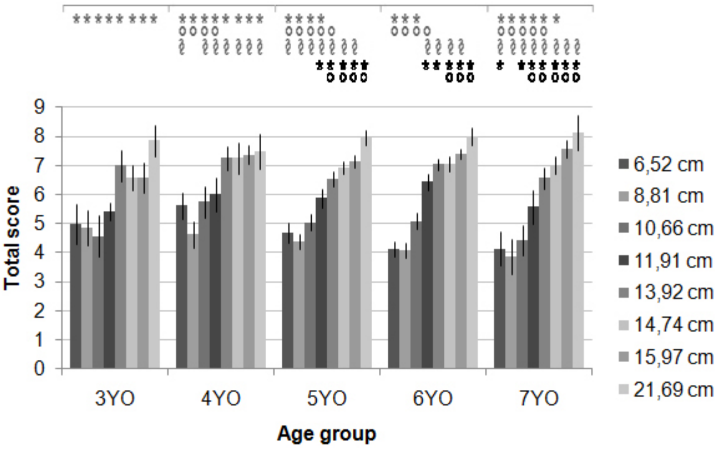

Fig 34 Total score (±SE) according to different perceived VTLs of the modified voices in the different age groups of NH chidlren. Bars with different symbols on the top are significantly different from each other (always p < 0.05), according to *post-hoc* multiple comparisons.

Mann-Whitney U tests for the different age groups (Bonferroni corrected for 5 comparisons) showed that the mean score of “big” VTLs (largest 4 values) was significantly higher than for “small” VTLs (smallest 4 values): 3 YO (N = 7, U = 4.0, p = 0.009), 4 YO (N = 8, U = 7.00, p = 0.009); 5 (N = 26, U = 46.5, p < 0.001), 6 YO (N = 23, U = 20.5, p < 0.001), 7 YO (N = 7, U = 1.00, p = 0.003).

Looking more in detail at the behavior of each age group within each class, we found that with the small VTLs the results were: 3 YO (χ^2 (N = 7, df = 3) = 2.46 p = 0.48); 4 YO (χ^2 (N = 8, df = 3) = 5.31 p = 0.15); 5 YO (χ^2 (N = 26, df = 3) = 11.71 p = 0.008). 6 YO (χ^2 (N = 23, df = 3) = 41.40 p < 0.001); 7 YO (χ^2 (N = 7, df = 3) = 10.36 p = 0.02).

With the big VTLs results were: 3 YO (χ^2 (N = 7, df = 3) = 4.32 p = 0.22); 4 YO (χ^2 (N = 8, df = 3) = 2.0 p = 0.57); 5 YO (χ^2 (N = 26, df = 3) = 22.11 p < 0.001). 6 YO (χ^2 (N = 23, df = 3) = 21.69 p < 0.001); 7 YO (χ^2 (N = 7, df = 3) = 8.28 p = 0.04). These results suggest that while 3 and 4-year olds have a two-class, “small vs big” system, from the age of 5 normal-hearing children are able to make finer distinctions when only scale information varies.

6.3.2.1.3 Partial correlation coefficients

Partial $\rho^2_{(VTL)}$ did not differ among different age groups (Kruskal-Wallis test: H (4, N = 71) = 8.83 p = 0.06) while partial $\rho^2_{(originalF0)}$ did (Kruskal-Wallis test: H (4, N = 71) =10.56 p = 0. 03), but none of the posthoc multiple comparisons were significant (Fig 35).

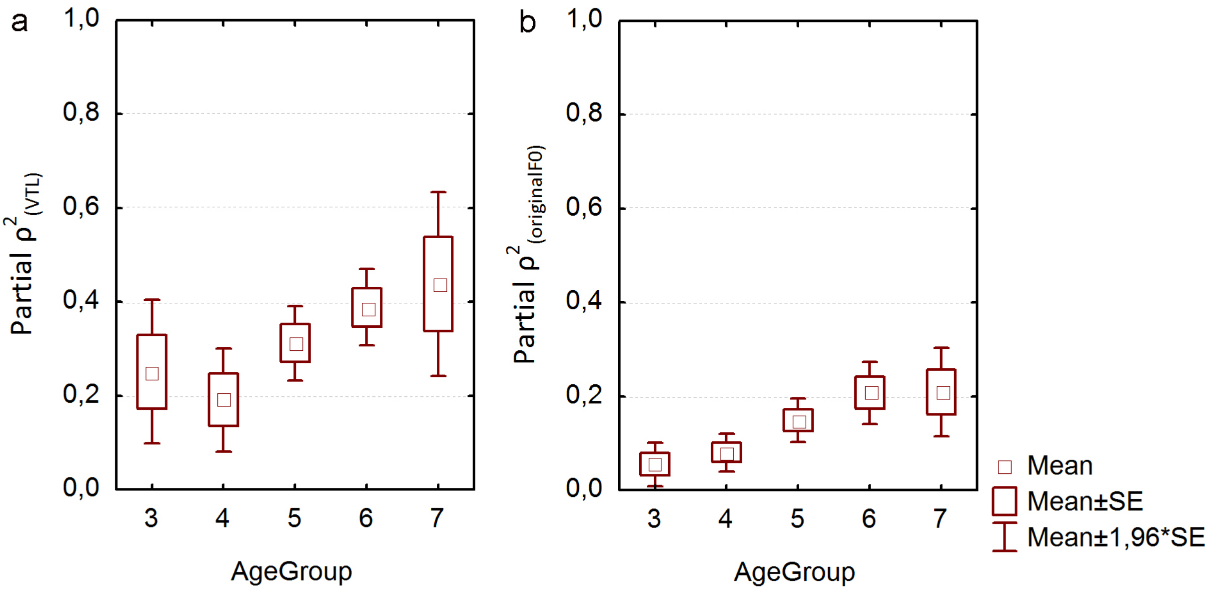

Fig 35 Box plot showing a) partial $\rho^2_{(VTL)}$ and b) partial $\rho^2_{(originalF0)}$ in the different children age groups.

Wilcoxon tests revealed that partial $\rho^2_{(VTL)}$ was significantly higher than partial $\rho^2_{(originalF0)}$ for 3 YOs (p = 0.04), 4 YOs (p = 0.05), 5 YOs (p < 0.001), 6 YOs (p = 0.003) but not for 7 YOs (p = 0.13).

In the younger age groups, VTL influenced size perception more than F0. It is important to note that this difference was not significant for adults, indicating that VTL was a more salient cue than F0 for size perception in younger children, but as they grew its relative importance decreased with age; in adults the two factors were equally important for this task.

Original VTL, varying F0 Condition

In all the groups of subjects, perceived size changed significantly as a function of F0 variation. CI (χ^2 (N = 18, df = 7) = 23.02, p = 0.002) AdultNH: (χ^2 (N = 10, df = 7) = 52.43, $p < 0.001$); NH (χ^2 (N = 71, df = 7) = 64.80, $p < 0.001$). The mean score decreased with increasing F0 (Fig 36). As with VTL, CI children had a pattern that produced a significant main effect, but contained no significant differences among individual stimuli.

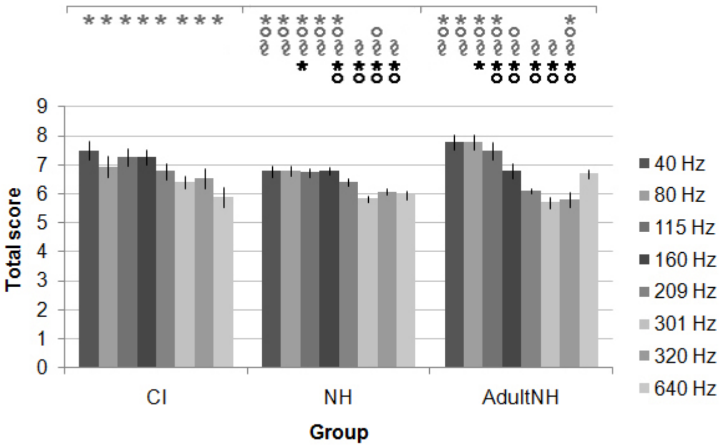

Fig 36 Total score (±SE) according to different F0 of the modified voices in the CI, NH and AdultNH group. Bars with different symbols on the top are significantly different from each other (always $p < 0.05$), according to *post-hoc* multiple comparisons.

In order to look more in detail at the pattern described in Fig 36, we analyzed the mean score on the first four F0s (“low”, from 40 Hz to 160 Hz) and the second four F0s (“high”, from 209 Hz to 640 Hz).

Mann-Whitney U tests for the three categories of subjects (Bonferroni corrected for 3 comparisons) showed that the mean score of “low F0s” is significantly higher than for “high F0s”: CI (U = 74.50 p = 0.005), AdultNH (U = 2.0, $p < 0.001$); NH (U = 1446.5 $p < 0.001$).

The results within the first four F0s were: AdultNH (χ^2 (N = 10, df = 3) = 15.39 p = 0.002); NH (χ^2 (N = 71, df = 3) = 1.48 p = 0.69); CI (χ^2 (N = 18, df = 3) = 2.40 p = 0.49).

Within the second four F0s results were: AdultNH (χ^2 (N = 10, df = 3) = 16.90 $p < 0.001$); NH (χ^2 (N = 71, df = 3) = 12.19 p = 0.007); CI (χ^2 (N = 18, df = 3) = 4.66 p = 0.19.

Even if the size perception changed as a function of F0 in all the groups of subjects, these analyses show that, just as with VTL cues, CI divide these stimuli into two “broad” categories: low F0 and high F0. NH children do not reliably change their size perception for F0s in the low frequency range, but they do with the high F0 values, while normally-hearing adults are able to make finer discriminations among stimuli across the full range of F0s.

6.3.2.1.4 Partial correlation coefficients

In this condition, partial $\rho^2_{(\mathrm{originalVTL})}$ differed among groups (H (2, N = 99) =26.08 p < 0 001) and this measure was lower for CI than the other two groups (both ps < 0.001). Partial $\rho^2_{(F0)}$ was also different among groups (Kruskal-Wallis test (2, N = 99) =8.77 p = 0.01; and the value for AdultNH was higher than for NHchildren (p = 0.009, Fig 37).

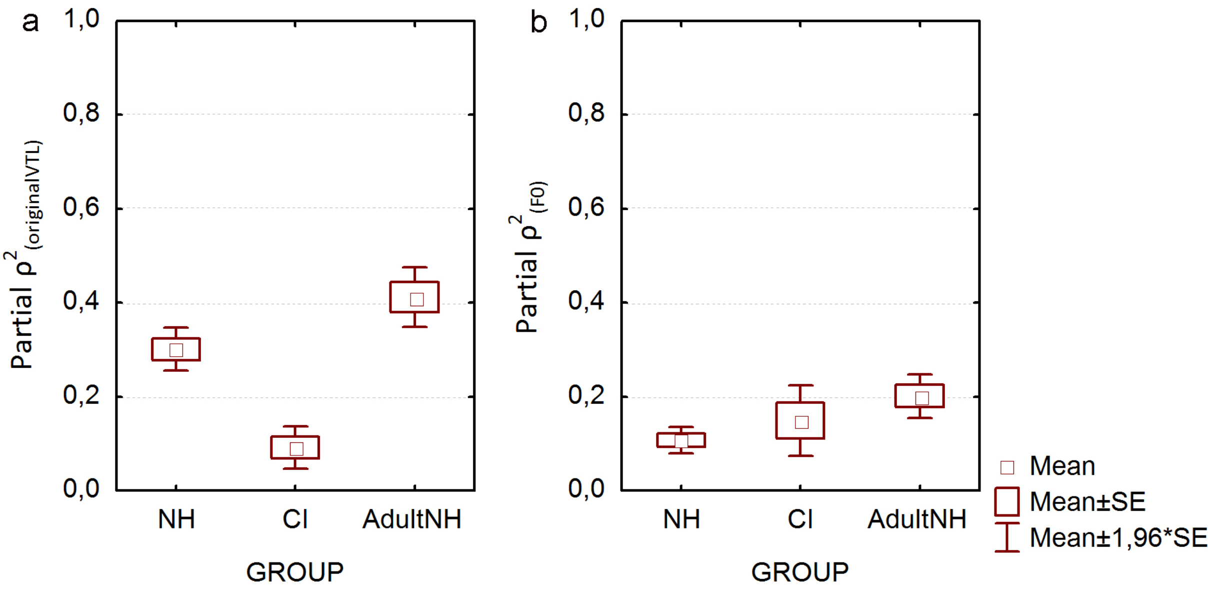

Fig 37 Box plot showing a) partial $\rho^2_{(\mathrm{originalVTL})}$ and b) partial $\rho^2_{(F0)}$ in the different groups of subjects

Partial $\rho^2_{(\mathrm{originalVTL})}$ was higher than partial $\rho^2_{(F0)}$ for AdultNH (Wilcoxon test: p = 0.007) and for NHchildren (p < 0.001), while it was not different for CI subjects (p = 0.28).

