IS SPEECH SPECIAL?
INSIGHTS FROM NEONATES AND NEUROIMAGING

by

ATHENA VOULOUMANOS

B.Sc. (Honours), McGill University, 1997

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(School of Medicine, Graduate Program in Neuroscience,
Department of Psychology)

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

January, 2004

Abstract

Language is a uniquely human adaptation that is hypothesised to require specialised anatomical substrates and dedicated processing mechanisms. Speech, the primary medium for language, is argued to rely on specialised substrates and processing as well. Yet, to date, evidence for the speech specialisation hypothesis has been equivocal. The four experiments in this thesis aim to advance the discussion in two ways. The differential processing of speech by humans was investigated through the use of functional neuroimaging tools (Experiment One), and through developmental studies of young infants' listening biases (Experiments Two-Four).

In Experiment One, functional neuroimaging tools are used to investigate the specificity of neural substrates recruited in detecting speech compared with closely matched non-speech controls. This study takes advantage of an event-related imaging design that provides a narrow window of observation for neural recruitment during individual stimulus events. The results of the first study demonstrate that adults activate specific neural substrates when detecting speech sounds, indicating that specialised substrates are involved from the early processing stages.

Experiments Two through Four take an ethological approach to speech specialisation and investigate whether young infants show a bias for listening to speech as compared to matched non-speech sounds. In Experiment Two, behavioural methods probe whether young infants of 2 to 7 months show listening preferences for speech compared with non-speech. Experiment Three seeks to establish the roots of a speech bias in the neonatal period. Finally, Experiment Four investigates the origin of the bias,

to determine whether the speech bias originates from prenatal experience, or is independent of specific experience. The results of these studies show that differential processing has its roots in early infancy, with infants demonstrating a preference for listening to speech from birth. The speech bias shown by neonates appears not to be based on specific experience with speech sounds, but instead is rooted in human biology. Human infants are prewired to preferentially attend to speech, granting speech a special status in relation to other sounds.

Table of Contents

List of Tables

List of Figures

Preface

The ideas presented in this thesis represent the work of the author, developed in discussion with her PhD advisor, Dr. J. Werker.

The functional MRI study described in Experiment One was conducted with the assistance of Drs. J. Werker, P. Liddle and K. Kiehl. The task was designed based on an oddball paradigm previously developed by Dr. K. Kiehl, and adapted for speech studies by Drs. K. Kiehl, and J. Werker, and the author. The data were collected and analysed by the author, with help from Dr. K. Kiehl. Participants were recruited by Dr. Liddle's imaging team, including T. Cairo, Dr. K. Laurens, and the author. This study was written by the author, in consultation with her collaborators, and published in *Journal of Cognitive Neuroscience* (full citation: Vouloumanos, A., Kiehl, K. A., Werker, J. F., and Liddle, P. F. [2001]. Detecting sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and non-speech. *Journal of Cognitive Neuroscience*, 13, 994-1005.).

The infant studies described in Experiment Two were designed, conducted and analysed by the author, in consultation with Dr. J. Werker. Participants were recruited by K. Corcoran, E. Job, A. Almas and C. Merkel. Reliability coding was provided by C. Merkel and F. Pons. The paper describing these studies was written by the author, in consultation with Dr. J. Werker, and is in press at *Developmental Science* (full citation: Vouloumanos, A., and Werker, J. F. [in press; accepted July 15, 2003]. Tuned to the signal: The privileged status of speech for young infants. *Developmental Science.*).

The neonate studies described in Experiments Three and Four were designed by the author and Dr. J. Werker. Neonates were recruited and tested by M. Bhatnagar, E. Moon, A. Valji, Dr. T. Burns and the author, and the data collected were analysed by the author. The paper reporting on these studies was written by the author, in consultation with Dr. J. Werker, and is under review.

Acknowledgements

And so it ends, sort of.

My supervisor Janet Werker took a chance one day in late July and made it all possible. Her guidance, support, and enthusiasm are inspirational. Geoff Hall, Peter Liddle, Doug Pulleyblank, and Richard Tees guided, challenged, and shaped.

I've had the honour of working with many extraordinary people. I thank them and wish them every happiness. The baby operation: Alisa Almas, Monika Bhatnagar, Tracey Burns, Kate Corcoran, Marisa Cruickshank, Suzanne Curtin, Hélène Deacon, Christiane Dietrich, Laurie Fais, Chris Fennell, Miguel Imperial, Marie Jetté, Liz Job, Karla Kaun, Nenagh Kemp, Rachel Lewis, Carla Merkel, Erin Moon, Michelle Patterson, Laura Sabourin, Rushen Shi, Ramesh Swamy, Whitney Weikum, and Katie Yoshida. Peter Liddle and Elton Ngan welcomed me into their world of neuroimaging. Todd Woodward, Sara Weinstein, Izabella Patyk, Kristin Laurens, Kent Kiehl, Miguel Imperial, Tara Cairo, and Cameron Anderson always made me feel like one of the crew. Chris Fennell, my compatriot and colleague, was there for both the beginning and the end, and the groaning in between. Kristin Laurens and Brandi Ormerod, my classmates, colleagues and friends, commiserated, toiled and triumphed.

My friends Tracey Burns, Kate Corcoran, Erin Putterman, and Dave Sagar opened their hearts, and many bottles of wine. George Hadjipavlou, Maegen Giltrow, Melissa Gillis, Flo, and the denizens of 2827 shared their happy home with me. Arash Nakhost and Claire Richardson are the best friends I could ever hope for.

Gary Marcus, my principal primate, helped me grow and learn and think, and loved me anyway. I can't imagine life without him.

My family—dad, mom, George, and Ari—you are everything. This thesis is dedicated to you.

Chapter 1. Introduction

1.1 Is speech special?

Language is considered a uniquely human adaptation (e.g., Lieberman, 1994; Pinker & Bloom, 1990) implemented by specialised processes (Chomsky, 1986) rooted in unique biological substrates (Lenneberg, 1967; Lieberman, 1984). Within the broader context of language, the perception of speech is argued to be an adaptation that elicits specialised processing (e.g., Liberman, 1996; Liberman, Harris, Hoffman, & Griffith, 1957; Trout, 2001). Though speech is not a necessary medium for language (signed and written language being obvious alternative media), perceiving speech is a necessary pre-requisite for understanding spoken language. Moreover, historically, speech may have been a primary motor driving the evolution for language (Lieberman, 1984) and is currently conceptualised as the "natural form of expression of a general linguistic disposition" (Remez, 1994, p. 158). As such, understanding the specialised processing that underlies the recognition and production of speech is essential.

In conceptualising the "special" status of speech for humans, speech perception has been described as a uniquely human linguistic adaptation. Speech elicits processing substrates (see section 1.2.1) and mechanisms (see section 1.2.2) that are different from other auditory events (e.g., Liberman, 1996; Liberman et al., 1957). This position will be referred to as the "domain-specific" perspective because it considers speech to be a unique domain of perception (and production). The competing conceptualisation considers speech perception to be subject to general auditory processing no different from

the perception of other sounds, requiring no specific hardware or mechanisms that may be unique to humans (Bregman, 1990; Diehl & Kluender, 1987; Kluender, 1994; Massaro, 1987). Because of their description of speech perception as general auditory processing, these accounts will be referred to as "domain-general auditorist" (cf., Trout, 2001). Massaro (1994) summarises these alternatives succinctly:

> A speech organ is necessary because speech is a highly specialized domain that necessarily requires a specialized processing system. In contrast, it might be hypothesized that understanding speech is just one domain of many that require discrimination, categorization and understanding. (p. 219-220)

1.2 Speech is special: traditional evidence

Evidence from traditional approaches builds a convincing case for the uniqueness of speech perception (for reviews see also Doupe & Kuhl, 1999; Liberman, 1996; Samuel & Tartter, 1986; Trout, 2001; Werker & Tees, 1992). Three broad classes of arguments for the special processing of speech will be reviewed. First, speech is suggested to be a unique domain of perception that undergoes unique kinds of processing. Second, speech processing is implemented by specific neural substrates in the left hemisphere. Third, specialised mechanisms and substrates are present in early infancy. A fourth implicit implication that follows from these points is that processing abilities and specialised neural substrates for speech are specific to humans (Lieberman, 1990; Trout, 2001).

1.2.1 Unique neural substrates for speech perception.

Distinct anatomical substrates in the left hemisphere show functional specificity for speech. Over a century of aphasia data demonstrate that lesions to specific areas of the left hemisphere can cause specific deficits in speaking or perceiving language (for reviews see Blumstein, 1994; Blumstein & Milberg, 2000; Damasio & Damasio, 2000). This basic network is supplemented by additional areas (predominantly in the left hemisphere) that vary between individuals (Ojemann, Ojemann, Lettich, & Berger, 1989). A left hemisphere advantage is observed when healthy adults are processing speech: behavioural studies demonstrate a robust right ear/left hemisphere advantage for speech compared with the left ear/right hemisphere advantage for music and other sounds (Kimura, 1967). The connectivity of the auditory system, in which contralateral connections dominate, presents a non-invasive behavioural assay for hemispheric dominance. When sounds are presented to the right ear, they are processed mainly by the left hemisphere, and vice-versa. Listeners are more accurate in reporting sounds presented to the right ear than to the left ear when the input is verbal. However, when music or environmental sounds are played, accuracy is greater for sounds presented to the left ear. These behavioural data are also supported by many neuroimaging studies which demonstrate greater activity in the left hemisphere for speech processing (reviewed in Binder & Price, 2001, and in section 2.2).