6.3.2.1.5 *NH children age by age*

While the mean score did not change significantly with changing F0 in 3-year-olds (χ^2 (N = 7, df = 7) = 7.56, p = 0.37), perceived size changed significantly in all the other age groups: 4YO: χ^2 (N = 8, df = 7) = 20.47, p = 0.005; 5YO: $\chi2$ (N = 26, df = 7) = 28.45, p < 0.001; 6YO: χ^2 (N = 23, df = 7) = 38.03, p < 0.001; 7YO: χ^2 (N = 7, df = 7) = 19.38, p = 0.007 (Fig 38).

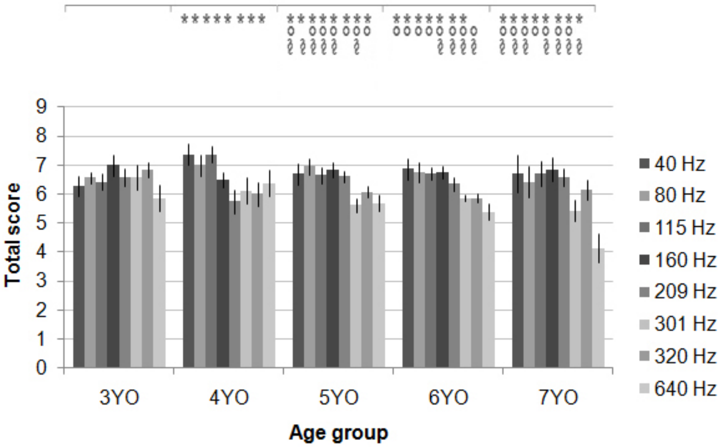

Fig 38 Total score (±SE) according to different F0 of the modified voices in the different age groups of NH children. Bars with different symbols on the top are significantly different from each other (always p < 0.05), according to *post-hoc* multiple comparisons.

Given that 3 YO are not able to change their size perception when only F0 changes, we wanted to look more in detail the behavior of older age groups (4 – 7 YO). Therefore we ran analyses grouping the first four and the second four F0 values (Mann-Whitney U test with Bonferroni-correction: level of significance p = 0.0125). Results were: 4 YO (N = 8; U = 10.0, p = 0.02), 5 YO (N = 26, U = 218.50, p = 0.03), 6 YO (N = 23, U = 138.5, p = 0.006), 7 YO (N = 7, U = 40.0, p = 0.11).

Moreover, we also examined Friedman ANOVAs on the mean score across different F0 values for each of the two classes.

Results for the “low F0” class were: 4 YO: χ^2 (N = 8, df = 3) = 7.47, p = 0.06; 5 YO: χ^2 (N = 26, df = 3) = 1.36, p = 0.71; 6YO: χ^2 (N = 23, df = 3) = 0.75, p = 0.86; 7YO: χ^2 (N = 7, df = 3) = 0.56, p = 0.90.

Results for the “high F0” class were: 4 YO: χ^2 (N = 8, df = 3) = 1.90, p = 0.59; 5 YO: χ^2 (N = 26, df = 3) = 6.95, p = 0.07; 6YO: χ^2 (N = 23, df = 3) = 5.95, p = 0.11; 7YO: χ^2 (N = 7, df = 3) = 13.00, p = 0.004.

From these results it is clear that, even if we found a main effect of F0 change on size perception in 4-year-old children, it did not follow a clearly-interpretable pattern. However, it appears that from the age of 6, children start to categorize F0 with at least two different classes.

6.3.2.1.6 Partial correlation coefficients

In this condition we did not find significant differences among age groups in partial $\rho^2_{(\mathrm{originalVTL})}$. (H (4, N = 71) =6.32 p = 0.18) nor in partial $\rho^2_{(\mathrm{F0})}$ (H (4, N = 71) =7.13 p = 0.13) (Fig 39).

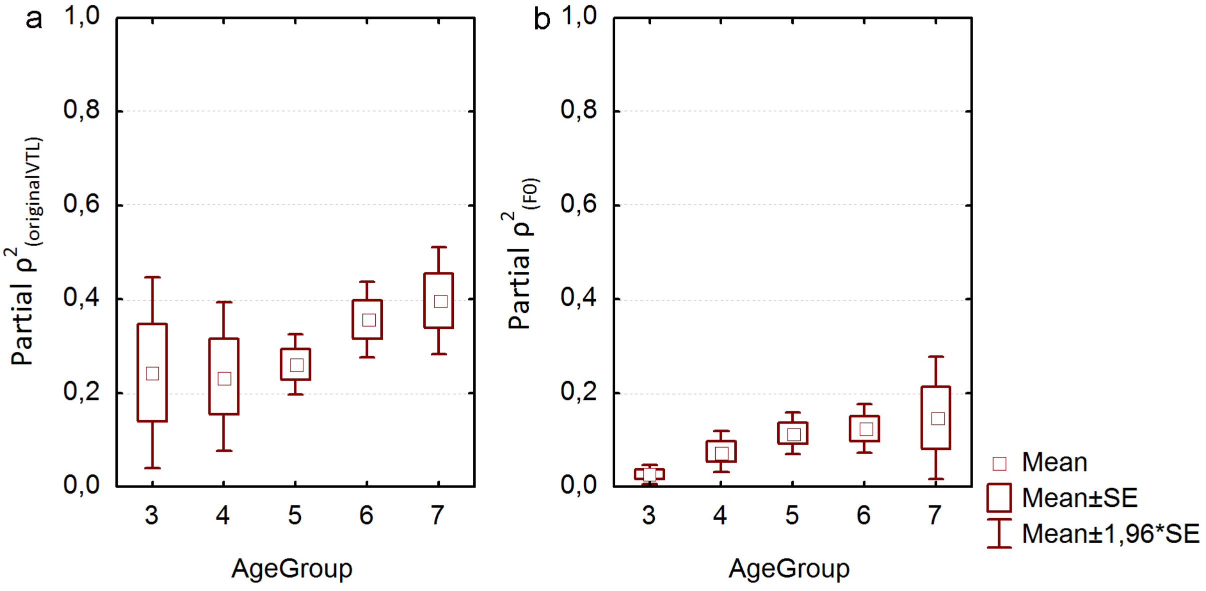

Fig 39 Box plot showing a) partial $\rho^2_{(\mathrm{originalVTL})}$ and b) partial $\rho^2_{(\mathrm{F0})}$ in the different age groups.

Partial $\rho^2_{(\mathrm{originalVTL})}$ was significantly higher than partial $\rho^2_{(\mathrm{F0})}$ for 3YOs (Wilcoxon test: p = 0.04), 5 YOs (p = 0.003), 6 YOs (p < 0.001). It was higher but did not reach significance for 4 YOs (p = 0.16) and 7 YOs (p = 0.06)

6.4 DISCUSSION

Accuracy on Natural Speech items.

In NH children, performance was always above chance, even if for younger children the performance on the child voice was lower than the performance on the adult male and female voices. As already seen in Chapter 3 (on natural sounds), accuracy on natural speech items significantly and positively correlated with age; in particular the correlations between age and accuracy on the child voice were significant and survived Bonferroni correction. The performance of CI subjects was lower than NH subjects on child and woman voices, but not different from the other groups on the low-pitched man's voice. The child voice was the most difficult to judge for this group as well, and the performance was at chance with this type of item.

In CI subjects, we found a relation between performance and age at testing on the adult male voice, as well as a positive correlation with implant use, but these tests did not survive Bonferroni correction. The influence of implant use and the residual low-frequency hearing reported by speech therapists supports the hypothesis that the performance may have been influenced by the differential experience that CI users have had with sounds confined to different frequency bands.

How F0 and VTL variation affected size perception

Perceived size changed significantly as a function of VTL variation. The mean score increased with increasing simulated VTL in all three groups of subjects; however, while the patterns of NH and AdultNH were similar and the mean score increased *gradually* with increasing VTL, the effect of changing VTL on size perception in CI users was radically different and showed that these subjects divided stimuli in two distinct classes: small VTLs and big VTLs. The same pattern was found with varying F0: CI divided stimuli in two classes: low F0s *vs* high F0s.

An analysis of different age groups in normal-hearing children revealed that perceived size changed significantly as a function of VTL variation, but the pattern for 3 and 4 YOs was based on a two category system as found in CI subjects, while older children had a more finely-graded system , more similar to that seen in adults.

Dealing with F0, our results show that 3 and 4 year-old normal hearing children seem unable to categorize stimuli when only this acoustic cue changes, while from the age of 5 they seem to be able to categorize the stimuli at least into two distinct classes. However the results on F0 are not straightforward, and must be interpreted with caution. Nonetheless, CI children seem to have at least a two-category system, comparable to 5-6 year-old normal hearing controls. Moreover, CI subjects were able to correctly identify the gender of older speakers when listening to natural speech. Hence, considering that the average age of CI children was 65.56 (± 4.43) months, their F0 performance appeared comparable to normal-hearing children of their own age, while with VTL information their performance appeared to be developmentally-delayed (similar to 3-year-old normally-hearing children).

NH children appeared to rely more on VTL than F0, and the difference between the contribution of the two acoustic cues lessened with age; on the other hand, NHAdults were equally affected by VTL and F0 in this task. With age the relative contribution of VTL gradually increased from 20% to 40%, reaching the adult value, while the contribution of F0 remained very low (around 20%).

In summary, CI subjects were able to correctly identify speakers when presented with natural voices and non-conflicting cues. However it was not possible to understand whether they proficiently used either VTL or F0 to perform the task, since both these acoustic cues accounted for very little of the variance in perceptual performance, and the difference between partial $\rho^2_{(VTL)}$ and partial $\rho^2_{(F0)}$ was never significant.

On the other hand in NH subjects, both adults and children, VTL seemed to be a generally more salient cue for size perception than F0, which has been considered a less reliable indicator of body size in previous studies (Gonzales 2004; Rendall et al. 2005, 2007). It is interesting that the relative contribution of both cues is high and almost comparable in adults, while in children it is clear that VTL is more important than F0. Moreover, NH children are able to make fine discrimination relying on F0 only in the high frequency range but not in the low frequency range. As stated previously, it is unclear whether this might be due to some kind of immaturity of brain processing mechanisms for F0 perception (Moore 2002), or with the development of experience-dependent strategies for weighting different kinds of perceptual information.

7 FINAL DISCUSSION

In this thesis we tried to focus on two important acoustic cues commonly found in most sounds produced by natural animate and inanimate objects: fundamental frequency and pattern of resonances.

We started by studying these parameters from the *production* point of view (Chapter 2) and showed that in Barbary Macaques, it was not easy to find a straightforward linkage between size information and vocal variation. We found that resonances and fundamental frequency can be a predictor of body size in males but not in females, and that some kind of modulation in sound production occurs according to different behavioral contexts; however, we could not find a clear relationship between resonances or fundamental frequency and vocal behavior at a more detailed level within our subjects.

The second and most important part of the thesis focused on developmental aspects of the perception of these acoustic cues, using a series of psychophysical experiments carried out with pre-school children and adults.

Contrary to the opinion of previous researchers, we showed that size extraction does not appear to be an automatic process (Smith et al. 2005); however, the present research was not designed to provide information on what stage of the auditory pathways these cues are extracted at. In Chapter 3, we further showed that performance is not equivalent with sounds from different source categories (humans, dogs, string instruments) even in adults. This is interesting, because all of these sounds present the same physical manifestations of how their acoustic properties change with their size. Hence, the present experiments suggest a strong experience-dependence that does not appear to generalize between objects, and that gradually improves size extraction accuracy with age.

Whenever natural sounds with congruent spectral and temporal size-related information are presented, children as young as 2 can extract size information reliably above chance, and from the age of 5 their performance is similar to adults (at least with speech stimuli), matching a low-pitched sound with a bigger sound source and *vice versa* (Mondloch and Maurer 2004).

Considering that the two parameters important for size extraction are F0(GPR) and resonances (formants in the speech items), we looked in greater detail in chapters 4, 5 and 6 at the relative importance of these two cues, with sounds that had a similar structure but were from two different sources (speech and dog growls).

According to this, we showed in Chapter 4 and 6 (with modified speech) that when only one of the two cues varies and the two parameters are incongruent, child performance differed notably from adult performance, and that this performance difference changed with age.

In particular, adults were able to rely on F0 for size judgments much more than children, especially when listening to high-pitched voices. It is unclear if this is related to something about the physiological maturation of the auditory system (Moore 2002), or whether this simply depends on experience. Normal hearing children seem to be able to categorize F0 at least into “low *vs* high” categories starting at the age of 5 years, while adults are able to make finer categorizations.