1.2.2 Unique perceptual organisation of speech stimuli.

Speech is recovered in situations that require sophisticated analysis. This sophisticated perception of speech sounds is captured by phenomena such as "trading relations" between the spectral and temporal cues of speech, in which listeners perceive the same category of speech sound (e.g., "ba") from signals with different acoustic properties (Repp, 1982; Underbakke, Polka, Gottfried, & Strange, 1988). Moreover, the context in which different acoustic properties appear affects the perceived category; for example, listeners hear either "ba" or "wa" depending on the length of the vowel following the consonant (Eimas & Miller, 1980; Miller & Liberman, 1979). Listeners can reconstitute speech sounds when their constituent frequencies are presented to separate ears. In these dichotic listening conditions, listeners demonstrate the ability to perceive a single stimulus concurrently as speech (e.g., a syllable) and non-speech (e.g., a tone or chirp), resulting in so-called duplex perception (Liberman, Isenberg, & Rakerd, 1981). Importantly, when the intensity of sounds is reduced, only speech is perceived, suggesting that not only are speech and auditory modes separable, but that speech perception is dominant (Whalen & Liberman, 1987). A comparable phenomenon of simultaneous modes of speech and non-speech perception is observed during the perception of sine-wave speech (Remez, Rubin, Pisoni, & Carrell, 1981), further suggesting that phonetic and auditory modes of perception are separable.

1.2.3 Early developmental roots

According to many domain-specific theories (e.g., the motor theory), the uniqueness of speech perception is based on processes that are innately specified, requiring only minimal experience to engage (Liberman & Mattingly, 1985). Finding evidence of speech phenomena and dedicated speech substrates in very young infants early in development would support the idea of innate origins, buttressing the argument for special processing of speech.

1.2.3.1 The perceptual abilities of young infants

Infants show remarkable perceptual sensitivity for speech that is evident in their precocious ability to discriminate and categorise many aspects of human language. For example, newborn infants are sensitive to word boundaries (Christophe, Dupoux, Bertoncini, & Mehler, 1994), distinguish between stress patterns of multisyllabic words (Sansavini, Bertoncini, & Giovanelli, 1997), categorically discriminate lexical versus grammatical words (Shi, Werker, & Morgan, 1999), and differentiate between good and poor syllable forms (Bertoncini & Mehler, 1981). Very young infants can distinguish between rhythmically dissimilar languages (Mehler et al., 1988; Nazzi, Bertoncini, & Mehler, 1998; Ramus, Hauser, Miller, Morris, & Mehler, 2000), and categorically discriminate between minimally different phonetic contrasts in their native (Bertoncini, Bijeljac-Babic, Blumstein, & Mehler, 1987; Dehaene-Lambertz & Baillet, 1998; Eimas, Siqueland, Jusczyk, & Vigorito, 1971) and non-native languages (Streeter, 1976; Trehub, 1976; Werker & Tees, 1984). These perceptual facilities for analysing properties of speech that are exploited by the phonological and syntactic systems of language suggest

that human infants have specialised speech mechanisms. However, even more convincing evidence comes from infant studies demonstrating the same kinds of sophisticated analyses and dedicated substrates seen in adults.

1.2.3.2 Sophisticated speech analysis

Like adults, infants can perform sophisticated analyses with speech sounds. Using a dichotic listening task with 3-4 month-old infants, Eimas and Miller (1992) showed that infants can integrate two sources of acoustic information played to different ears into a coherent phonetic percept. Sophisticated organisation for speech perception is also seen in the trading relations shown by infants for shifting boundaries for perceiving voice onset time differences when other voicing cues are manipulated (Eimas & Miller, 1980; Miller & Eimas, 1983). Moreover, infants show context effects similar to those of adults when perceiving category boundaries (Miller & Eimas, 1983). Speech is processed in a specialised way from early on in human development.

1.2.3.3 Neural specialisation for speech perception

Perhaps the most convincing evidence for a specialised perceptual mechanism comes from studies showing left hemisphere functional specialisation for speech processing. In dichotic listening tasks, newborns and young infants already possess some degree of right ear/left hemisphere advantage for speech discrimination (Bertoncini et al., 1989; Entus, 1977; but see Best, Hoffman, & Glanville, 1982). Newborns also show a rightward turning bias while listening to binaurally presented female speech, but only when information in the "speech range" is present; attenuating frequencies below 3500

Hz eliminates the turning asymmetry (Ecklund-Flores & Turkewitz, 1996). These behavioural demonstrations are complemented by neuroimaging studies eliciting greater activation in the left hemisphere for speech compared with backwards speech in 3 month-olds (Dehaene-Lambertz, Dehaene, & Hertz-Pannier, 2002) and in newborns (Peña et al., 2003). Electrophysiological measures reveal functional anatomical specialisation in newborns for speech in the left hemisphere in some studies (Dehaene-Lambertz & Baillet, 1998; Dehaene-Lambertz & Dehaene, 1994; Molfese & Molfese, 1979a; Molfese & Molfese, 1980), but not others (Cheour et al., 1998; Cheour et al., 1997; Novak, Kurtzberg, Kreuzer, & Vaughan Jr., 1989; reviewed in Werker & Vouloumanos, 2001). Most, but not all, results are in agreement that very young infants show the same cerebral lateralisation for aspects of speech perception that many adults show, as well as adults' facility in perceptual grouping of the speech signal.

1.3 Challenges to the traditional conception of "speech is special"

To review the discussion thus far, the evidence for speech perception as a special adaptation comes from a variety of sources: a) humans performing a constellation of sophisticated operations on the speech signal, b) the activation of unique and/or dedicated neural substrates, and c) the manifestation of these abilities early during human development. But many scholars remain unconvinced. Strong challenges have been motivated by theories of general auditory processing (e.g., Bregman, 1990) as well as experimental findings questioning the uniqueness of speech phenomena (e.g., Diehl & Kluender, 1989a).

In Bregman's (1990) theory of auditory scene analysis, any and all sound is analysed by low-level acoustic mechanisms, which evaluate frequency and amplitude of the signal according to a set of principles (e.g., proximity and similarity) in order to create a perceptual group. As such, these elementary grouping principles allow the listener to segregate multiple streams of information into individual voices (e.g., a speaker at a cocktail party). The organisation that falls out from these principles affects how listeners perceive all sounds, from music to speech. Citing contextual effects and the influence of multiple sources of information (e.g., auditory and visual) on our perception of speech, Massaro (1994) suggests that general abilities in cognition, perception, and learning underlie speech perception. In both Bregman's and Massaro's views, speech perception is similar to other forms of pattern recognition. Taking a different approach based on parallels between the way humans and non-humans appear to process speech, Kluender (1994) concludes that speech perception is a consequence of the structure of the (vertebrate) auditory system.

1.3.1 Special mechanisms and neural substrates seem neither speech-specific…

Evidence shows that apparent speech-specific phenomena such as categorical perception, context effects and trading relations may not be exclusive to speech. Adults (Pisoni, 1977) and infants (Jusczyk, Pisoni, Walley, & Murray, 1980; Simos & Molfese, 1997) show categorical discrimination when labelling non-speech sounds composed of two tones with different lags in onset asynchrony. Some of the context effects that seem central and special in phonetic perception, such as the distinction between "ba" (a stop

consonant produced by completely obstructing airflow with the lips and tongue, Ladefoged, 1993), and "wa" (a glide consonant in which the airflow obstruction is incomplete, Ladefoged, 1993), that depend on the length of the following vowel are also true of complex non-speech sinusoidal tones (Jusczyk, Pisoni, Reed, Fernald, & Myers, 1983; Pisoni, Carrell, & Gans, 1983). The phenomenon of duplex perception, perhaps one of the most convincing arguments in the "speech is special" arsenal, has also been demonstrated for non-speech sounds, such as slamming doors (Fowler & Rosenblum, 1990) and music (Pastore, Schmuckler, Rosenblum, & Szczesiul, 1983). Even trading relations are evident between spectral and temporal cues of non-speech sounds (Parker, Diehl, & Kluender, 1986). The parallels between the sophisticated analysis of speech and non-speech sounds suggests to domain-general auditorists (e.g., Bregman, 1990; Massaro, 1994) that basic auditory processes are recruited in the organisation of phonetic information.

1.3.2 ... nor substrate-specific

The case for neural specialisation for speech is weakened by evidence for bilateral recruitment of adult neural substrates during speech perception and inconsistent localisation in infants. Instead of localised left hemisphere activation, functional imaging studies sometimes reveal bilateral activation for speech and non-speech sounds (reviewed in 2.2, and in Binder & Price, 2001). Electrophysiological studies with developmental populations suggest that neural responses during the discrimination of particular speech contrasts are not necessarily confined to the left hemisphere; for example, infants tend to

show left hemisphere responses for changes in place of articulation (Dehaene-Lambertz & Baillet, 1998; Dehaene-Lambertz & Dehaene, 1994; Molfese, 1980; Molfese, Buhrke, & Wang, 1985; Molfese & Molfese, 1979a), but changes in voice onset time elicit either bilateral (Novak et al., 1989) or right hemisphere (Molfese & Molfese, 1979b) responses. Some studies recruit the left hemisphere indiscriminately as both speech sounds and complex tone arrays elicit left hemisphere activation in 4 month-olds, albeit with activation peaks in different electrodes (Dehaene-Lambertz, 2000).

1.3.3 … nor species-specific

Keen perceptual sensitivity to speech properties might not even be the exclusive domain of humans. Experiments with other species show that many animals appear just as sensitive to perceptual properties of speech. Animals including budgerigars (Dooling, Okanoya, & Brown, 1989), quail (Kluender, Diehl, & Killeen, 1987), chinchillas (Kuhl & Miller, 1975), and macaques (Kuhl & Padden, 1982) discriminate between minimally different phonemes in a categorical fashion, like humans. Rhesus monkeys and Japanese macaques experience context effects in a stop/glide distinction ("ba" vs. "wa") in the same way as humans (Stevens, Kuhl & Padden, 1988, as cited in Kluender, 1994). Arguably more complex discrimination abilities, such as distinguishing between rhythmically different languages which requires the coordination of multiple cues, are present in cotton-top tamarin monkeys (Ramus et al., 2000). Effects such as the perceptual magnet effect which seemed to describe uniquely human patterns of vowel discrimination at the time of discovery (Kuhl, 1991), with appropriate testing conditions

such as a pre-exposure, may not be human-specific (Pons, Trobalon, Bosch, & Sebastián-Gallés, under review). Though these shared abilities may reflect analogous processes and neural structures (resulting from independent evolution), it has been argued on the grounds of parsimony that they are more likely to result from homologous processes stemming from a common evolutionary source (Diehl & Kluender, 1989a, 1989b). These facts led Kluender (1994), to the conclusion that "much of the evidence for [humans'] abilities to discriminate speech sounds may fall out gracefully from general properties of auditory systems" (p. 189).