Therefore, size perception in adults seems to be equally affected by both F0 and VTL when these information sources vary one at a time and are incongruent, and their performance is finely-graded. Children appear to process these two acoustic cues in very different ways: the VTL information is processed in a two-category system starting at the age of 3, and is more finely categorized as children get older. It is not as clear how F0 influences size perception: it seems that at least a two-category system is present starting at the age of 5 years, but the results obtained in these experiments were not clear-cut and more work on the development of F0 processing needs be done in the future. In general, normal- hearing children relied more on changes in VTL than on F0 for their size perception, but the relative importance of F0 increased as children aged. Again, these experiments cannot discriminate between explanations based on maturation and those based on experience.

In Chapter 5 we used dog growls, which have a pulse-resonant physical structure similar to human speech. We modified these sounds to have incongruent F0/VTL information. In this case both adults and children relied more on VTL to infer size. In particular, VTL variation affected size only for adults’ percepts while children were less sensitive than adults to variation in perceived VTL. These results replicated those of Taylor et al. (2008). F0 variation did not affect the judgments of either of these groups of subjects.

The study on CI subjects (Chapter 6) showed that they were able to correctly identify speakers when presented with natural voices and non-conflicting cues, even if they did so at chance levels with the highest-pitched stimuli (child voices). Considering the low-frequency residual hearing reported by their speech therapists, this suggests an experience-dependent effect: they did best with voices that provided information within the frequency bands where they had the greatest amount of experience .

It was not possible to clearly understand whether CI subjects proficiently used either VTL or F0 to perform the task, since both these acoustic cues accounted for very little of the variance in perceptual performance. However, variation of both F0 and VTL changed CI size perception: they were able to use F0 to perform the task as well as their normal-hearing peers (as already shown by Kovačić and Balaban 2009, 2010), but they were able to use also VTL, even if in a coarse way by which they identified two categories of objects: big VTLs vs small VTLs. Therefore, CI subjects tested in these experiments appeared to be developmentally–delayed in scale-processing, while their performance in an F0 perception task was not different from normal hearing children of comparable ages.

The experiments run in this thesis leave some open questions, that could be addressed in future studies. First, it seems important to devote further investigation to the experience-dependence hypothesis suggested by the data presented here. In order to do this, a size perception experiment using completely novel stimuli (not necessarily speech) with controlled variation of F0 and resonances should be carried out. In this experiment, adult subjects should be trained (controlling for the length of the training session) and tested before and after training to see how experience with these novel stimuli affects size perceptual performance.

Another point would be to extend the age span of tested children and increase the degree of detail involved in measuring size perception performance. One idea would be to focus on different frequency ranges of F0 (low and high) and different patterns of resonances (distant resonances vs close resonances), and to increase the number of size rating categories to more to than 3 (perhaps using a 7-to-10 point scale). To avoid the confounding effects of gender in the speaker identification task, tone could use just one type of speaker (e.g. only men of different heights). The position of the visual stimuli could also be randomized.

It would also be of interest to test younger children (such as 1-year-old toddlers or infants) with natural and modified speech sounds using natural behaviors such as looking (measured via eye-tracking) to overcome their limited attentional resources. On the other hand, we should also test also older, school-age children (>6 YO) with these experiments to see when they reach adults' performance in their ability to use F0.

In conclusion, this series of psychophysical experiments has opened up a new line of research to study, from the developmental point of view: the experience-dependent process by which size attribution from acoustic cues emerges between infancy and adulthood

8 REFERENCE LIST

Boyd PJ (2011) Potential benefits from deeply inserted cochlear implant electrodes. Ear and Hearing 32(4):411-427.

Blamey PJ, Sarant JZ, Paatsch LE, Barry JG, Bow CP, Wales RJ, Wright M, Psarros C, Rattigan K (2001) Relationships among speech perception, production, language, hearing loss, and age in children with impaired hearing. Journal of Speech Language Hearing Research 44:264-285.

Bradbury JW, Vehrencamp SL (1998) Principles of animal communication. Sinauer:Sunderland.

Bryant GA, Haselton MG (2008) Vocal cues of ovulation in human females. Biology Letters 5:12-15.

Brumm H, Kipper S, Riechelmann C, Todt D (2005) Do Barbary macaques ‘comment’ on what they see? A first report on vocalizations accompanying interactions of third parties. Primates 46:141-144.

Carello C, Anderson KL, Kunkler-Peck AJ (1998) Perception of object length by sound. Psychological Science 9:211-214.

Charlton BD, Reby D, McComb K (2007) Female red deer prefer the roars of larger males. Biology Letters 3:382-385.

Charlton BD, Reby D, McComb K (2008) Effects of combined source (F0) and filter variation (formants) on red deer hind responses to male roars. Journal of the Acoustical Society of America 123: 2936–2943.

Clark G (2004) Cochlear Implants, in Speech processing in the auditory system:422-462.Springer-Verlag, New York.

Cohen L (1993) The scale representation. IEEE Transactions of Signal Processing 41:3275-3292.

Dell'Aira K (1995) Studio dei determinanti del comportamento allo-materno in due colonie di primati non umani. MS Thesis, Biological Sciences, University of Parma.

Deag JM, Crook JH (1971) Social behavior and 'agonistic buffering' in the wild barbary macaque *Macaca sylvanus* L. Folia Primatologica 15(3):183-200.

Ey E, Pfefferle D, Fischer J (2007) Do age- and sex-related variations reliably reflect body size in non-human primate vocalizations? A review. Primates 48(4):253-267.

Fant G (1960) Acoustic theory of speech production. The Hague, Mouton.

Feinberg DR, Jones BC, Little AC, Burt DM, Perrett DI (2005) Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. Animal Behaviour 69:561-568.

Fischer J, Hammerschmidt K (2002) An overview of the Barbary macaque, *Macaca sylvanus*, vocal repertoire. Folia Primatologica 73(1):32-45.

Fischer J (2004) Emergence of individual recognition in young macaques. Animal Behaviour 67:655-661.

Fitch WT, Hauser MD (1995) Vocal production in nonhuman primates: acoustics, physiology, and functional constraints on "honest" advertisement. American Journal of Primatology 37:191-219.

Fitch WT (1997) Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. Journal of the Acoustical Society of America 102:1213-1222.

Fitch WT (2000) The phonetic potential of nonhuman vocal tracts: comparative cineradiographic observations of vocalizing animals. Phonetica 57:205-218.

Fitch WT, Fritz JB (2006) Rhesus macaques spontaneously perceive formants in conspecific vocalizations. Journal of the Acoustical Society of America 120:2132-2141.

Fitch WT, Giedd J (1999) Morphology and development of the human vocal tract: a study using magnetic resonance imaging. Journal of the Acoustical Society of America 106:1511-1522.

Fitch WT, Reby D (2001) The descended larynx is not uniquely human. Proceedings of the Royal Society of London 268:1669-1675.

Fugate JM, Gouzoules H, Nygaard LC (2008) Recognition of rhesus macaque (*Macaca mulatta*) noisy screams: evidence from conspecifics and human listeners. American Journal of Primatology 70(6):594-604.

Gamba M, Giacoma C (2007) Quantitative acoustic analysis of the vocal repertoire of the crowned lemur. Ethology Ecology and Evolution, 19 323-43.

Garcia D, Hall DA, Plack CJ (2010) The effect of stimulus context on pitch representations in the human auditory cortex. Neuroimage 51(2):808-816.

Ghazanfar AA, Turesson HK, Maier JX, van DR, Patterson RD, Logothetis NK (2007) Vocal-tract resonances as indexical cues in rhesus monkeys. Current Biology 17:425-430.

Ghazanfar AA, Hauser MD (2001) The auditory behaviour of primates: a neuroethological perspective. Current Opinion in Neurobiology 11(6):712-720.

Gonzales J (2004) Formant frequencies and body size of speaker: a weak relationship in adult humans. Journal of Phonetics 32:277-287.

Goto M, Hashiguchi H, Nishimura T, Oka R (2003) RWC music database: music genre database and musical instrument sound database. Proceedings of the International Symposium for Music Information Retrieval:229–230.

Gouzoules H, Gouzoules S (1989) Design features and developmental modification of pigtail macaques, *Macaca nemestrina*, agonistic screams. Animal Behaviour 37:383-401.

Gouzoules H, Gouzoules S (1990) Body size effects on the acoustic structure of pigatil macaque (*Macaca nemestrina*) screams. Ethology 85:324-334.

Gouzoules H, Gouzoules S (1990) Matrilineal signatures in the recruitment screams of pigtail macaques, *Macaca nemestrina*. Behaviour 115:328-347.

Gouzoules H, Gouzoules S (1995) Recruitment screams of pigatil monkeys (*Macaca nemestrina*): ontogenetic perspectives. Behaviour 132:431-450.

Gouzoules H, Gouzoules S (2000) Agonistic screams differ among four species of macaques: the significance of motivation-structural rules. Animal Behaviour 59(3):501-512.

Gouzoules H, Gouzoules S, Tomaszycki M (1998) Agonistic screams and the classification of dominance relationships: are monkeys fuzzy logicians? Animal Behaviour 55:51-60.

Gouzoules S, Gouzoules H, Marler P (1984) Rhesus monkey (*Macaca mulatta*) screams: representational signaling in the recruitment of agonistic aid. Animal Behaviour 32:182-193.

Griffiths TD, Uppenkamp S, Johnsrude I, Josephs O, Patterson RD (2001) Encoding of the temporal regularity of sound in the human brainstem. Nature Neuroscience 4:633-637.

Gros-Louis J (2004) The function of food-associated calls in white-faced capuchin monkeys, *Cebus capucinus*, from the perspective of the signaler. Animal Behaviour 67:431-440.

Hall DA, Plack CJ (2007) The human 'pitch center' responds differently to iterated noise and Huggins pitch. Neuroreport 18:323-327.

Hall DA, Plack CJ (2009) Pitch processing sites in the human auditory brain. Cerebral Cortex 19(3):576-85.

Hammerschmidt K, Fischer J (1998) Maternal discrimination of offspring vocalizations in Barbary macaques (*Macaca sylvanus*). Primates 39:231-236.

Hammerschmidt K, Fischer J (1998) The vocal repertoire of Barbary macaques: A quantitative analysis of a graded signal system Ethology, 104(3) 203-16.

Hammerschmidt K, Newman JD, Champoux M, Suomi SJ (2000) Changes in rhesus macaque 'coo' vocalizations during early development. Ethology 106(10):873-886.

Hammerschmidt K, Ansorge V, Fischer J (1994) Age-related variations in the vocal repertoire of barbary macaques in Current Primatology II:287-295. Roeder JJ, Thierry B, Anderson JR, Herrenschmidt N, Eds. Universite Louis Pasteur, Strasbourg.

Hasek CS, Singh S, Murry T (1980) Acoustic attributes of preadolescent voices. Journal of the Acoustical Society of America 68:1292-1265.

Hauser MD, Evans CS, Marler P (1993) The role of articulation in the production of rhesus monkey, *Macaca mulatta*, vocalizations. Animal Behaviour 45:423-433.

Hewitt G, MacLarnon A, Jones KE (2002) The functions of laryngeal air sacs in primates: a new hypothesis. Folia Primatologica 73(2-3):70-94.

Holt RF, Kirk KI (2005) Speech and language development in cognitively delayed children with cochlear implants. Ear and Hearing 26(2):131-148.

Houben MMJ, Kohlrausch A, Hermes DJ (2001) Auditory cues determining the perception of the size and speed of rolling balls. Proceedings of the 2001 International Conference on Auditory Display, Espoo, Finland, July 29-August 1 2001,105-110.

Inoue M (1988) Age gradations in vocalization and body weight in Japanese monkeys (*Macaca fuscata*). Folia Primatologica 51:76–86.

Irino T, Patterson RD (2002) Segregating information about the size and shape of the vocal tract using a time-domain auditory model: The stabilized wavelet-Mellin transform. Speech Communication 36:181-203.

Ives DT, Smith DR, Patterson RD (2005) Discrimination of speaker size from syllable phrases. Journal of the Acoustical Society of America 118:3816-3822.

Kawahara H, Irino T (2004) Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation. in Speech Separation by Humans and Machines: 167–180. Eds Divenyi P. Kluwer Academic, Boston.

Kawahara H, Masuda-Kasuse I, de Cheveigne A (1999) Restructuring speech representations using pitch-adaptive time-frequency smoothing and instantaneous-frequency-based F0 extraction: Possible role of repetitive structure in sounds. Speech Communication 27:187–207.

Kovačić D, Balaban E (2009) Voice gender perception by cochlear implantees. Journal of the Acoustical Society of America 126(2):762-775.

Kovačić D, Balaban E (2010) Hearing history influences voice gender perceptual performance in cochlear implant users. Ear and Hearing 31:806-814.

Laneau J, Wouters J (2004) Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees. Journal of the Acoustical Society of America 116:3606-3619.

Lee S, Potamianos A, Narayan S (1999) Acoustics of children's speech; developmental changes of temporal and spectral paramaters. Journal of the Acoustical Society of America 105:1455-1468.