1.3.4 The traditional "speech is special" debate: some interim observations

The debate is far from over. It is clear from a number of traditional approaches already contributing to the debate that no one piece of data will conclusively end it. The advent of new brain imaging technologies allowing more precise monitoring of neural activation to individual events (Rosen, Buckner, & Dale, 1998), used in conjunction with careful non-speech controls could open a vital window onto the question of neural specificity for speech. In addition, since infants may be considered relatively naïve listeners of human language (compared with adults), they can provide additional sources of information about the perception of speech as a special sound. However, rather than considering the problem from the standpoint of traditional speech science, addressing the special status of speech from an ethological perspective can shed a different light on the problem (see also Jusczyk & Bertoncini, 1988).

1.4 Specialised signal processing: a perspective from ethology

The ethological tradition of studying animals in their natural habitat can offer a new perspective on the special status of speech for humans. "An animal's environment is rich in cues that can convey useful information. An early finding in ethology, however, showed that animals attend to only a relatively small subset of these cues – sign stimuli." (Ryan, Phelps, & Rand, 2001, p.143). In other words, some kinds of information are more salient than others. Gould and Marler (1987) suggest that through a process of innately guided learning, organisms have innate predispositions for specialised learning of particular kinds of information. Rather than prescribing the organism's behaviour, these biases would help orient organisms towards particular aspects of their environment.

From an evolutionary perspective, organisms may be biased to process particular information that increases their overall fitness and chance for survival. This preferred information subset includes biological signals that allow animals to recognise their conspecifics and kin. Examples of recognition signals, including olfactory, vocal and visual cues, abound. For example, olfactory cues allow mammals, including humans (Russell, 1976) and squirrel monkeys (Kaplan & Russell, 1974), to recognise their mother and kin as infants (reviewed in Porter, 1998). Since olfactory cues are readily available in the prenatal environment, they do not lend themselves easily to parcelling out predispositions compared with the learned preferences of young organisms. Unlike olfaction, cues in the visual and auditory modalities are more readily examined for particular predispositions for biologically significant information.

1.4.1 Visual preferences in chicks

Studies on the development of social preferences in chicks have been instrumental in developing models for understanding the interaction between predispositions and learned preferences (for review see Bolhuis & Honey, 1998). The famous studies on imprinting by Lorenz (1952) demonstrated one kind of preference that emerges in chicks based on their specific first post-hatching exposure to a moving visual stimulus. This process, involving a specific area in the chick's forebrain, allows the infant organism to acquire a preference based on exposure to a *specific* stimulus (Johnson, Bolhuis, & Horn, 1985; Johnson & Horn, 1986). So whether they were trained with a rotating red cylinder or with a stuffed jungle fowl, chicks prefer to approach the familiar stimulus in the short term. The other mechanism, independent of this brain area, directs the young organism to particular stimuli in the absence of any specific exposure. When chicks are given *no* specific visual training, they prefer to approach the stuffed fowl[1] (Johnson et al., 1985). Two processes are involved in setting the preferences of young chicks, a learned preference for objects to which chicks have specific exposure (Johnson, Bolhuis, & Horn, 1992) and a predisposition to approach objects with particular configurations (specifically, the configural information present in the head and shoulders area of animals, Johnson & Horn, 1988). In this sense, chicks are not equipped with a bias for a particular stimulus, but instead have a bias for a general description of a preferable configuration. This general predisposition suffices for chicks in their natural environment to orient themselves to important and useful stimuli.

1.4.2 Vocal communication biases

As for visual information, organisms are biased to attend to information for species-specific communication early in their development. The degree of description available in the auditory domain can be quite finely detailed. For example, green treefrogs and barking treefrogs can discriminate and recognise their respective species' calls despite significant similarities in call structure (Gerhardt, 1988). Abilities to discriminate and recognise signals from one's own species are often complemented by a preference for conspecific signals. In selecting mates, for example, female túngara frogs prefer the calls of males of their own species to the calls of other species (except for the synthesised calls of their immediate—hypothesised—ancestors, Ryan & Rand, 1995). When given equal exposure to conspecific and heterospecific calls, chicks and quail prefer to approach the calls of their own species (Park & Balaban, 1991). Elegant transplantation experiments, in which chicks implanted with quail neural tissue are induced to prefer quail calls, demonstrate that species-specific call predispositions are implemented at the neural level (Long, Kennedy, & Balaban, 2001). Many organisms show biases for conspecific communication signals even in the absence of prior exposure. Even before learning how to sing, white-crowned sparrows (Nelson, 2000; Whaling, Solis, Doupe, Soha, & Marler, 1997) and zebra finches raised in acoustic isolation (Braaten & Reynolds, 1999) prefer to listen to their own species' song. Songbirds reared in acoustic isolation produce extremely abnormal songs, which contain, however, recognisable species-typical features, suggesting that birds possess some innate description of the calls of their species (reviewed in Marler, 1997). Though there is clear

evidence for innate predispositions for listening to conspecific calls and song, these biases exist in tandem with the ability to learn from a broader class of stimuli. For example, though song and swamp sparrows prefer to learn their own song, in its absence they are able to learn songs of other species (Marler & Peters, 1977, 1989). During an early sensitive period, zebra finches prefer to learn conspecific song, though, like sparrows, they too are able to learn songs of other species (Boehner, 1990). In the perception of conspecific vocalisations, as in visual perception, innate biases for particular kinds of information coexist with powerful learning abilities.

1.5 Parallel biases in humans

There are striking similarities between the processing systems for information in the visual and auditory modalities of animals. In both modalities, animals show clear experience-independent predispositions that are refined by learning through specific exposure. Not all preferences are the same. For humans, speech and faces are the most complex meaningful species-specific signals in their respective modalities. Both stimuli are likely to engage a similar combination of experience-independent and experience-dependent biases. Tapping into the biases of human infants, for speech and faces among other stimuli, has been a complex but ultimately fruitful area of scientific inquiry.

1.5.1 Using selective attention to tap into infant biases: methodological considerations

Preference measures are notoriously thorny in infant studies and have been extensively investigated (e.g., Hunter & Ames, 1988; Najm-Briscoe, Thomas, & Overton,

2000; Pascalis & de Haan, 2003; Roder, Bushnell, & Sasseville, 2000; Rose, Gottfried, Melloy-Carminar, & Bridger, 1982; Wetherford & Cohen, 1973). However, preference measures are used in two very different ways with human infants. In one class of studies, infants are first familiarised or habituated with stimuli of one type, and then probed on their representations or their memory in a test phase that compares relatively familiar with relatively novel stimuli. In this type of study (e.g., Jusczyk & Aslin, 1995; Marcus, Vijayan, Bandi Rao, & Vishton, 1999; Saffran, Aslin, & Newport, 1996), both familiarity and novelty preferences constitute valid evidence. The familiarity or novelty preference shown by infants in a given study can be predicted with some success by models that take the age of the infant and the complexity of the stimuli or task into account (Cohen, DeLoache, & Rissman, 1975; Hunter & Ames, 1988). Studies exploring infant biases, however, belong to a second class in which infants are tested on their inherent preferences for particular stimuli over others, without any explicit prior exposure to the study items (e.g., Cooper & Aslin, 1990; Fernald, 1985; Moon, Cooper, & Fifer, 1993; Ward & Cooper, 1999; Werker & McLeod, 1989, and many others). This latter method provides a measure of the intrinsic attention-holding power of an auditory stimulus for an infant of a particular age in particular comparison conditions (see Cohen, 1972).

Selective preference has emerged as a powerful means with which to address questions about infant biases. When infants are presented with a choice of two sounds or visual displays without familiarisation immediately prior to testing, and they are allowed to control the presentation of the stimuli, the favoured stimulus can be considered the "preferred" stimulus. A preference, as revealed by infants' selective behaviour, is an

indicator of privileged processing. By extrapolating from the test situation in the laboratory to the infant's natural environment, we may make inferences about which objects and sounds in the world are salient for the infant.

1.5.2 Human face perception

Like other animals, human infants demonstrate biases for their conspecifics and kin. The exploration of infant biases in face perception has developed ahead of the study of speech biases. From this work, it has become apparent to ethologists and developmentalists that the visual world of young humans is mediated through biases similar to those of chicks (e.g., Morton & Johnson, 1991). In particular, human infants seem to show both experience-independent predispositions and experience-based preferences. Minutes after their birth, neonates preferentially orient towards and look at face-like stimuli compared with non-faces (Goren, Sarty, & Wu, 1975; Johnson, Dziurawiec, Ellis, & Morton, 1991; Kleiner, 1987; Maurer & Young, 1983; Valenza, Simion, Cassia, & Umilta, 1996). In the most comprehensive account of these data, Morton and Johnson (1991) proposed that the structural information in face-like stimuli—i.e., the configuration of crude facial features such as the eyes and mouth—provides the preferred information (see also Johnson et al., 1991; Simion, Cassia, Turati, & Valenza, 2001; Valenza et al., 1996). The least domain-specific account of these data proposes that stimulus energy (amplitude spectrum) is the relevant dimension and that newborn preferences can be predicted based on their visual system's contrast sensitivity (Easterbrook, Kisilevsky, Hains, & Muir, 1999; Kleiner, 1987). But

even if sensory information were all that mattered, the basic conclusion remains the same: newborns without prior experience are biased to orient towards faces. This initial predisposition becomes refined with experience: within hours of birth, infants can discriminate and prefer their mother's face to a stranger's face (Bushnell, Sai, & Mullin, 1989; Walton, Bower, & Bower, 1992). The initial bias for faces is complemented by (or facilitates, Morton & Johnson, 1991) experienced-based preferences for specific face stimuli.