Lieberman P (1968) Primate vocalization and human linguistic ability. Journal of the Acoustical Society of America 44:1574-1584.

Lieberman P (1969) On the acoustic analysis of primate vocalizations. Behavioral Research Methods and Instrumentation 1:169-174.

Lieberman P (1984) The biology and evolution of language. Harvard University Press, Cambridge.

Lieberman P (2006) Limits on tongue deformation--Diana monkey formants and the impossible vocal tract shapes proposed by Riede et al. (2005). Journal of Human Evolution 50(2):219-21.

Loizou PC (1998) Introduction to cochlear implants. IEE Engineering in Medicine and Biology Magazine: The Quarterly Magazine of the Engineering in Medicine and Biology Society 18:32-42.

Marks LE, Hammeal R, Bornstein M (1987) Perceiving similarity and comprehending metaphor. Monographs of the Society for Research in Child Development 52.

Matsumura S (1999) The evolution of egalitarian and despotic social systems among macaques. Primates 40(1):23-31.

McComb K (1991) Female choice for high roaring rates in red deer, *Cervus elaphus*. Animal Behaviour 41:79–88.

McKay CM, McDermott HJ (1994) Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees. Journal of the Acoustical Society of America 96:2664-2673.

McKay CM, McDermott HJ (1995) Pitch matching of amplitude-modulated current pulse trains by cochlear implantees: the effect of modulation depth. Journal of the Acoustical Society of America 97:1777-1785.

Mehler J, Bertoncini J, Barriere M (1978) Infant recognition of mother's voice. Perception 7:491-497.

Mènard N, Vallet D (1993) Dynamics of fission in a wild barbary macaque group (*Macaca sylvanus*). International Journal of Primatology 14:479-500.

Mondloch CJ, Maurer D (2004) Do small white balls squeak? Pitch-object correspondences in young children. Cognitive, Affective and Behavioural Neuroscience 42:133-136.

Morton ES (1977) On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. American Naturalist 111(981):855-869.

Moore DR (2002) Auditory development and the role of experience. British Mediacl Bulletin 63 (1): 171-181.

Nicholas JG, Geers AE (2007) Will they catch up? The role of age at cochlear implantation in the spoken language development of children with severe to profound hearing loss. Journal of Speech, Language and Hearing Research 50:1048-1062.

Nie K, Barco A, Fan-Gang Z (2006) Spectral and temporal cues in cochlear implant speech perception. Ear and Hearing 27(2):208-217.

Oda R, Masataka N (1995) Function of copulatory vocalizations in mate choice by females of Japanese macaques (*Macaca fuscata*). Folia Primatologica 64(3):132-139.

Owren MJ, Bernacki R (1988) ‘The acoustic features of vervet monkey (*Cercopithecus aethiops*) alarm calls. Journal of the Acoustical Society of America 83:1927–1935.

Owren MJ, Seyfarth RM, Cheney DL (1997) The acoustic features of vowel-like grunt calls in chacma baboons (*Papio cyncephalus ursinus*): Implications for production processes and functions. Journal of the Acoustical Society of America 101:2951–2963.

Owren MJ, Rendall D (2003) Salience of caller identity in rhesus monkey (*Macaca mulatta*) coos and screams: perceptual experiments with human (*Homo sapiens*) listeners. Journal of Comparative Psychology 117(4):380-390.

Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD (2002) The processing of temporal pitch and melody information in auditory cortex. Neuron 36:767-776.

Peters M, Ploog D (1973) Communication among primates. Annual Reiew of Physiology 35:221-242.

Peterson GE, Barney HL (1952) Control methods used in the study of vowels. Journal of the Acoustical Society of America 24:175-184.

Pfefferle D, Fischer J (2006) Sounds and size: identification of acoustic variables that reflect body size in hamadryas baboons, *Papio hamadryas*. Animal Behaviour 72: 43-51.

Prud'Homme J (1991) Group fission in a semifree-ranging population of Barbary Macaques (*Macaca sylvanus*). Primates 32(1):9-22.

Reby D, McComb K (2003) Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags. Animal Behaviour 65:519-530.

Reby D, McComb K (2007) Female red deer prefer the roars of larger males. Biology Letters 3(4):382-385.

Reby D, McComb K, Cargnelutti B, Darwin C, Fitch WT, Clutton-Brock T (2005) Red deer stags use formants as assessment cues during intrasexual agonistic interactions. Proceedings of Biological Sciences 272:941-947.

Rendall D, Kollias S, Ney C, Lloyd P (2005) Pitch (F0) and formant profiles of human vowels and vowel-like baboon grunts: the role of vocalizer body size and voice-acoustic allometry. Journal of the Acoustical Society of America 117:944-955.

Rendall D, Vokey JR, Nemeth C (2007) Lifting the curtain on the Wizard of Oz: biased voice-based impressions of speaker size. Journal of Experimental Psychology: Human Perception and Performance 33:1208-1219.

Riede T, Fitch T (1999) Vocal tract length and acoustics of vocalization in the domestic dog (*Canis familiaris*). Journal of Experimental Biology 202:2859-2867.

Riede T, Zuberbühler K (2003) The relationship between acoustic structure and semantic information in Diana monkey alarm vocalization. Journal of the Acoustical Society of America 114:1132-1142.

Riede T, Bronson E, Hatzikirou H, Zuberbuhler K (2005) Vocal production mechanisms in a non-human primate: morphological data and a model. Journal of Human Evolution 48(1):85-96.

Riede T, Bronson E, Hatzikirou H, Zuberbuhler K (2006) Multiple discontinuities in nonhuman vocal tracts - A response to Lieberman (2006). Journal of Human Evolution 50(2):222-225.

Riede, T., M. J. Owren, and A. C. Arcadi (2004) Nonlinear acoustics in pant hoots of common chimpanzees (*Pan troglodytes*): frequency jumps, subharmonics, biphonation, and deterministic chaos. American Journal of Primatology 64(3):277-291.

Rosen S (1992) Temporal information in speech: Acoustic, auditory, and linguistic aspects. Philosophical Transactions of the Royal Society of London, Series B 336:367–373.

Rosen S, Faulkner A, Wilkinson L (1999) Adaptation by normal listeners to upward spectral shifts of speech: Implications for cochlear implants. Journal of the Acoustical Society of America 106(6):3629-3636.

Semple S (1998) The function of Barbary macaque copulation calls. Proceedings of the Royal Society of London Series B-Biological Sciences 265(1393):287-291.

Siegel S, Castellan NJ (1988). Nonparametric statistics for the behavioral sciences. McGraw-Hill: London.

Siebert ER, Parr LA (2003) A structural and contextual analysis of chimpanzee screams. Annals of the New York Academy of Sciences 1000:104-109.

Slocombe KE, Zuberbuhler K (2005) Agonistic screams in wild chimpanzees (*Pan troglodytes schweinfurthii*) vary as a function of social role. Journal of Comparative Psychology119(1):67-77.

Slocombe KE, Zuberbuhler K (2007) Chimpanzees modify recruitment screams as a function of audience composition. Proceedings of the National Academy of Sciences of the United States of America 104(43):17228-17233.

Small MF (1990) Promiscuity in Barbary Macaques (*Macaca sylvanus*). American Journal of Primatology 20(4):267-282.

Smith DR, Patterson RD, Turner R, Kawahara H, Irino T (2005) The processing and perception of size information in speech sounds. Journal of the Acoustical Society of America 117:305-318.

Smith DR, Walters TC, Patterson RD (2007) Discrimination of speaker sex and size when glottal-pulse rate and vocal-tract length are controlled. Journal of the Acoustical Society of America 122:3628-3639.

Smith DR, Patterson RD (2005) The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. Journal of the Acoustical Society of America 118:3177-3186.

Sokal RR, Rohlf FJ (1995) Biometry (3rd edn). WH Freman and company, New York.

Taylor AM, Reby D, McComb K (2008) Human listeners attend to size information in domestic dog growls. Journal of the Acoustical Society of America 123(5):2903-2909.

Titze IR (1989) Physiologic and acoustic differences between male and female voices. Journal of the Acoustical Society of America 85:1699-1707.

Tokuda I, Riede T, Neubauer J. Owren MJ, Herzel H (2002) Nonlinear analysis of irregular animal vocalizations. Journal of the Acoustical Society of America 111(6):2908-2919.

Tomasello M, Call J (1997) Primate Cognition. Oxford: University Press.

Tomaszycki ML, Gouzoules H, Wallen K (2005) Sex differences in juvenile rhesus macaque (*Macaca mulatta*) agonistic screams: life history differences and effects of prenatal androgens. Developmental Psychobiology 47(4):318-327.

van Dinther R, Patterson RD (2006) Perception of acoustic scale and size in musical instrument sounds. Journal of the Acoustical Society of America 120:2158-2176.

Vannoni E, McElligott AG (2008) Low frequency groans indicate larger and more dominant fallow deer (*Dama dama*) males. PLoS ONE 3:e3113.

von Kriegstein K, Warren JD, Ives DT, Patterson RD, Griffiths TD (2006) Processing the acoustic effect of size in speech sounds. Neuroimage 32:368-375.

von Kriegstein K, Smith DR, Patterson RD, Ives DT, Griffiths TD (2007) Neural representation of auditory size in the human voice and in sounds from other resonant sources. Current Biology 17:1123-1128.

Wilson BS, Dorman MF (2008) Cochlear implants: A remarkable past and a brilliant future. Hearing research 242:3-21.

Zollinger SA, Riede T, Suthers RA (2008) Two-voice complexity from a single side of the syrinx in northern mockingbird *Mimus polyglottos* vocalizations. Journal of Experimental Biology 211(12):1978-1991.

Zuberbühler K (2005) The phylogenetic roots of language: evidence from primate communication and cognition. Current Directions in Psychological Science 14(3):126-130.

9 APPENDIX: IMITATION OF ACTIONS AND ITS RELATION WITH HANDEDNESS IN PRESCHOOL CHILDREN

Linda Sebastianutto, Caterina Spiezio, Evan Balaban and Raffaella Ida Rumiati

9.1 INTRODUCTION

Imitating a model performing a skill is a more efficient and less time-consuming way of learning than a trial-and-error strategy. As to the hand used for imitating, two main types of imitation have been distinguished: specular and anatomic imitation (Koski *et al.* 2003). In the first type of imitation, the subject performed the action as if he/she was in front of a mirror, therefore when a demonstrator facing the subject performed an action with the left hand the subject imitated the same action with the right hand. In the anatomical imitation, the subject imitated a gesture using the same hand used by the model.

Specular imitation relies on a simple system that directly matches the observed and executed actions, it does not require a cognitive effort such as understanding the spatial relations of the model's actions, and try to match it with the executed action (Koski *et al.* 2003). Specular imitation does not seem to engage attentional resources and this may explain why this is the children's preferred solution and why a natural tendency to use this configuration has been described also in adults (Brass et al. 2000, 2001).

Specular responses predominate over anatomic ones children when they are 10-year-old (Wapner and Cirillo 1968). To date, some studies showed that children, when required to imitate cross-midline gestures, produce ipsilateral responses. This phenomenon is called "the contralateral inhibition effect" and it has been also reported in aphasic adult patients and in epileptic and mentally retarded children (Bekkering *et al.* 2000; Schofield 1976 *a, b*). In normal children this tendency linearly decreases with age and it has been shown that in preschool children (4 year-old group) cross-lateral responses are produced mainly with the preferred hand. With age there is also a decreasing linear trend to produce all the imitative responses with the preferred hand (Gleissner *et al.* 2000; Schofield 1976 *b*).

In all the cited studies, however, pre-scholar children (Bekkering *et al.* 2000; Gleissner *et al.* 2000) and scholar children (Wapner and Cirillo 1968) imitated meaningless actions from modified versions of the original Head's Hand, Eye and Ear Test (1926) whereas adults (Koski *et al.* 2003) imitated a simple finger lifting movement.

According to the dual-route model proposed by Tessari and Rumiati (2004), healthy adults imitate meaningful (MF) and meaningless (ML) actions using two different processes: while ML actions imitation is processed using a direct motor route, imitation of MF actions relies on a semantic route, since the observation of model performing such an action allowed the retrieval of the correspondent motor pattern stored in the semantic memory. Selective impairment of either the direct or the semantic route has been described in brain damaged patients (Goldenberg and Hagmann, 1997; Bartolo *et al.* 2001): impaired imitation of MF actions has been explained as a damage of the lexical-semantic route, while an impairment in imitating ML action as a damaged sublexical-direct route (Tessari et al. 2007).

While the direct route is supposed to be already present very early in the development (Meltzoff and Moore 1977), the ontogenetic development of the semantic route has never been discussed so far.