1.5.3 Interactions between predispositions and learning: lessons from face perception

Very young infants show evidence for face preferences of two different types: innate predispositions encourage infants to track face-like stimuli, and experience-based preferences guide them towards particular faces. In chicks, innate predispositions and experience-based imprinting are implemented by independent neural systems. In humans, the neural systems involved are under current debate (Cassia, Simion, & Umilta, 2001; Morton & Johnson, 1991). Morton and Johnson (1991) propose that subcortical brain systems (CONSPEC) orient the attention of chicks and humans infants to conspecific information, in particular to structural face-like information, which is then available to learning mechanisms (CONLERN). Recently, a connectionist network has been developed to provide a mechanistic account for experience-independent face preferences (Bednar & Miikkulainen, 2000). Though this connectionist model questions the need for postulating two separate mechanisms to account for innate versus learned preferences, it successfully demonstrates how internally-generated patterns of activity

based on a (prespecified) template could organise the visual system prior to external experience to produce innate preferences for particular configurations.

1.5.4 Human speech perception

The concept of innately guided learning has been explicitly applied to speech perception and language acquisition to suggest that human infants may have a bias for listening to speech (Gould & Marler, 1987; Jusczyk, 1997; Jusczyk & Bertoncini, 1988) . This bias may be prewired and specific (Gould & Marler, 1987), or it may depend to some extent on prenatal experience (Jusczyk, 1997). In either case, a bias for speech would orient infant humans towards important auditory information in their environment.

1.5.4.1 Evidence for specific speech preferences in young infants

Neonates already demonstrate preferences for particular sounds in their midst. Human newborns prefer their mother's voice compared to the voices of other women (DeCasper & Fifer, 1980; Mehler, Bertoncini, & Barriere, 1978) even without any postnatal experience (Querleu et al., 1984). Newborns also prefer hearing their native language compared with other languages (Mehler et al., 1988; Moon et al., 1993; Nazzi et al., 1998). As infants have more experience with language, they develop detailed language-specific preferences. For example, by 10 months, English infants prefer strong-weak stress patterns most common to English speech (Jusczyk, Cutler, & Redanz, 1993) and prefer English phonotactic sequences (Jusczyk, Friederici, Wessels, & Svenkerud, 1993). Human infants are thus sensitive and attentive to parameters of the speech signal. The more experience with speech infants have, the more fine-grained their preferences

for the patterns that are specific to their native language. This selective attention to functional aspects of speech might be fundamental for tuning to the native language between 8 - 10 months of age (Jusczyk, Cutler et al., 1993; Werker & Tees, 1984).

Much evidence points to the rapid development of experience-based preferences for many aspects of speech perception. Aspects of heard speech are therefore encoded and remembered by young infants. A more fundamental question, however, is whether infants have a general predisposition for listening to speech compared with other sounds.

1.5.4.2 Evidence for a general predisposition for speech in young infants

Early studies found that newborn infants move in synchrony with speech but not non-human sounds (Condon & Sander, 1974; Kato et al., 1983), suggesting a powerful role for speech in moulding infant behaviour. However, subsequent analyses using more rigorous methodology failed to replicate this finding (Dowd & Tronick, 1986). Newborns were shown to orient towards isolated speech sounds ('bébé') compared with silence (Alegria & Noirot, 1978), though infants also demonstrate orienting behaviour to non-speech sounds such as rattle noises (Muir & Field, 1979). Speech and non-speech sounds were never directly compared in these orienting tasks, and thus these studies provide no evidence that infants listen to speech sounds any differently than to other sounds. A couple of methodologically-oriented studies compared the degree to which speech and a non-speech sound held the attention of 4-month-olds (Colombo & Bundy, 1981) and newborns (Butterfield & Siperstein, 1970). Though speech appeared favoured, serious methodological considerations bring these conclusions into doubt. The non-speech control used in both studies was white noise, a spectrally-rich but temporally-

impoverished sound. In addition, the engaging speech condition in the newborn study consisted of folk music (with both vocal and musical components) which cannot be considered a valid "speech" stimulus (Butterfield & Siperstein, 1970). Moreover, the prenatal experience of infants was never considered. The bias of young infants for speech has therefore not been established. One possibility is that sounds attract infants' attention indiscriminately, and then only later, through specific experience with particular sounds, are more detailed preferences established. The other possibility is that the preference for speech is specific, fundamental and prewired.

1.6 Rationale for the thesis

Speech and faces are two of the most complex natural stimuli in the human environment. Constructs useful for understanding the predispositions and learned preferences that inform infants' parsing of faces in their environment could be applied to their perception of speech. Infants may have general unlearned predispositions for listening to speech, just like they have predispositions for orienting towards faces. While research on face perception has succeeded in disentangling the effects of visual experience from infant's innate predispositions, research on speech perception, thus far, has not. This thesis describes four experiments that investigate the predispositions of humans for listening to speech and the specificity of neural substrates recruited in speech detection. In Experiment One, functional neuroimaging tools are used to describe the substrates that are recruited by adults' detection of a speech sound compared with carefully matched non-speech controls that provide a stringent test for the specificity of

neural substrates for speech detection. In Experiment Two, behavioural methods probe whether young infants of 2 to 7 months prefer listening to speech compared with non-speech sounds. Such a preference in early infancy would suggest early deployment of a speech bias but wouldn't address the origins of the bias. In the framework explored in this thesis, the particular operations that speech undergoes in the human brain might be based on experience, but the initial bias for preferentially listening to speech might be specified in the human biological endowment, and independent of specific experience. These possibilities are explored in the two final experiments. Experiment Three investigates the presence of a speech bias in the neonatal period. Finally, Experiment Four determines whether prenatal experience can account for the speech bias, in order to reveal its origins.

Chapter 2. Differential neural activation for speech and non-speech

2.1 Preface

Traditionally, speech perception has been argued to elicit the activation of specialised substrates in the left hemisphere. Evidence from dichotic listening tasks (Kimura, 1967), and studies in aphasia (reviewed in Blumstein, 1994; Damasio & Damasio, 2000) suggest that specific substrates of the left hemisphere are involved in particular language functions. However, functional neuroimaging studies on healthy participants in the past decade have proved inconclusive. Inconsistent results are most likely due to poor non-speech controls that preserve only some characteristics of speech and to the use of blocked designs that record average brain activity, rather than the brain's proximate response to individual stimuli (e.g., Binder, Frost, Hammeke, Rao, & Cox, 1996; Binder, Rao, Hammeke, Yetkin et al., 1994; Celsis et al., 1999; Fiez et al., 1995; Mummery, Ashburner, Scott, & Wise, 1999; Price et al., 1996; Zatorre, Evans, Meyer, & Gjedde, 1992). As a result, the functional specificity of neural substrates involved in language processing is still in question.

The following experiment addresses some of the weaknesses of previous studies by comparing neural activation elicited by the detection of speech sounds with the detection of non-speech sounds that preserve the characteristic frequency and timing information of speech. Brain activation was imaged using an event-related design, in which the haemodynamic response for each individual sound is recorded, opening a more

precise window on neural recruitment. This study is reprinted, with permission from MIT Press, from the Journal of Cognitive Neuroscience (13 [7], 994-1005).

2.2 Introduction

In the acoustic chaos of the external world, one sound to which humans attend effortlessly and automatically is spoken language. Do speech signals trigger different neural processors than do other environmental sounds? In this study we address this question by investigating the neural substrates mediating the initial processing of speech. We examine cortical activation during listeners' detection of rare stimuli in a sound sequence by comparing the detection of a speech stimulus to that of carefully matched non-speech stimuli.

Processing auditory input as speech is a crucial first step before further linguistic analysis (e.g., phonetic, semantic, or syntactic) can be performed. One class of theory holds that speech, like all sounds, is initially processed by general psychoacoustic mechanisms (e.g., Cole & Jakimik, 1980). Another class of theories claims that language processing involves specialised linguistic mechanisms (e.g., Chomsky, 1986; Chomsky, 2000). According to strong modularity theories, speech is instantaneously shunted into a different processing pathway than other acoustic stimuli (e.g., Fodor, 1983; Liberman, 1996). Evidence from duplex studies supports this hypothesis: When the steady state vocalic base of a syllable is presented to one ear, and rapidly changing formant transitions are presented to the other, subjects simultaneously hear speech (integrating the transitions with the base to perceive a consonant-vowel syllable) and non-speech (the

formant transitions alone sound like chirps, Whalen & Liberman, 1987). Duplex perception therefore suggests that different processors act in parallel on the same auditory input to perform distinct, but simultaneous, operations. This raises the possibility that unique neural substrates mediate these parallel processes.

Despite extensive studies using a number of different approaches, the neural substrates involved in language processing are incompletely identified. As early as last century, studies on aphasia revealed a functional asymmetry in that language processing relies preferentially on the left hemisphere. Divided along a crude dichotomy, receptive language depends on the posterior area of the superior temporal gyrus, or Wernicke's area, while productive language relies on the inferior frontal lobe, or Broca's area (for a review see Damasio & Geschwind, 1984). This functional asymmetry is complemented by a structural asymmetry favoring the left hemisphere, predominantly that of temporal regions associated with language functions (Galaburda, Sanides, & Geschwind, 1978; Geschwind & Levitsky, 1968). Recent neuroimaging studies on language processing have suggested a more distributed cortical network whose extent and pattern of activation vary across different studies (for a relevant discussion on the variability of substrates activated during phonetic processing see Demonet, Fiez, Paulesu, Petersen, & Zatorre, 1996; Poeppel, 1996). There is still disagreement, however, about the *precise* neural substrates that perform specific linguistic operations. This arises in part from the limited resolution of neuroimaging techniques, and in part from the inherent difficulty in isolating one or another aspect of the linguistic process (e.g., semantic, syntactic,

phonetic) from the other mental operations, linguistic and non-linguistic (e.g., attention, memory), that are simultaneously performed.