Conceptual knowledge of tool use and the related motor patterns are segregated in the left-hemisphere. Left-brain regions were active during production of pantomimes, when subject performed the action both with the right with the left hand (Johnson-Frey 2004). Besides imitating pantomimes of objects use (or transitive actions) (the MF actions in Tessari and Rumiati 2004), there is another kind of meaningful actions that it is worth to consider, i.e. the communicative – or intransitive - gestures. Observation of intransitive gestures is correlated with activation of language related areas in the left hemisphere (Bates and Dick 2002; Gallagher and Frith 2004). Both communicative gestures and tool use are processed in the left hemisphere. Carmo and Rumiati (2009) found that healthy adults imitate better intransitive actions than transitive gestures, since the latter are considered to be more difficult to imitate given the intrinsic association with the object representation. In young children both recognition and production of pantomimes are difficult (O'Reilly 1995), but no one has addressed the imitation of these kind of gestures in healthy pre-school children, nor the comparison between transitive and intransitive gestures in these subjects.

Both in Tessari and Rumiati's (2004) and in Koski *et al.*'s study (2003) adult subjects were told to use the right hand for imitation. In the first study the model always showed actions performed with the left hand, while in the second study the pictures of either the left or the right hand were presented. On the other hand in all the cited studies with children, the model used the left, the right or both hands and the participants were free to choose the hand to perform imitation.

The direction of handedness seems to appear very early in the human development, it gradually develops with age and becomes fully stable and consistent when children are 10-years old (Gesell and Ames 1947). Children show very early in the development a hand preference in doing several actions such as grasping, reaching and manipulating tools (Fagard and Marks 2000). In toddlers (18-36 months-old) there is an age-related increase of right-handedness: 50% of two-year old children and 70% of three-year old children were categorized as right-handers (Fagard and Marks 2000). Moreover, the frequency with which the dominant hand is used seems to be dependent on the difficulty of the task: more complex actions are performed using the dominant hand with a higher probability (Bryden 2000).

How this progressive establishment of hand preference is related with the imitation of MF and ML actions? The aim of this study is to address the following questions taking into account both imitation and handedness development:

1) Do children imitate relying on two different processes, namely the direct-matching process and the semantic process?

2) Does the tendency to produce specular responses exist also in imitation of already –known actions?

3) Should the tendency to use the preferred hand in imitation parallel the establishment of handedness?

4) Is imitation of transitive gestures more difficult than imitation of intransitive gestures?

We know from the cited studies that a progressive use of the non-preferred hand is involved in the imitation of ML actions. It is plausible to expect that, if a semantic route develops within the first five years, an imitation of MF gestures should be performed always with the preferred hand, independently from the Model's hand. When a child observes a familiar action performed by a demonstrator, s/he should retrieve the motor pattern stored in memory and imitate the action with the hand s/he usually uses for that action. Therefore a specular response should be less frequent when a child imitates an already known action. A possible confounding factor could be the difficulty that young children have in both recognizing and producing pantomimes (O'Reilly 1995). For this reason this kind of action could be specularly imitated because it may not be recognized as a familiar action in the first place. On the other hand, since communicative gestures are supposed to be recognized as familiar actions, children should always perform them with their preferred hand.

Right-handed children should imitate familiar actions with the right hand. The left-handed subjects - most likely a small proportion of the sample size, since 10% is the estimated proportion in human population (Iwasaki 2000) - should imitate MF actions with their left hand. There should be also a correlation between the hand used for imitation and the degree of handedness of the subjects.

Overall, specular imitation should predominate at least for imitation of ML actions in all the age groups, while age-dependent differences should emerge in the imitation of MF actions.

9.2 METHODS

9.2.1 *PARTICIPANTS*

Sixteen 3-year-old (35.9 ± 2.6 months), sixteen 4-year-old (47.8 ± 3.5 months) and sixteen 5-year-old children (62.8 ± 3.4 months) from two primary schools in Trieste, one in Udine and one in Vigevano (Pavia) for a total of 48 children took part in the experiment. Each child was tested individually in a quiet room, except for the handedness assessment in spoon/fork use (see below). In this case the experimenter observed the children with focal sampling method during lunchtime.Eight more children (five 3-year-old; two 4-year-old and one 5-year-old) were eliminated from the study because they did not complete the whole experimental session.

9.2.2 *HANDEDNESS ASSESSMENT*

Children were required to perform five different tasks that they were free to perform using the preferred hand. The tasks were: drawing with different crayons, inserting coins in a piggy bank, throwing a small ball, putting scattered toys in a box, and foraging. For each task 10 trials were collected. Since acts could occur in series, only the first act of a trial was scored (*bout*).

To avoid a bias in hand selection, stimuli were presented symmetrically to the children in drawing and inserting coins in a piggy bank. Children were presented with only one object (crayon, coin) at a time.

For each of the five tasks, binomial z scores were calculated for each participant on the basis of the frequency of left- and right-hand use.

Children with z scores greater than 1.95 were classified as right-handed, whereas subjects with *z* scores lower than -1.95 were classified as left-handed. Subjects with *z* scores between -1.95 and 1.95 were classified as ambipreferent. The level of significance was an alpha of $p < 0.05$ (Hopkins and Cantalupo 2003). A child was scored as right-handed if z scores were significant for right hand in at least four tasks, while s/he was scored as left handed if z scores were significant for left hand in at least four tasks. In all the other cases the child was considered ambipreferent (Gleissner *et al.* 2000).

A measure of the "*handedness strength*" for each subject was also calculated. This measure was obtained by assigning a score for each task (-1 for a left hand preference, 0 for no preference, +1 for a right hand preference); the sum of scores over the five tasks was calculated for each subject. All the possible sums ranged from –5 (completely left handed subject) to +5 (completely right handed subject).

9.2.3 MEMORY ASSESSMENT

Children's working memory was tested using a digit span test (forward) and an action span test (with meaningless actions only) (Tessari and Rumiati 2002)

9.2.4 IMITATION OF MEANINGFUL AND MEANINGLESS ACTIONS.

Subjects were required to imitate the experimenter performing MF and ML actions. The within-subject variables were Model's hand (MH: right *vs* left), Type of actions (T: MF *vs* ML) and Content (C: Communicative gestures *vs* pantomimes), whereas Age was the between-subject variable.

Stimuli. The model performed twenty-one MF and twenty-one ML actions (see Appendix 1), facing the subject. All the actions were unimanual and did not cross the body midline in order to avoid the contralateral inhibition effect (Bekkering *et al.* 2000; Gleissner *et al.* 2000; Schofield 1976). Among the MF actions, 10 were communicative gestures and 11 were pantomimes of object use. Most of these actions normally are used in the clinical practice to assess the dominant hand of the child (Wille 1986). ML actions were obtained from meaningful actions by modifying the relation between the body and the hand/arm.

Design and Procedure. The two types of actions were always presented in separated blocks each of which was performed twice by the experimenter, the first time using one hand and the second time using the other hand. Within each block, the sequence of actions did not change. The order of presentation of blocked stimuli was obtained by systematically crossing two factors: the hand used by the experimenter (right first or left first) and the type of action (MF first or ML first), thus obtaining 16 different orders of presentation (see Appendix 2). Within each age group, each subject was randomly assigned to one order of presentation; model-first hand and stimulus-type were shown in a counterbalanced order.

At the beginning of each block subjects were given the following instructions: “Now I will show some movements that you know (or “you do not know” with ML actions). Do as I do”. After the child imitated the gesture, the experimenter asked the child to pick up a reward put on a table 50 cm behind the child and put it in a small plastic bag. Each experimental session was video-recorded and later scored by two independent raters.

9.2.5 QUALITATIVE ANALYSIS

The accuracy in imitation was evaluated using a 3-point scale (0 = unrecognizable; 1 = recognizable; 2 = identical to model’s action).

9.2.6 ERROR ANALYSIS

The following classes of errors were considered, according to Tessari and Rumiati (2004) and Bekkering (2001):

Contralateral error: a gesture that crosses the body midline is produced instead of an ipsilateral one.

Spatial error of the hand: the overall movement is correct but the position of the hand is wrong

Spatial error of the arm: the movement is recognizable but the position of the arm or the direction/plane of the movement is wrong

Semantic:

Prototypicalization: the subject produces his/her own version of the meaningful movement instead of the version given by the model

Body Part as a Tool: the subject perform the action using the hand/arm/finger as it were the actual tool.

Visuosemantic: the action produced is visually similar and semantically related to the perceived action.

Visual:

Partial perseveration: the produced movement is a combination of two actions presented in the list

Global perseveration: subjects performed an action included in the list instead of the one presented by the model.

Lexicalization: a visually similar, meaningful gesture not included in the list is produced instead of the meaningless action performed by the model.

Substitution: a visually similar meaningful action, not included in the list, is produced instead of the presented meaningful action.

Omission: the subject does not produce any movement

Unrecognizable gesture: the subject produces a movement that raters fail to recognize.

9.2.7 HAND PREFERENCE

The hand used by the subject to imitate each item was scored.

9.2.8 ACTION RECOGNITION

At the end of the presented actions, children were required to recognize the imitated actions. Actions were presented in random order. The experimenter asked whether the child knew the meaning of the presented action. In case of positive answer, two different procedures were used for communicative gestures and pantomimes.

For communicative gestures two possibilities (one right and one wrong) were given by the experimenter. The order of the correct and the wrong response were counterbalanced for each subject.

For the pantomimes a multiple choice response was required: the experimenter put in front of the child all the ten real objects she pretended to use performing pantomimes. The child was required to indicate the correct object.

9.3 RESULTS

9.3.1 INTER-RATER'S AGREEMENT

In order to assess the agreement of the two rater's observation, a Cohen's kappa coefficient was calculated for each subject on both the accuracy and the error analyses. The obtained Kappa coefficients were 0.87 ± 0.07 and 0.90 ± 0.08 (M ± SD) respectively, corresponding to a very good agreement. Hence, analyses were performed on the average scores of the two raters.

9.3.2 HANDEDNESS ASSESSMENT

According to the results obtained in the five tasks, subjects were classified as right- or left-handed or as ambipreferent subjects as presented in Table 12.

Age Group	Right Handed	Left Handed	Ambipreferent
3 YO	8	0	8
4 YO	8	1	7
5 YO	12	0	4

Table 12 Distribution of subjects according to their hand preference

Considering the measure of *handedness strength*, results are presented in Fig 40. As can be expected, the great proportion of subjects shows a strong right hand preference. No completely left handed subject was present in the sample. It should be noted, however, that 3 year-old children showed a higher proportion of left handed subjects if compared with the other two age groups.

Fig 40 Distribution of handedness strength among subjects

9.3.3 *MEMORY SPAN*

Simple linear regression analysis was performed in order to investigate if the performance in action or digit span significantly changed with age (Fig 41). Digit span significantly increased with age ($r^2 = 0.43$, $N = 48$, $p < 0.001$), as well as action span did ($r^2 = 0.49$, $N = 48$, $p < 0.001$). As it can be expected, these results confirm that working memory increased with age.

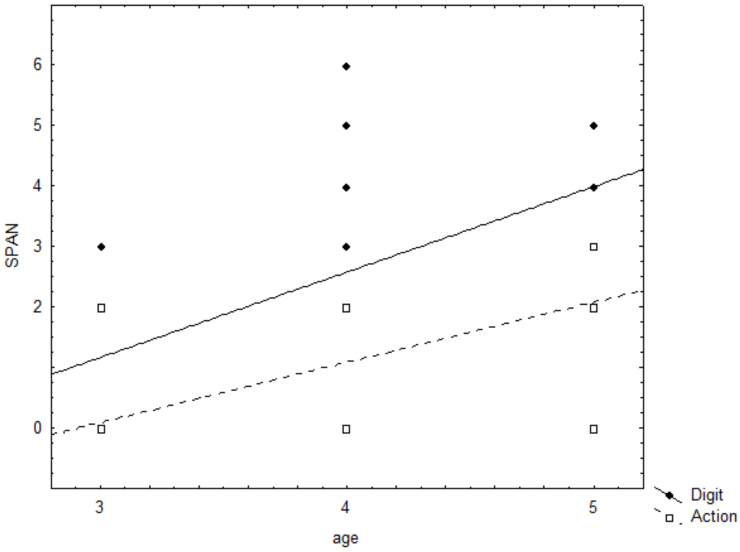

Fig 41 Relation between age and memory span

9.3.4 IMITATION

9.3.4.1 Accuracy

A repeated measure ANOVA was performed, with type of action (MF *vs* ML), Model's Hand (MH, right *vs* left) and CONTENT (transitive *vs* intransitive) as within-subject factors and Age as between subject factor. Accuracy was the dependent variable.