The processing of sound as speech is fundamental to all levels of analysis of spoken language. In contrast to higher-order linguistic operations, this perceptual step is more easily isolated through careful manipulation of the physical properties of the signal. By honing in on the functional neuroanatomy of speech perception, neuroimaging studies provide evidence for the contribution of different neural substrates in the steady state processing of speech. Neuroimaging studies comparing speech to simple non-speech foils such as noise bursts (Binder et al., 2000; Binder, Rao, Hammeke, Yetkin et al., 1994; Zatorre et al., 1992) or pure tones in passive listening (Binder et al., 2000; Binder et al., 1996; Celsis et al., 1999; Fiez et al., 1995), and active decision-making tasks (Binder et al., 1996; Demonet et al., 1992; Fiez et al., 1995) generally reveal bilateral activation to both speech and non-speech in the superior temporal gyrus (STG), with some areas in the left STG significantly more activated by speech stimuli (Binder et al., 1996; Demonet et al., 1992; Zatorre et al., 1992). The difference in activation between speech and non-speech seems to be somewhat more pronounced in the active tasks (Binder et al., 1996; Fiez et al., 1995).

However, tones and white noise bursts are unlike speech on many spectral and temporal dimensions and, as a result, they are arguably imperfect non-speech controls. In more recent studies researchers have used non-speech foils that are more closely matched in terms of the temporal and spectral properties that characterise speech. One approach has been to use reversed speech, which is acoustically matched in terms of duration,

amplitude, and spectral properties, but lacks the distinct temporal attributes of speech. Here, both isolated word tokens (Binder et al., 2000; Hickok, Love, Swinney, Wong, & Buxton, 1997; Price et al., 1996), and entire reversed sentences (Wong, Miyamoto, Pisoni, Sehgal, & Hutchins, 1999) have been contrasted. Another approach has been to use signal-correlated noise, which preserves the amplitude envelope, tempo, rhythm and syllabicity of speech, but lacks spectral information (Mummery et al., 1999). One recent study has compared speech to musical non-speech counterparts matched in duration and amplitude of systematically varying complexity (Benson et al., 2001). Results from these studies using more closely matched stimuli are more heterogeneous, with authors reporting speech-specific activation in the left superior temporal sulcus (STS) either posteriorly (Benson et al., 2001; Mummery et al., 1999) or anteriorly (Price et al., 1996), in the left posterior middle temporal gyrus (MTG, Price et al., 1996), bilaterally in the ventral STS and STG (Binder et al., 2000; Mummery et al., 1999) and the right supramarginal gyrus (Wong et al., 1999).

The heterogeneity of these findings might stem from the fact that the non-speech foils were matched to speech *either* spectrally *or* temporally, but not both. As of yet, no study has compared speech with non-speech controls that are matched in *both* spectral and temporal attributes, despite the fact that these dimensions together convey the identifying characteristics of natural speech. To contrast speech with a non-speech stimulus that preserves many of the spectral and temporal characteristics of speech without actually sounding like speech, we created complex sine-wave analogues. These sine-wave analogues consist of time-varying sinusoidal waves that track the resonant

center frequencies of natural speech and reproduce the changes in these frequency peaks across time. While eliminating characteristics of the voicing source, the broader band formant information and parts of the harmonic spectrum, these analogues preserve the critical frequency and temporal information of speech. The overall pattern of change in these energy peaks resembles the resonance changes produced by the human vocal tract when articulating speech (Remez, Rubin, & Pisoni, 1983).

Sine wave analogues are the ambiguous figures of the speech world. The defining characteristics of speech are so well preserved in these analogues that human listeners can be led into perceiving them as either speech or non-speech. In a classic series of speech perception studies, Remez and his colleagues demonstrated that sine wave analogues of *continuous* speech could be perceived as speech (Remez et al., 1981), allowing listeners to recover the message. Yet, this perception depends critically upon the listener's expectations; listeners who were not instructed to expect speech rarely heard the analogues as such (Remez et al., 1983). More recent studies show that sine-wave analogues of *isolated* words are even less likely to be heard as speech (Remez, Pardo, Piorkowski, & Rubin, 2001).

As non-speech counterparts, our use of sine-wave analogues of isolated nonsense words is therefore ideal: While the fidelity of the sine-wave analogues to the speech signal is such that analogues can be processed as speech under certain experimental conditions, by presenting analogues of *nonsense* words in isolation to naïve listeners, we ensured that our analogues were not perceived as speech.[2]

To date, most previous studies have imaged brain function during speech perception tasks using blocked design tasks (for notable exceptions see Hickok et al., 1997). The block design is effective at describing differences in steady state processing of speech versus non-speech. To observe neural activation in response to the detection of a stimulus in the auditory stream, the event-related design is more appropriate (for a discussion of the relative merits of using event-related designs see D'Esposito, Zarahn, & Aguirre, 1999; Stevens & Schwartzreich, 2000) and has been applied to auditory tasks (e.g., Belin, Zatorre, Hoge, Evans, & Pike, 1999; Hickok et al., 1997). This type of presentation allows modeling of the haemodynamic response to each individual stimulus presentation, thus offering a smaller and more precise window into the initial processing of speech. Moreover, it reduces the effects of possible confounds such as habituation or anticipation (Dale, 1999; Rosen et al., 1998).

In this study, we used an event-related fMRI design to investigate the neural substrates activated when a listener detects speech in comparison to a spectrally and temporally matched non-speech stimulus (see Figure 1). A secondary comparison contrasted cortical activation elicited by these complex stimuli with that elicited by simple tones. Because this task only requires listeners to indicate when they detect a stimulus that is different from the background, it does not require in-depth analysis of the signal. Differences in the patterns of neural activation to speech versus complex non-speech analogues should thus reflect the operation of the distinct processors activated during listeners' initial processing of the different stimuli.

2.3 Methods

2.3.1 Participants

Fifteen healthy right-handed adult volunteers (6 females, 9 males, mean age 25.5) participated in the study (handedness assessed as per Annett, 1967). Participants provided written informed consent and were screened for MRI compatibility before entry into the scanning room. All experimental procedures met with University ethical approval.

2.3.2 Procedure

Sounds were presented through insert earphones embedded within 30 dB sound attenuating MR compatible headphones using custom presentation software (http://nilab.psychiatry.ubc.ca/vapp). Because of the difficulty in accurately measuring *absolute* intensity values at the exit point from insert earphones, all sounds, speech, complex non-speech and tones, were equated for intensity *relative* to each other. Moreover, sounds were clearly audible above the noise of the scanner, as evidenced by listeners' near-perfect performance. Participants heard two stimulus runs, an 'experimental' run and a 'tone' run. The background non-target stimulus in both runs was a 1000 Hz tone, occurring with a probability of .8. The stimuli of interest were presented in a pseudorandom oddball design (separated by 3-5 non-target stimuli). In the 'experimental' run, the infrequent sounds consisted of speech (.1) and complex non-speech (.1). In the 'tone' run, 1500 Hz high tones (.1) and 500 Hz low tones (.1) were the infrequent sounds. Each run was 12.5 min, with a 2 s stimulus onset asynchrony (SOA)

for a total of 380 total stimuli per run. Order of presentation of the stimulus runs was counterbalanced across participants. Participants made a motor response on an MRI-compatible fiber-optic response device (Lightwave Medical, Vancouver, BC) using their left index finger for every infrequent sound they heard. Reaction times were monitored on-line.

2.3.3 Stimuli

The stimuli of interest were of four different types: (a) speech; (b) complex non-speech; (c) high tones; and (d) low tones. A) Speech stimuli consisted of six tokens of a monosyllabic nonsense word ("lif"[3]) spoken by a native female English speaker. Tokens varied in intonational contour (average minimum and maximum pitch: 202 Hz and 350 Hz respectively), and in duration (525 - 711 ms). B) Complex non-speech stimuli consisted of time varying sine-wave analogues of the speech tokens in which all regions of significant energy were tracked (namely the fundamental frequency and the first three formants; see Figure 1). Sinusoidal waves tracking these energy peaks were created individually in Mathcad 3.1 (Mathsoft Inc., Cambridge, MA). Fundamental frequency (corresponding to pitch) was tracked individually for each of the six speech tokens. Because the first three formants were virtually identical across the multiple natural repetitions, one set of formants from a representative word token was tracked. This set was composed of the first formant of the initial consonant segment ("l"), and the first three formants of the vocalic segment ("i"). The sine analogue to ("f") was created using a white noise generator and filtered. This representative set was then added onto the sine

wave analogue of the pitch contour of each segment using Signalyze 3.12 (Agora Language Marketplace, Charlestown, MA) to create six different stimuli. Analogues thus retained the duration, pitch contour, amplitude envelope, relative formant amplitude, and relative intensity of their speech counterparts (see Figure 1). C) Low tones were six pure sinusoidal waves of 500 Hz generated using Sound Edit Pro, version 2 (Macromedia Inc, San Francisco, CA) matched in duration to the speech stimuli. D) High tones were six pure sinusoidal waves of 1500 Hz generated using Sound Edit Pro, version 2 (Macromedia Inc, San Francisco, CA) matched in duration to the speech stimuli.