The effects of factor age (F(2, 45) = 13.69, p < 0.001), Model's hand (F(1,45) = 5.81, p = 0.02) and Content (F(1,45) = 109.36, p < 0.001) were significant. Accuracy increased with age (M ± SE: 3YO: 46.13 ± 2.34; 4 YO: 54.81 ± 2.337; 5 YO: 63.42 ± 2.34). Performance was better on intransitive actions (M ± SE: 59.1 ± 1.47) than on transitive actions (M ± SE: 50.5 ± 1.35) and it was better when the model uses the left hand (M ± SE: 56.00 ± 1.44) instead of the right hand (M ± SE: 53.57 ± 1.44).

The interaction Model's Hand X Type X Content X Age was also significant (F(2, 45) = 5.02, p = 0.01- Fig 42).

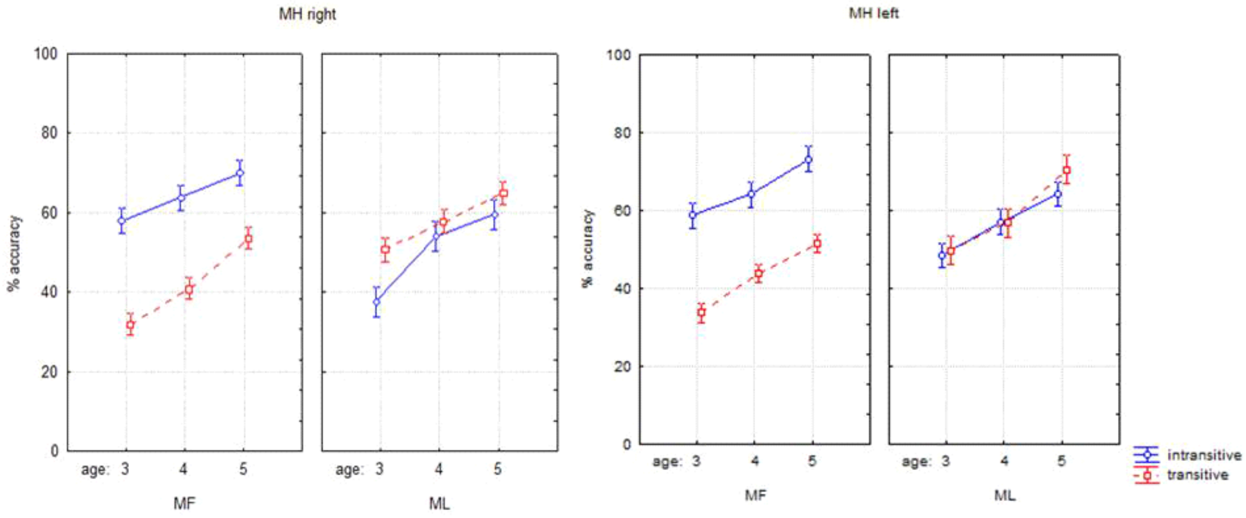

Fig 42 Significant interaction Model's Hand X Type X Content X Age (M±Se)

The same analysis was done only on Right Handed subjects. In this case the effects of factor Age (F(2, 25) = 11.54, p < 0.001), Type of action (F(1, 25) = 9.52, p = 0.005) and Content (F(1, 25) = 61.08, p < 0.001) were significant. Accuracy increased with age (M ± SE: 3YO: 47.72 ± 2.87; 4 YO: 55.32 ± 2.87; 5 YO:65.22 ± 2.34). The imitation was better for ML actions (M ± SE:57.98 ± 1.73) respect to MF actions (M ± SE: 54.19 ± 1.62) and on intransitive gestures (M ± SE:60.49 ± 1.82) respect to transitive gestures (M ± SE: 51.67 ± 1.48).

The interaction Type X Content was also significant (F(1, 25) = 149.38, p < 0.001) - Fig 43. Post-hoc Tukey HSD test revealed that all the comparisons made were significant (intransitive gestures: MF *vs* ML p < 0.001; transitive gestures: MF *vs* ML p < 0.001; MF: intransitive *vs* transitive p < 0.001; ML: intransitive *vs* transitive p = 0.04).

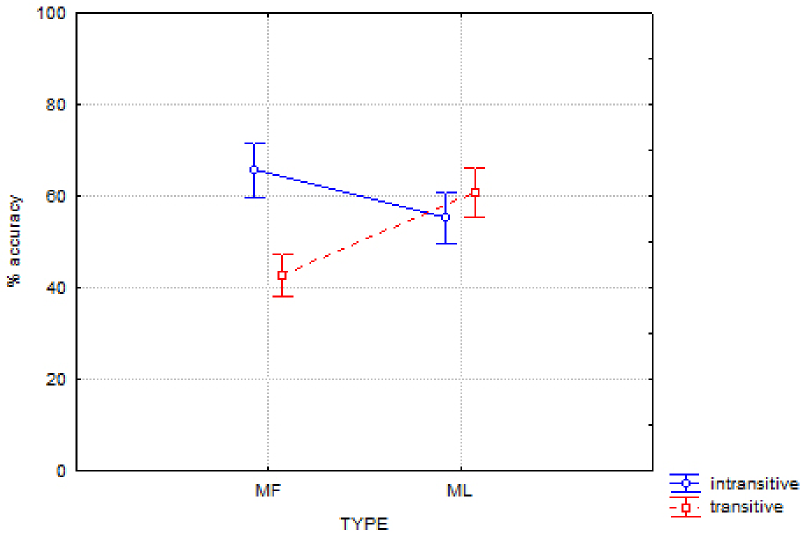

Fig 43 The interaction Type X Content (M±SE)

In order to better understand the 4-way interaction, two separate analyses were carried out for transitive and intransitive actions, respectively. Within subject factors were Model's Hand and Type of action while Age was the between subject factor.

Looking only at transitive gestures, significant effects of Age (F(2, 45)=15.86, p < 0.001) and Type of action (F(1, 45)=123.98, p < 0.001) were found. Imitation of transitive gestures was better on ML gestures (M ± SE: 58.40 ± 1.71) respect to MF action (M ± SE: 42.61 ± 1.31), a parallel analyses done only on right handed subjects revealed the same significant effects. No effect of Model's hand was found.

The analyses on intransitive actions revealed a significant effect of Age (F(2, 25) = 7.30. p = 0.003), and Model's hand (F(1, 45) = 6.83, p = 0.01) of type of action (F(1, 25) = 42.40. p < 0.001), however the imitative performance on intransitive gestures was better on MF actions (M ± SE: 64.64 ± 1.67) respect to ML actions (M ± SE: 53.49 ± 1.60), and better when the models used her left hand (M ± SE: 60.96 ± 1.54) respect to the right hand (M ± SE: 57.16 ± 1.73). The same type of analysis only on right handed subjects showed the significant effects of Age and Type of actions and a significant TYPE X AGE interaction (F(2, 25) = 3.41, p = 0.05).

These results suggested that imitation of transitive and intransitive actions could rely on two different system: while imitation of intransitive gestures relies on the semantic route, the imitation of pantomimes relies on the direct route.

9.3.4.2 Error analysis

	3 YO				*4 YO*				*5 YO*			
	ML	**MF**	**MF-trans**	**MF-intrans**	**ML**	**MF**	**MF-trans**	**MF-intrans**	**ML**	**MF**	**MF-trans**	**MF-intrans**
Contralateral	0.34 (0.47)	0.22 (0.41)	0.19 (0.40)	0.03 (0.13)	0.63 (0.96)	0.34 (0.60)	0.13 (0.34)	0.22 (0.55)	0.72 (0.88)	0.72 (1.46)	0.34 (0.57)	0.38 (1.01)
Spatial hand	14.88 (3.37)	15.22 (3.92)	8.75 (2.46)	6.47 (2.47)	15.16 (3.36)	13.84 (3.79)	8.81 (2.73)	5.03 (2.27)	11.50 (3.98)	12.44 (4.67)	8.91 (2.81)	3.53 (2.35)
Spatial arm	21.34 (3.01)	16.47 (3.68)	10.00 (2.20)	6.47 (2.67)	18.94 (4.62)	19.00 (3.46)	11.41 (2.79)	7.59 (2.79)	16.78 (5.55)	16.78 (3.44)	10.44 (2.12)	6.34 (1.96)
Prototipicalization		0.69 (1.17)	0.16 (0.625)	0.28 (0.52)		0.69 (0.73)	0.22 (0.41)	0.47 (0.76)		0.59 (0.82)	0.50 (0.77)	0.09 (0.27)
BPT		3.78 (2.07)	3.59 (2.27)			2.31 (1.31)	2.25 (1.28)			1.16 (0.98)	1.09 (0.95)	
Visuosemantic		0.84 (0.72)	0.81 (0.75)	0.03 (0.11)		0.44 (0.48)	0.41 (0.49)	0.03 (0.13)		0.31 (0.54)	0.19 (0.36)	0.13 (0.29)
Partial perseveration	2.56 (2.48)	2.84 (2.48)	1.94 (1.78)	0.91 (1.24)	1.31 (1.18)	2.34 (1.89)	1.34 (0.94)	1.00 (1.25)	0.53 (0.87)	1.00 (0.71)	0.56 (0.68)	0.44 (0.57)
Global perseveration	0.94 (0.98)	1.41 (0.92)	0.44 (0.66)	0.97 (0.74)	0.44 (0.89)	1.00 (0.84)	0.41 (0.71)	0.59 (0.61)	0.13 (0.39)	0.22 (0.31)	0.06 (0.25)	0.16 (0.24)
Lexicalization	1.34 (1.26)				1.34 (0.98)				1.25 (1.06)			
Substitution		0.13 (0.50)	0.06 (0.25)	0.06 (0.25)		0.09 (0.27)	0.09 (0.27)	0 (0)		0.22 (0.26)	0.13 (0.22)	0.09 (0.20)
Omission	1.03 (1.49)	0.69 (1.08)	0.44 (0.89)	0.25 (0.58)	0.09 (0.27)	0.25 (0.77)	0.18 (0.75)	0.06 (0.25)	0.00 (0)	0.00 (0)	0.00 (0)	0.00 (0)
Unrecognizable	1.16 (1.87)	3.31 (2.43)	2.56 (1.66)	0.75 (0.93)	1.78 (2.64)	2.38 (1.73)	1.84 (1.31)	0.53 (0.96)	0.72 (1.40)	1.28 (1.30)	1.16 (1.15)	0.13 (0.29)

Table 13 Mean number (and standard deviation) of errors during the imitation of 42ML and 42MF actions in each experimental session

Non parametric analyses were run in order to test whether the number of different errors varied according to age, model's hand (MH), Type of action or Content. Results are summarized in Table 14.

	Effect of age	**Effect of MH**	**Effect of type of action**[5]	**Effect of content (only MF)**
Contralateral	NS (If MH right, p < 0.001;if MH left, p NS)	p = 0.002	NS	NS
Spatial_hand	NS	NS	NS	p < 0.001
Spatial_arm	NS (if MF p NS; if ML p = 0.04)	p = 0.03	NS	p < 0.001
Sem_prototipicalization	NS	NS		NS
BPT	p < 0.001	NS		
Visuosemantic	p = 0.05	NS		p < 0.001
Partial perseveration	p < 0.001	NS	p = 0.007	p = 0.03
Global perseveration	p < 0.001	NS	p = 0.01	p = 0.03
Lexicalization	NS	NS		
Substitution	p = 0.04	NS		NS
Omission	p < 0.001	NS	NS	NS
Unrecognizable	NS (if MF p = 0.02, if ML p NS)	NS	p < 0.001	p < 0.001

Table 14 Results of non parametric tests to check the effect of Model's hand, Type of action and Content on different types of error

Here follows a detailed discussion of the significant results.

- *Contralateral errors*

No effect of age on the mean number of contralateral errors was found. Differences emerged distinguishing the model’s hand: more errors occurred when the model used the right hand (M = 0.86; SD = 1.55) than when the model used the left hand (M = 0.13; SD = 0.38). An effect of age in these two cases was tested: when the model used the right hand the number of errors increased with age (3YO: M = 0.31, SD = 0.60; 4YO: M = 0.91, SD = 1.29; 5YO: M = 1.38, SD = 2.20).

- *Spatial error of the hand*

There was a significant effect of Content: children produce more errors with transitive gestures (M = 8.82; SD = 2.61) respect to intransitive gestures (M = 5.01; SD = 2.61).

- *Spatial error of the arm*

More errors were made when subjects imitate ML (M = 19.02, SD = 4.81) than MF actions (M = 17.2, SD = 3.64).

Only on imitation of ML actions, subjects showed a tendency to produce less errors as they grew (3YO: M = 31.34, SD = 3.01; 4YO: M = 18.94, SD = 4.62; 5YO: M = 16.78, SD = 5.55). There was a significant effect of Content: children produce more errors with transitive gestures (M = 10.61; SD = 2.41) respect to intransitive gestures (M = 6.80; SD = 2.52).

- *BPT*

The occurrence of BPT errors decreased with age (3YO: M = 3.78, SD = 2.07; 4YO: M = 2.31, SD = 1.31 ; 5YO: M = 1.16, SD = 0.98). These results are in accord with O’Reilly (1995).