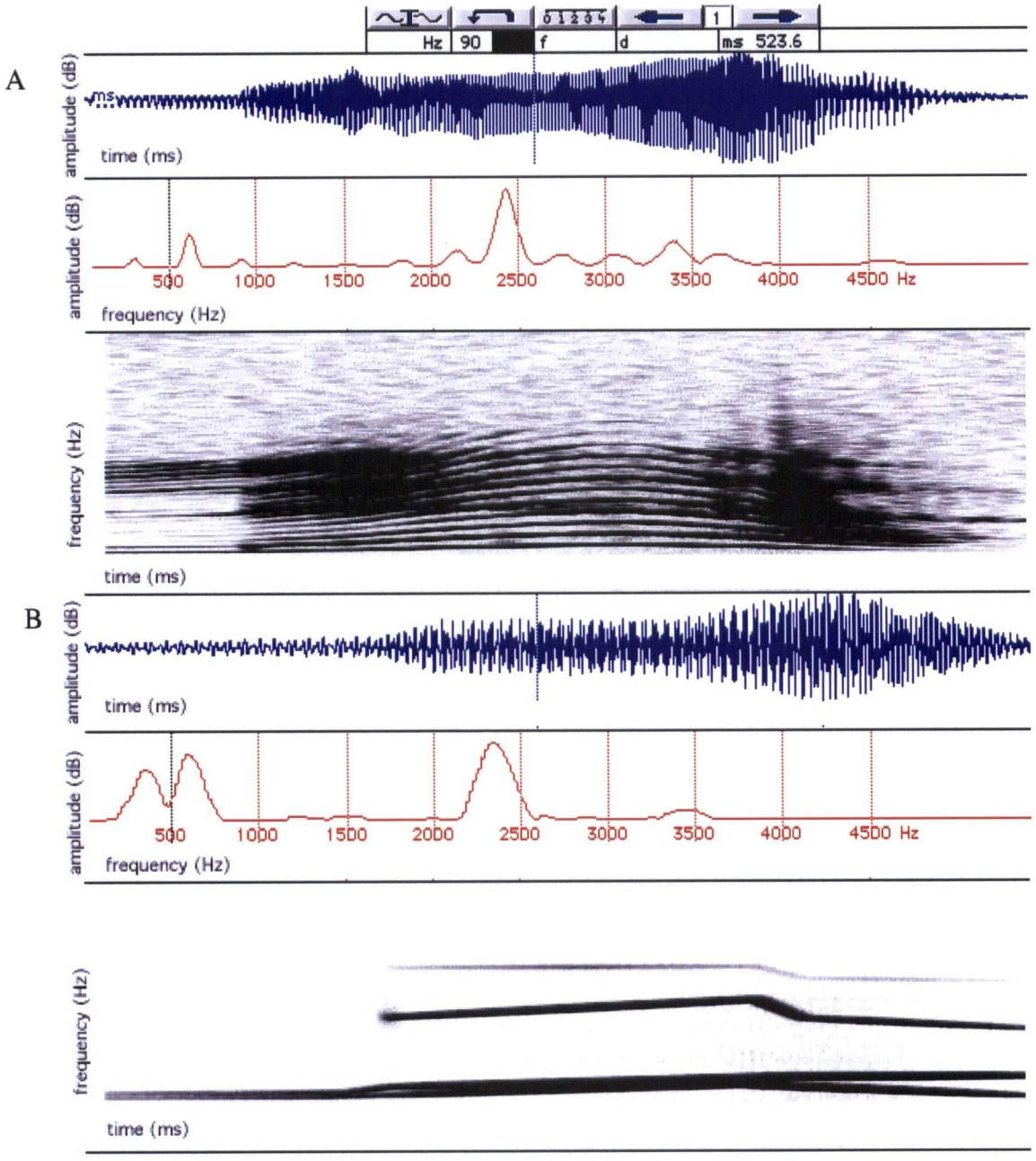

Figure 1. Speech (A) and complex non-speech (B) stimuli.

Similarities between the two types of stimuli are illustrated in waveform diagrams (blue), spectra depicting the relative amplitudes of different frequencies (red), and spectrograms showing changes in frequency across time (black).

2.3.4 Imaging parameters

Echo-planar images (EPI) were collected on a standard clinical GE 1.5 T system fitted with a Horizon Echo-speed upgrade. Conventional spin-echo T_1 weighted sagittal localisers were used to view the positioning of the participant's head and to graphically prescribe the functional image volumes. Functional image volumes were collected with a gradient echo (GRE) sequence (TR/TE 3000/40 ms, 90° flip angle, FOV 24 x 24 cm, 64 x 64 matrix, 62.5 kHz bandwidth, 3.75 x 3.75 mm in plane resolution, 5.00 mm slice thickness, 29 slices, 145 mm total brain coverage). This sequence is sensitive to the blood oxygen-level dependent (BOLD) contrast (Ogawa, Lee, Kay, & Tank, 1990). Each stimulus run consisted of 246 boldscans (full-brain scans). The first 12 s collected at the beginning of each run were discarded from the analyses, to avoid the T_1 saturation effects that occur in the early scans.

2.3.5 Image processing

Functional images were reconstructed offline. Statistical parametric mapping software (SPM99, Wellcome Department of Cognitive Neurology, London, UK) was used for image realignment and normalisation into modified Talairach stereotaxic anatomical space (using affine and non-linear components, as implemented in SPM99). Images were smoothed using a Gaussian kernel (8 mm FWHM) to compensate for intersubject anatomical differences, and to optimise the signal-to-noise ratio. Event-related responses to the stimuli of interest were modelled using a synthetic haemodynamic response composed of two gamma functions and their temporal

derivatives (for a discussion of the relative advantages and disadvantages of this modeling method, see (Kiehl, Laurens, Duty, Forster, & Liddle, 2001). The peak of the response was modeled at 6 s poststimulus time, consistent with the results of other event-related fMRI studies (Hickok et al., 1997). A high pass filter (cutoff period 89 s) was incorporated into the model to remove noise associated with low frequency confounds. A low pass filter (at the Nyquist frequency, with a period of 6 s) was also applied to remove noise associated with alternations of the applied radio frequency field. Three contrasts were used to create SPM{t} maps, later transformed into SPM{Z} maps, for three comparisons of interest: (a) activation for speech sounds relative to the complex non-speech stimuli, (b) activation for the speech sounds relative to simple tones, and (c) activation for the complex non-speech sounds relative to the simple tones.

2.3.6 Statistical analyses

Statistical analyses were performed in SPM99 using a fixed-effects model. Because multiple voxels were examined, a correction for multiple comparisons based on the theory of Gaussian fields was employed. The areas of activation reported are significant at the voxel level, with z-scores greater than 4.63, corresponding to a corrected significance level of $p < .05$. We further explored hemispheric differences in activation by comparing suprathreshold voxels in each hemisphere for individual listeners. Within the SPM program, we imposed a mask of the middle temporal gyrus on each listener's SPM{t} map for the main comparison of interest (see [a] above) . A custom script was used to extract suprathreshold voxels ($z = 2.63$, $p < .05$ uncorrected) in

the left and right hemispheres of every listener. A paired t-test was conducted on these hemispheric voxel counts to obtain an index of hemispheric asymmetry.

The standard fixed effects model of analysis was used to analyze patterns of activation within subjects because of the greater power it affords us in detecting details of the activation patterns. In order to demonstrate that our main findings can be generalised to the population, we performed an exploratory analysis using a random-effects model in SPM99 on the main comparison of interest, that of speech sounds relative to complex non-speech stimuli. We consider this analysis exploratory because the sample size in this study (n = 15), though larger than that of most neuroimaging studies, does not allow an adequate level of power to perform a full-fledged random-effects analysis. Images analyzed using this model were smoothed with a 14mm FWHM Gaussian filter. The areas of activation reported using this analysis are significant at the voxel level, with z-scores greater than 4.63, corresponding to a corrected significance level of $p < .05$.

2.4 Results

2.4.1 Behavioural performance

Accuracy for speech and non-speech detection was near perfect, with participants attaining an average score of 99.6% correct motor responses for speech, and an average score of 99.8% correct for non-speech. Participants' performance accuracy in the tone detection run was equally high with an average score of 99.6% for high tones, and 100% for low tones.

Participants responded fastest to speech (M = 314 ms, SE = 10), slower to high (M = 326 ms, SE = 21) and low tones (M = 328 ms, SE = 18), and slowest to complex non-speech (M = 337 ms, SE = 13). A series of paired two-tailed t-tests revealed that the only significant difference was between speech and complex non-speech ($t(14) = 2.71$, $p <$ .02). No other comparisons were significant.

2.4.2 Cortical activation

Speech versus complex non-speech. Speech stimuli elicited significantly greater activation than complex non-speech stimuli in the superior and middle temporal gyri (see Figure 2, Table 1ai). The individual haemodynamic responses confirm that activation was present in all listeners (illustrated in Figure 3). The differential activation of the MTG was bilateral, but the extent of differential activation was greater in the left hemisphere (45 versus 24 voxels exceeding a cluster-level height threshold of $z = 4.63$, p < .05 corrected).

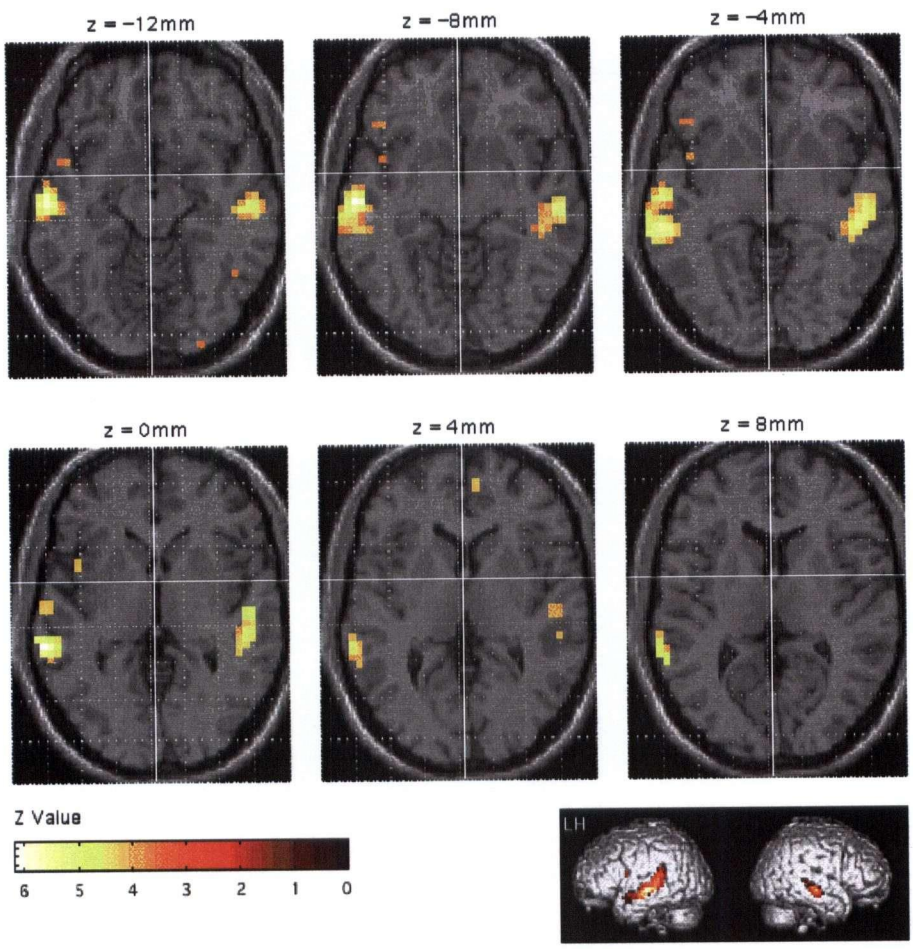

Figure 2. Cortical activation for speech versus complex non-speech.