- *Visuosematic*

The occurrence of visuosematic errors decreased with age (3YO: M = 0.84, SD = 0.72; 4YO: M = 0.44, SD = 0.48 ; 5YO: M = 0.31, SD = 0.54). These errors occurred more often on transitive gestures (M = 0.47; SD = 0.60) respect to intransitive gestures (M = 0.05; SD = 0.19).

- *Perseverative errors*

Both partial and global perseverations show the same trend: their number decreased with age and they were more frequent on MF actions.

1. **Partial Perseveration:** effect of age (3YO: M = 5.41, SD = 4.11; 4YO: M = 3.66, SD = 2.41 ; 5YO: M = 1.53, SD = 1.28). This kind of error occurred more with MF actions (M = 2.06, SD = 1.97) if compared with ML actions (M = 1.47, SD = 1.83). Partial perseveration occurred more often on transitive gestures (M = 1.28, SD = 1.33) respect to intransitive gestures (M = 0.78, SD = 1.08)

2. **Global perseveration:** effect of age (3YO: M = 2.34, SD = 1.50; 4YO: M = 1.44, SD = 1.45; 5YO: M = 0.34, SD = 0.57). This kind of error occurred more frequently with MF action (M = 0.87, SD = 0.88) if compared with ML actions (M = 0.50. SD = 0.85). Global perseveration occurred more often on intransitive gestures (M = 0.57, SD = 0.65) respect to transitive gestures (M = 0.30. SD = 0.59).

Linear regression analysis (

Fig 44 Correlation between perserverative errors and digit and action span) showed a significant negative correlation between the number of perseverative errors and both action (r^2 = 0.27, N = 43 p < 0.001) and digit span (r^2 = 0.27, N = 46 p < 0.001).

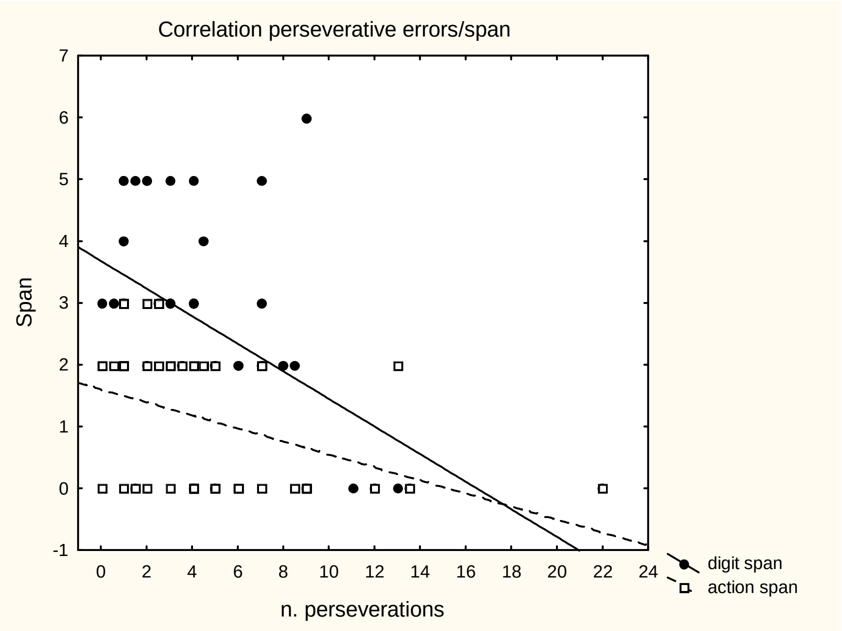

Fig 44 Correlation between perserverative errors and digit and action span

- *Substitution*

There was a significant effect of age (3YO: M = 0.13, SD = 0.50; 4YO: M = 0.09, SD = 0.27; 5YO: M = 0.22, SD = 0.26).

- *Omission*

Omissions occurred mainly in 3-year-olds, and the number decreased with age (3YO: M = 1.72, SD = 2.16; 4YO: M = 0.34, SD = 1.01; 5YO: M = 0. SD = 0).

- *Unrecognizable*

Unrecognizable gestures were produced more often when children imitated MF actions (M = 2.32, SD = 2.02) respect to ML action (M = 1.22 , SD = 2.04).

The number of unrecognizable gestures produced during imitation of MF action decreased with age (3YO: M = 3.31, SD = 2.43; 4YO: M = 2.38, SD = 1.73; 5YO: M = 1.28, SD = 1.30).

More unrecognizable gestures where produced during imitation of transitive gestures (M = 0.65, SD = 1.19) respect to intransitive gestures (M = 0.57, SD = 1.04)

9.3.5 HAND PREFERENCE

In general, subjects used their preferred hand (usually the right) 81.75% (SD 23.03) of the times, while they used their non preferred hand 16 % of the times (SD 22.60). This difference was significant (paired-wised t (47) = 10.01, p < 0.001).

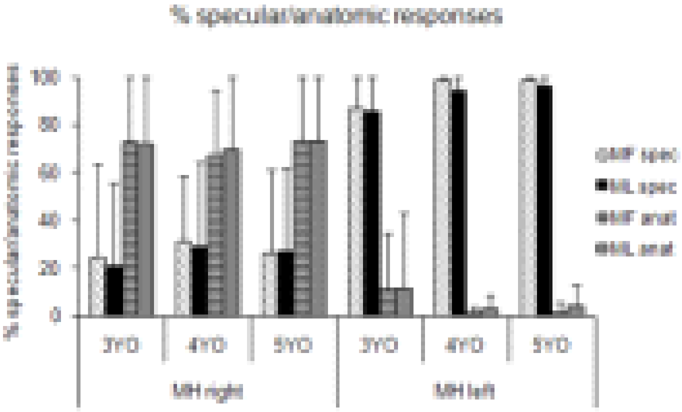

Fig 45 Mean percentages (+sd) of specular/anatomic responses in the three age groups.

A repeated measure ANOVA was performed, with type of action (MF *vs* ML), Content (transitive vs intransitive gestures) and Model's Hand (MH, right *vs* left) as within-subject factors and Age as between-subject factor. The percentage of right hand use was the dependent variable.

A significant effect Model's Hand was found (F(1, 45) = 34.38, $p < 0.001$): children imitated with their right hand more often when the Model used her left hand (M ± SE: MH left: 93.66 ± 2.21; MH right 72.27 ± 4.50).

A significant interaction CONTENT X TYPE X Age was also found F(2, 45) = 3.29, $p = 0.05$ (see Fig 46)

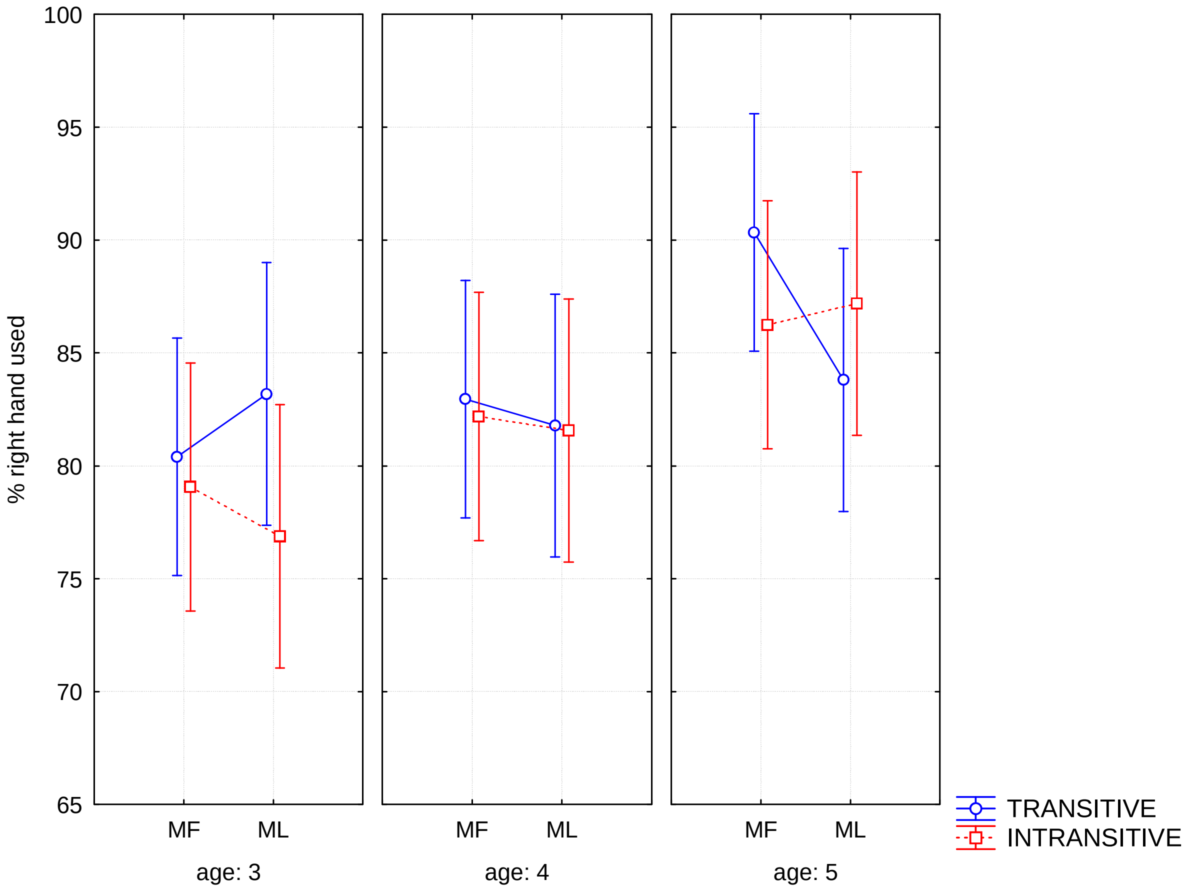

Fig 46 The significant interaction Content X Type X Age (M±SE)

The handedness strength was significantly and positively correlated with the number of responses given with the right hand ($r^2 = 0.19$, N = 48, $p < 0.001$) and negatively correlated with the number of responses given with the left hand ($r^2 = 0.18$, N = 48, $p = 0.002$) (Fig 47).

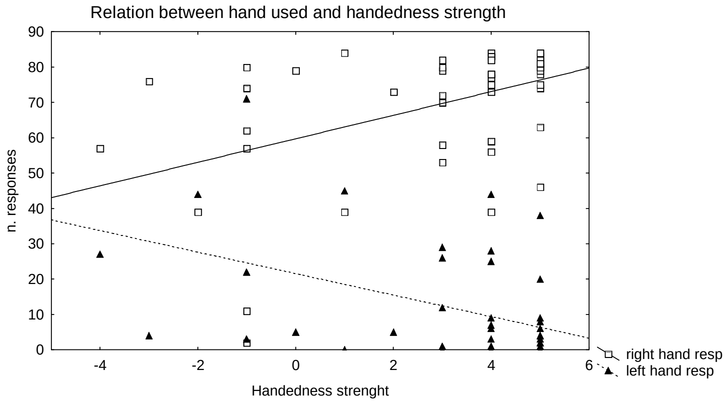

Fig 47 Correlation between hand preference and handedness strength

9.3.6 ACTION RECOGNITION

Action recognition was required to a subset of the final sample (N = 42).

A repeated measure ANOVA with Type of action (MF vs ML) and Content (transitive vs intransitive) as within-subject factor and Age as between subject factor was run. The percentage of correct responses (i.e. to associate the right meaning to a MF action or do not find any related meaning to a ML action) was the dependent variable.

The factor Age was significant (F(2, 39) = 3.83, p = 0.03): the percentage of correct responses in 3-years old were (M ± SE) 81.75 ± 2.26, in 4-years-old were 89.12 ± 1.79, in 5 years-old were 88.68 ± 1.79. A significant effect of type of action was also found (F(1, 39) = 9.76, p = 0.003): children gave more correct responses on ML actions (88.65 ± 1.33) respect to MF actions (84.38 ± 1.30).

The factor content was also significant (F(1, 39)=80.32, p < 0.001): intransitive gestures were recognized much better than transitive gestures (91.96 ± 1.25 and 81.97 ± 1.23 respectively).

There was a significant interaction Content x Age (F(2, 39) = 6.52, p = 0.004) – see Fig 48.