Axial images illustrating cortical activation to speech relative to complex non-speech are shown at 4 mm intervals (fixed-effects analysis; display threshold $z = 3.72$; left hemisphere is on the left; colour bar indicates corresponding z-score). (Insert. Cortical surface rendering of areas activated for speech versus complex non-speech using a random-effects analysis [display threshold $z = 3.72$; colour bar indicates corresponding z-score]).

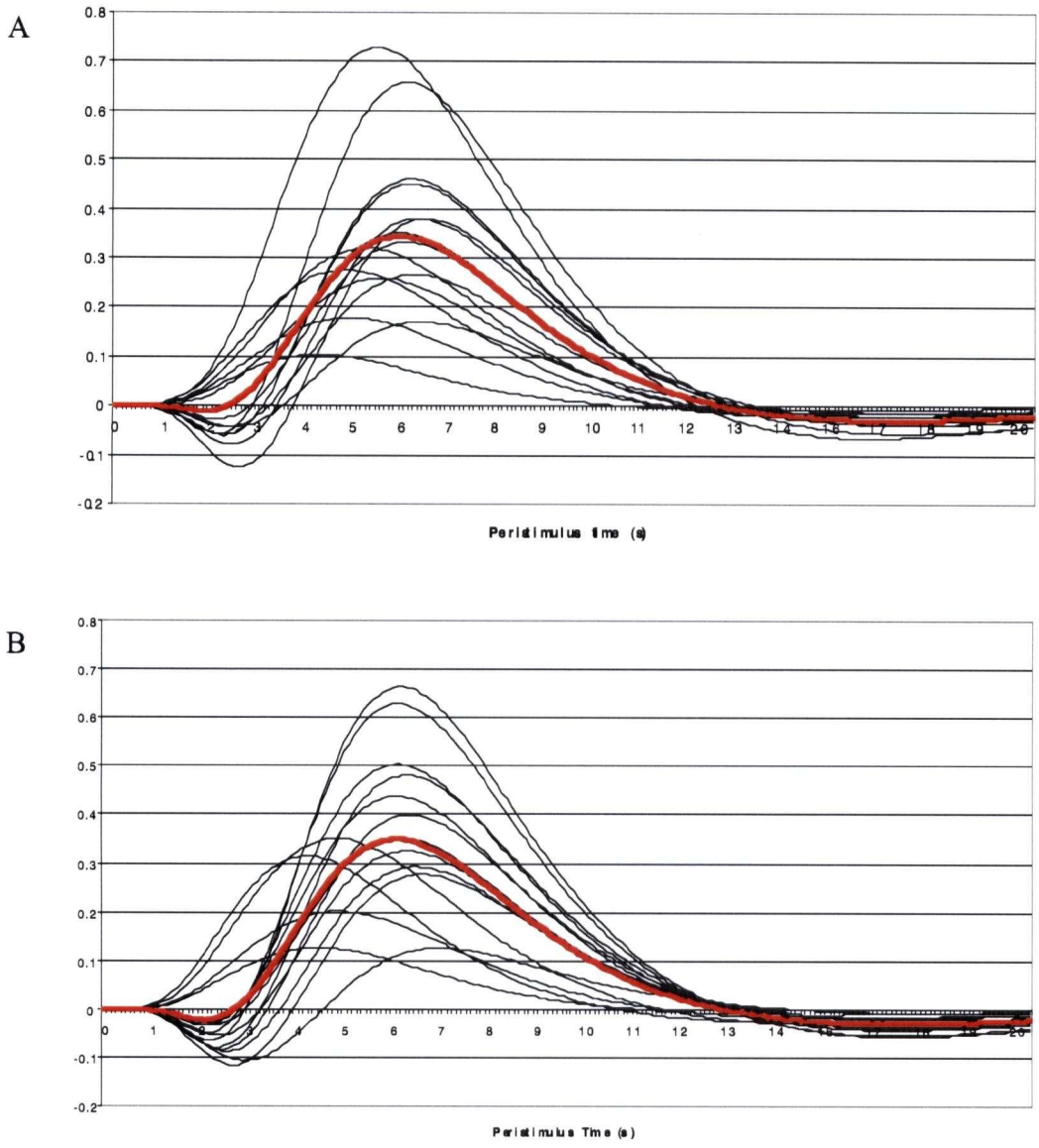

Figure 3. Modelled haemodynamic responses of individual listeners.

Haemodynamic responses for speech stimuli for individual listeners from a voxel of peak activation in: (A) Left middle temporal gyrus (Talairach co-ordinates: -60, -16, -8); and (B) Left posterior superior temporal gyrus (Talairach co-ordinates: -64, -44, 12). Haemodynamic responses are plotted in arbitrary units (mean responses are in red).

A one-tailed paired- t-test run with the individual contrasts within the omnibus fixed-effects analyses was used to isolate the number of voxels activated for each listener. This analysis confirmed that the hemispheric advantage was in the expected direction, but did not reach significance (LH: $M = 67$ voxels, $SE = 15$, vs. RH: $M = 51$, $SE = 9$, $t(14) = 1.156$, $p = .13$). There was a marked asymmetry in the topography of STG activation between hemispheres: the area of differential activation in the left hemisphere focused around a peak in the posterior STG (y = -44), while that in the right hemisphere was centered anteriorly and ventrally around the middle STG (y = -16). In addition to the temporal lobe activation, speech stimuli, but not complex non-speech, activated a small cluster in the right inferior frontal gyrus (IFG; see Table 1ai). There were no areas that were more activated for complex non-speech relative to speech. An exploratory random-effects analysis confirmed the robustness and generalisability of these results (see insert in Figure 2, Table 1aii). This more stringent analysis confirmed the bilateral activation of the MTG, and highlighted the degree of left hemisphere lateralisation. Differential activation of the STG was observed *only* in the left hemisphere, in the posterior area classically associated with receptive language. The activation observed in the right inferior frontal cluster did not survive the random-effects analysis.

Complex stimuli versus simple tones. The relative cortical activation to speech and complex non-speech was compared with that elicited by the simple tones. Compared with simple tones, speech activated the STG and MTG bilaterally (see Figure 4a, Table 1b). The extent of this differential activation was greater than that observed when speech was compared to complex non-speech.

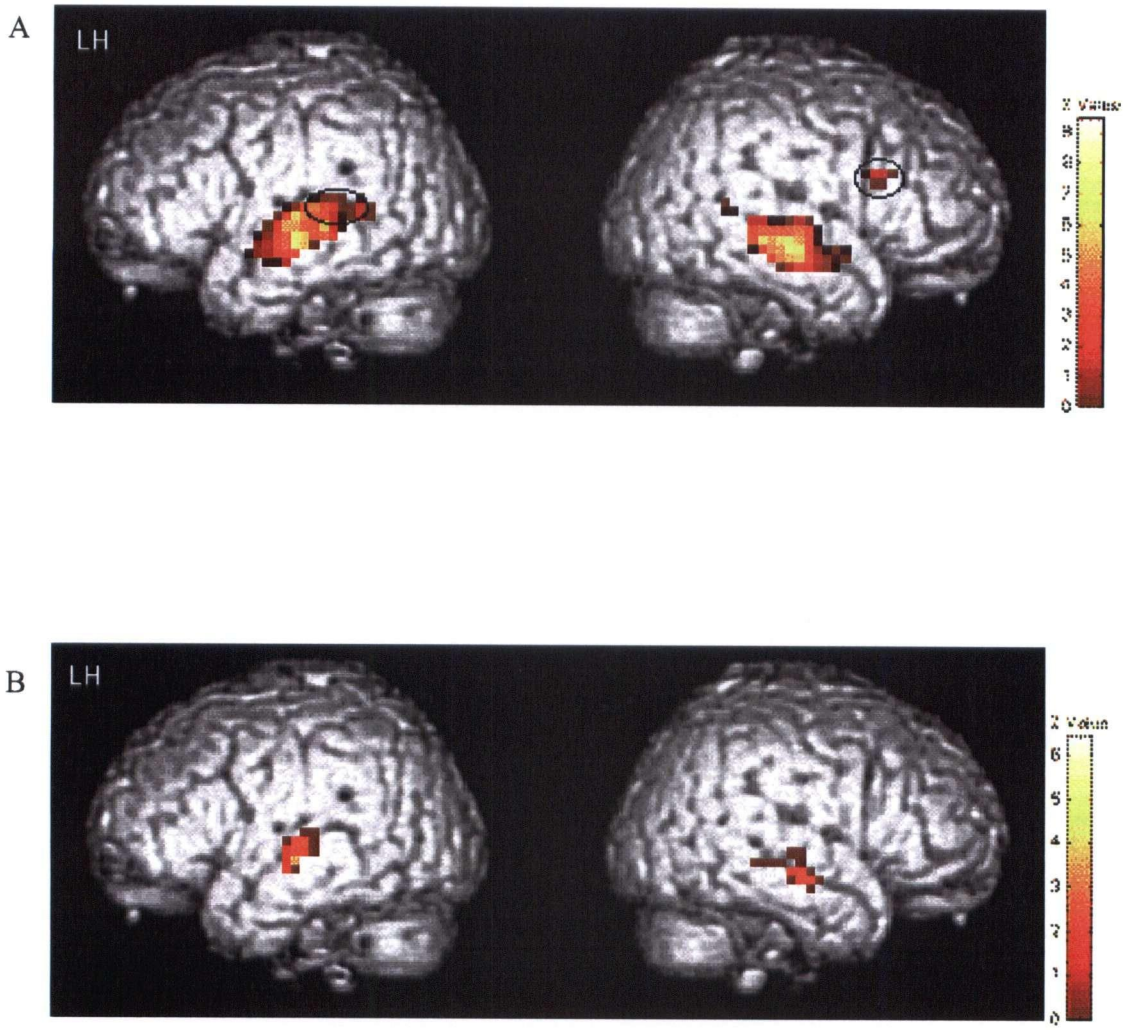

Figure 4. Cortical activation for complex sounds versus simple sounds.