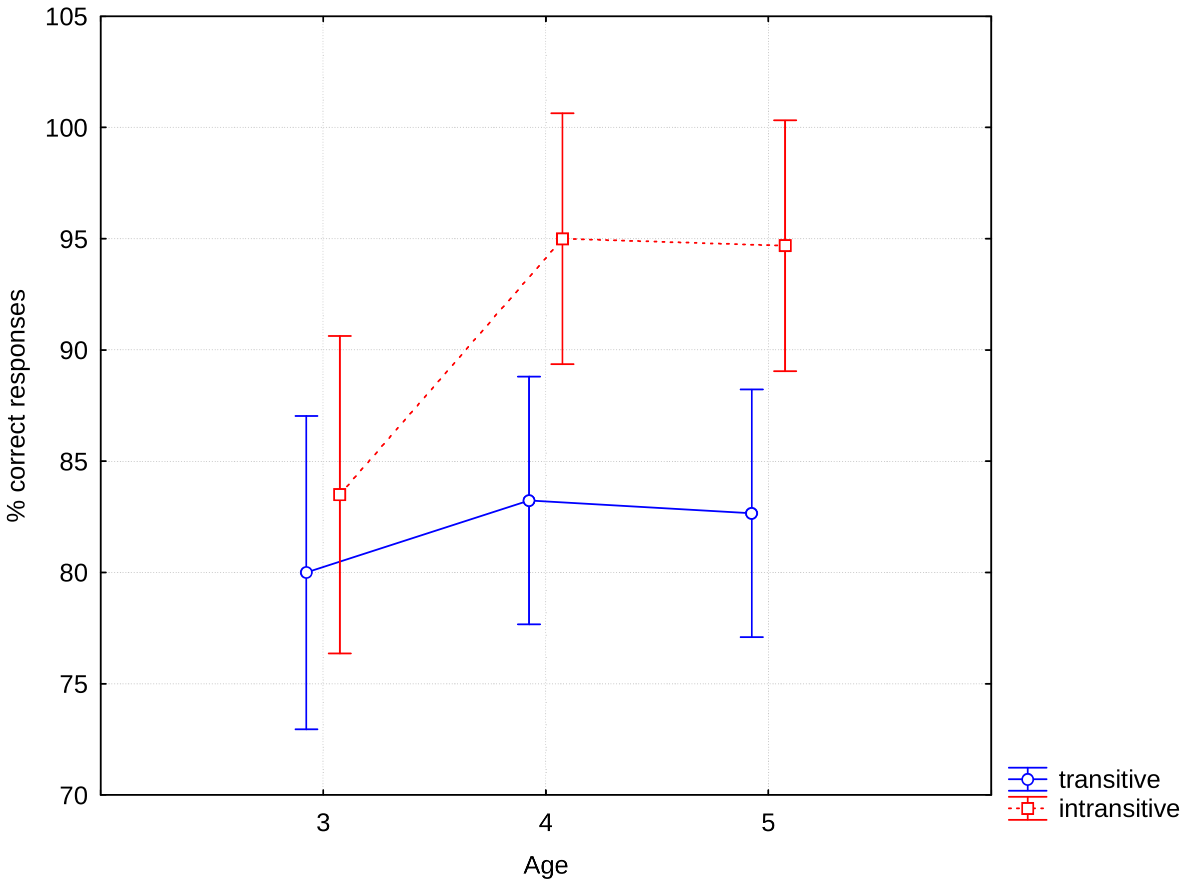

Fig 48 Significant interaction Content x Age (M±SE)

Post-hoc tests revealed that while there were non-significant differences between recognition of transitive and intransitive gestures in 3-year olds, intransitive gestures were recognized significantly better than transitive gestures both in 4- and in 5 –years old (always p < 0.001).

There were no significant differences among age groups in the recognition of transitive gestures, while 4- and 5-years old recognize intransitive gestures significantly better than 3 years-olds (all ps = 0.01).

Overall these results suggest that children as young as 3-years old have have semantic representation of both transitive and intransitive gestures.

A positive linear correlation between action recognition and imitative performance was found for MF actions ($p < 0.001$, $r^2 = 0.39$).

9.4 DISCUSSION

9.4.1 IMITATION

Our results suggest that children as young as 3 can imitate relying on two different routes, i.e. the semantic and the direct route. While the direct route is supposed to be present very early in the development (Meltzoff and Moore 1977), nobody showed so far evidences of existence of a semantic route.

With this study we showed that at least from the age of 3 children have a semantic representation of both transitive and intransitive gestures, but also that, when ML and MF action are presented in separate blocks, the latter are imitated using a semantic route. The double dissociation showed for imitation of transitive and intransitive gestures (i.e. better imitative performance on MF intransitive gestures respect to ML intransitive gestures and the opposite pattern in imitation of transitive gestures), could at first glance suggest that these two types of actions are imitated using two separate strategies: the semantic route for imitation of intransitive gestures and the direct route for imitation of transitive gestures. However no one has described so far a switch between different imitative strategies within the same block of trials.

Looking more in detail at the error pattern, it turned out that the lower accuracy found on MF transitive gestures was due to semantic errors such as Body Part as a Tool, indicating eventually that transitive gestures too are imitated using a semantic route.

Imitation of transitive gestures was more difficult than imitation of intransitive gestures. This pattern was already described in healthy adults under time constraint by Carmo and Rumiati (2009). These authors suggested that imitation of pantomimes requires a higher cognitive demand, since it is more difficult because of the intrinsic association with object representation. This could be referred to our sample as well.

9.4.2 HANDEDNESS

In our sample the handedneess assessment showed that, even if not completely lateralized, all the children but one had the right dominant hand. Subjects in our sample used mainly the preferred hand to imitate, independently from the hand used by the model facing them.

This is quite new, since many studies showed that specular imitation is the children's preferred solution and a natural tendency to use it is present also in adults (Brass *et al.* 2000. 2001). Specular imitation relies on a simple system that directly matches the observed and executed actions, does not require a cognitive effort such as understanding the spatial relations of the model's actions (Koski *et al.* 2003) and does engage attentional resources. In theory specular responses should predominate over anatomic responses in children as young as those in our sample, as many earlier studies showed (Wapner and Cirillo 1968, Gleissner *et al.* 2000; Schofield 1976 *b*). Our results did not confirm this trend, however we could still observe a tendency to imitate specularly over than anatomically in two results: the higher imitative performance when the Model used her left hand (i.e. more errors are produced when subjects were required to understand the spatial relation of the model's action) and the lower number of responses given with the right hand when the Model used her right hand.

9.4.3 *GENERAL CONCLUSION*

Our study sheds a new light in the ontogenetic perspective of imitative strategies: future studies are needed in order to understand *when* the semantic route actually develops, using meaningful stimuli instead of the simple, meaningless stimuli used so far in all the developmental studies on imitation.

As to the difference between transitive and intransitive gestures, even if we found a complementary pattern of imitation, we did not find any evidence of different mechanisms dedicated to the imitation of these two kinds of actions.

9.5 REFERENCES

Bartolo A, Cubelli R, Della Sala S, Drei S, Marchetti C (2001) Double dissociation between meaningful and meaningless gesture reproduction in apraxia. Cortex 37:696-699.

Bates E, Dick F (2002) Language, gesture, and the developing brain. Developmental Psychobiology 40:293-310.

Bekkering H, Wohlschläger A, Gattis M (2000) Imitation of gestures in children is goal-directed. The Quarterly Journal of Experimental Psychology 53A (1):153-164.

Brass M, Bekkering H, Prinz W (2001) Movement observation affects movement execution in a simple response task. Acta Psychologica 106:3-22.

Brass M, Bekkering H, Wohlscläger A, Prinz W (2000) Compatibility between observed and executed finger movement: comparing symbolic, spatial, and imitative cues. Brain and Cognition 44:124-143.

Bryden PJ (2000) Lateral preference, skilled behaviour and task complexity: hand and foot. In: Mandal MK, Bulman-Fleming WB, Tiwari G. Side Bias: a Neuropsychological Perspective. London: Kluwer Academic Press:225-248.

Carmo JC, Rumiati RI (2009) Imitation of transitive and intransitive actions in healthy individuals. Brain and Cognition 69:460-464.

Fagard J, Marks A (2000) Unimanual and bimanual tasks and the assessment of handedness in toddlers. Developmental Science 3(2):137-147.

Gallagher HL, Frith CD (2004) Dissociable neural pathways for the perception and recognition of expressive and instrumental gestures. Neuropsychologia 42(13) :1725-1736.

Gesell A, Ames LB (1947) The development of handedness. The Journal of Genetic Psychology 70:155-175.

Gleissner B; Meltzoff AN, Bekkering H (2000) Children’s coding of human action: cognitive factors influencing imitation in 3-year-old. Developmental Science 3(4):405-414.

Head H (1926) Aphasia and kindred disorders of speech. (Vol.1). Cambridge University Press: Cambridge.

Hopkins WD, Cantalupo C (2004) Handedness in chimpanzees (*Pan troglodytes*) is associated with asymmetries of the primary motor cortex but not with homologous language areas. Behavioral Neuroscience 118 (6):1176–1183.

Iwasaki S (2000) Age and generation trends in Handedness: an Eastern perspective. In: Mandal MK, Bulman-Fleming WB, Tiwari G. Side Bias: a Neuropsychological Perspective:83-100. Kluwer Academic Press, London.

Johnson-Frey SH (2004)The neural bases of complex tool use in humans. Trends in Cognitive Sciences 8 (2):71-78.

Koski L, Iacoboni M, Dubeau MC, Woods RP, Mazziotta JC (2003) Modulation of cortical activity during different imitative behaviors. Journal of Neurophysiology 89:460-471.

Mandal MK, Bulman-Fleming WB, Tiwari G (2000) Side Bias: a neuropsychological perspective. Kluwer Academic Press, London.

Meltzoff AN, Moore MK (1977) Imitation of facial and manual gestures by human neonates. Science 198:75-78.

O'Reilly AW (1995) Using representations: comprehension and production of actions with imagined objects. Child Development 66:999-1010.

Rumiati RI, Tessari A (2002) Imitation of novel and well-known actions. The role of short term memory. Experimental and Brain Research 1:425-433.

Schofield WN (1976) Do children find movements which cross the body midline difficult? Quartely Journal of Experimental Psychology 28:571-582.

Schofield WN (1976) Hand movements which cross the body midline: findings relating age differences to handedness. Perceptual and Motor Skills 42:643-646.

Tessari A, Canessa N, Ukmar M, Rumiati RI (2007) Neuropsychological evidence for a strategic control of multiple routes in imitation. Brain 130 (4):1111-1126.

Tessari A, Rumiati RI (2004) The strategic control of multiple routes in imitation of actions. Journal of Experimental Psychology: Human Perception and Performance 30:1107-1116.

Wapner S, Cirillo L (1968) Imitation of a model's hand movement: age changes in transposition of left-right relations. Child Development 39:887-894.

Wille AM (1996) Pratica dell'osservazione e della valutazione delle sindromi di maldestrezza del bambino. La terapia psicomotoria dei disturbi minori del movimento:111-151 Editore Marrapese, Roma.

9.5.1.1 List of actions

MF actions

1. “Hello” (with open hand)
2. Point someone
3. Show that something is good (finger on the cheek)
4. No
5. Silence
6. Hearing (hand rounding the ear)
7. “Sleep”
8. "Come here" moving the whole hand
9. Silent reproach (shaking slowly the open hand near the cheek)
10. Kiss
11. Comb hair
12. Shot with a toy gun
13. Brush teeth
14. Hammer a nail
15. Reaching something from the table
16. Drink from a glass
17. Cut with scissors
18. Dig

19. Phone

20. Eat with a spoon

21. Lock with a key

ML actions

1. Shake horizontally the hand with arm relaxed on the side
2. Point the middle of the forehead with forefinger
3. Same movement with finger under the chin
4. Shaking the hand with thumb and little finger rose. Hand turned downwards
5. The forefinger horizontally covers the eyes
6. Same position but hand on the chest
7. Hand on the back of the neck
8. Same gesture on the horizontal plan moving inwards
9. Shaking the fist near the cheek
10. Kiss with palm oriented outwards
11. Fist on the forehead
12. Rounding the hand
13. Hand on the mouth
14. Shaking horizontally the fist
15. Touch the object with the back of the hand

16. The thumb touches the forefinger forming a circle

17. Open and close the horizontal hand

18. Fist on the leg

19. Hand on one eye

20. Fist on the mouth

21. With extended arm, rounding the palm up and down two times

9.5.1.2 Order of presentation

Order n.				
1	Right MF	Right ML	Left MF	Left ML
2	Right MF	Right ML	Left ML	Left MF
3	Right ML	Right MF	Left ML	Left MF
4	Right ML	Right MF	Left MF	Left ML
5	Left MF	Left ML	Right MF	Right ML
6	Left ML	Left MF	Right MF	Right ML
7	Left ML	Left MF	Right ML	Right MF
8	Left MF	Left ML	Right ML	Right MF
9	Right MF	Left ML	Right ML	Left MF
10	Left ML	Right MF	Left MF	Right ML
11	Right MF	Left ML	Left MF	Right ML
12	Left ML	Right MF	Right ML	Left MF
13	Left MF	Right ML	Left ML	Right MF
14	Left MF	Right ML	Right MF	Left ML
15	Right ML	Left MF	Left ML	Right MF
16	Right ML	Left MF	Right MF	Left ML