Cortical surface rendering of differential activation for: (A) Speech relative to simple tones. Circumscribed areas map onto the posterior superior temporal gyrus in the left hemisphere, and the inferior frontal gyrus in the right hemisphere; and (B) Complex non-speech relative to simple tones (all comparisons: fixed-effects analysis; display threshold $z = 4.63$; colour bars indicate corresponding z-score).

Table 1. Areas of significant activation for all comparisons.

Areas of significant activation for three comparisons of interest: (a) speech compared with complex non-speech sounds using (i) a fixed-effects analysis, and (ii) a random-effects analysis, (b) speech compared with simple tones, and (c) complex non-speech compared with simple tones. Co-ordinates (x, y, z) are reported in the modified Talairach space used by SPM99.

Region	Talairach Co-ordinates (mm) x	y	z	z-score
Comparison of Interest				
(ai) Speech vs. complex non-speech				
L Middle Temporal Gyrus	-60	-16	-8	6.17***
L Middle Temporal Gyrus	-64	-36	0	5.76***
L Superior Temporal Gyrus	-64	-44	12	4.72*
R Middle Temporal Gyrus	56	-28	-4	5.32**
R Middle Temporal Gyrus	52	-20	-16	5.31**
R Superior Temporal Gyrus	52	-16	0	4.63*
R Inferior Frontal Gyrus	40	24	16	4.66*
(aii) Speech vs. complex non-speech (randomFX)				
L Superior Temporal Gyrus	-64	-32	4	5.07**
L Middle Temporal Gyrus .	-60	-20	-8	4.83*
R Middle Temporal Gyrus .	52	-20	-8	4.80*
(b) Speech vs. simple tones				
L Middle Temporal Gyrus	-60	-16	-8	9.49****
L Middle Temporal Gyrus	-64	-24	4	8.64****
L Superior Temporal Gyrus	-44	-32	8	5.73****
R Middle Temporal Gyrus	56	-12	-12	9.09****
R Middle Temporal Gyrus	56	-28	-4	8.16****
R Inferior Frontal Gyrus	48	16	24	5.52***
R Superior Temporal Gyrus	52	-48	8	5.37**
R Insula	32	24	12	4.70*
(c) Complex non-speech vs. simple tones				
L Superior Temporal Gyrus	-64	-24	4	6.44****
L Superior Temporal Gyrus	-60	-16	-4	5.48***
R Middle Temporal Gyrus	56	-12	-12	6.39****
R Middle Temporal Gyrus	60	-28	-4	4.94*
R Superior Temporal Gyrus	64	-16	0	4.87*

* $p < .05$, ** $p < .01$, *** $p < .001$, **** $p < .0001$
L, left hemisphere; R, right hemisphere

Compared to simple tones, complex non-speech activated bilateral foci in the STG and MTG, which were smaller in extent and in magnitude than the differential activation to speech (see Figure 4b, Table 1c). There were no areas that were relatively more activated for simple tones compared to either speech or to complex non-speech.

2.5 Discussion

The present study demonstrates that even in a simple detection task, speech elicits greater and topographically different cortical activation than complex non-speech analogues. Specifically, the detection of speech activated the middle temporal gyrus bilaterally, a unique locus in the left posterior superior temporal gyrus in the vicinity of Wernicke's area, and a small locus in the right inferior frontal gyrus (see Figures 2, 4a). This pattern of results suggests commonalities in the neural substrates processing complex auditory stimuli, as well as some degree of functional specialisation for speech even at this early processing stage.

Recruitment of classic receptive language areas : A specialisation for speech detection from the initial stages of processing

When compared with complex non-speech, the detection of speech elicited activation in classic receptive language areas along the Sylvian fissure, including the auditory association cortex (Brodmann areas [BA] 22) of the left STG, the posterior part of which is classically referred to as Wernicke's area. In addition, speech elicited differential activation bilaterally in the MTG (BA 21/22). There was a trend for more extensive activation to speech in the left MTG compared with this area's right

hemisphere homologue. This greater extent of differential activation in the left temporal lobe, and the unique locus of differential activation in the posterior STG, suggest a left hemisphere lateralisation for speech processing.

The bilateral, but left hemisphere weighted, activation of the temporal lobes that we observed during speech detection is consistent with the results of many studies directly comparing speech and non-speech processing. Most studies have reported both bilateral cortical activity, as well as loci of activation lateralised to the left hemisphere. Bilateral activation of the temporal lobes has been reported in comparisons of speech with noise bursts (Binder et al., 2000; Binder, Rao, Hammeke, Frost et al., 1994; Zatorre et al., 1992), signal-correlated noise (Mummery et al., 1999), musical non-speech counterparts (Benson et al., 2001), and reversed speech (Binder et al., 2000). Many of these studies also report the lateralisation of a unique locus in the left hemisphere that is differentially activated by speech though the exact location is variable: some studies report activity in the supramarginal gyrus (Benson et al., 2001; Celsis et al., 1999), others activate the posterior STG (Binder, Rao, Hammeke, Yetkin et al., 1994; Mummery et al., 1999; Zatorre et al., 1992), or the anterior STG (Price et al., 1996). The more anterior activation observed in the latter study might be due to their use of real words (which presumably activate additional processes beyond simple speech detection) compared with the nonsense words used in other studies, including our own.

This study corroborates previous neuroimaging and neuropsychological research indicating that human language is processed by unique neural substrates. The recruitment of the left posterior STG in a variety of speech processing tasks (see Table 2) suggests

that it plays an important role in that process. The activation we observe in this study using a simple oddball detection task which (a) does not require overt identification of the stimulus as speech, and (b) does not require in-depth linguistic analysis of the speech stimuli, suggests that the role the left posterior STG plays is fundamental during the first steps of linguistic processing. This pattern of findings is inconsistent with theories in which the recognition and analysis of speech is like the perception of all sounds (Diehl & Kluender, 1986). Instead, these results argue for modular theories of speech perception (e.g., Fodor, 1983; Liberman, 1996), which claim that the processing of language is specialised from the initial detection of linguistic stimuli. Though our results strongly suggest a neural specialisation for processing human speech, we will discuss several additional accounts that could enrich the interpretation of these findings.

Potential effects of attention and familiarity

One account of the differential cortical activation elicited during speech detection speaks to the increased attentional resources recruited during its processing. Though attention is likely to play some modulatory role on the magnitude of the haemodynamic response during auditory processing in our task, we think it unlikely to fully account for the pattern of differential activation we observe.

Table 2. Studies contrasting speech and non-speech processing.

Co-ordinates (x, y, z), in standard Talairach space, of the peak activation area in the posterior superior temporal gyrus (STG) of the left hemisphere.

			Talairach Co-ordinates		
Study	**Task**	**Non-speech comparison**	**x**	**y**	**z**
Benson et al. (2000)	passive listening	musical non-speech	-62	-23	16
Binder et al. (1994)	passive listening	white noise	n/a	n/a	n/a
Binder et al. (2000)	nonspecific button press	tones	-52	-42	6
Celsis et al. (1999)	change detection	tones	-30	-52	30
Mummery et al. (1999)	passive listening	signal correlated noise	-54	-38	8
Zatorre et al. (1992)	nonspecific button press	white noise	-58	-21	8
Vouloumanos et al.	simple detection	complex matched non-speech	-63	-42	13

In designing our task as one of simple detection with both speech and complex non-speech as infrequent oddballs embedded among background tones, we effectively equated task-related attentional demands between the two types of oddball stimuli. Listeners were required to monitor the auditory stream and perform the same operation irrespective of stimulus type. Moreover, in using the event-related design which allows for pseudorandomisation of stimuli of interest, listeners' attention levels should have been maintained throughout as there is no reliable way for them to anticipate the next stimulus type (standard or oddball), as may be the case in block designs.

Although this study explicitly controlled task-related attentional demands, it could be argued that speech stimuli recruited greater attentional resources through their status as familiar sounds in auditory space. We have attempted to minimise this confound by: (a) using a nonsense word, which was novel to the listener, and (b) comparing speech to high and low tones, which are familiar sounds frequently used as dial tones, busy signals, or alerting signals at traffic crossings adapted for persons with disabilities. Comparisons of speech to relatively novel complex non-speech and to relatively familiar tones yielded a similar differential pattern of activation. This suggests that familiarity, if implicated at all, would only contribute modestly to the pattern of activation we observe.

The results of neuroimaging studies investigating auditory attention during speech processing corroborate the modesty of its potential contribution. Selective attention tasks report patterns of temporal lobe activation in part similar to the pattern we observe in our speech detection task, and in part notably different. As in our study, tasks of selective attention revealed a left hemisphere advantage in the MTG for attended conditions

(Hashimoto, Homae, Nakajima, Miyashita, & Sakai, 2000; Pugh et al., 1996). However, in the STG, the attention-related activation elicited by these tasks was bilateral (Grady et al., 1997; Hashimoto et al., 2000; Pugh et al., 1996), in stark contrast to the left-lateralised activation locus we observe in the posterior STG (see Figure 2). The differential activation in the posterior STG of the left hemisphere elicited during speech processing is therefore unlikely to be modulated by attention.

On the other hand, attentional mechanisms are likely to modulate the activation we observe in the right inferior frontal gyrus. This area has reliably been recruited in studies investigating attention and general arousal (Hashimoto et al., 2000; Pugh et al., 1996; Stevens & Schwartzreich, 2000; Tzourio et al., 1997). Although this activation might be speech-specific, a number of alternative explanations have been proposed. In a different task that preserves some features of the "change detection" aspect of our task, Celsis et al. (1999) also observed a small region in the right frontal gyrus that was activated by a deviant sequence containing a square tone (compared to a sine tone). The authors suggest pitch monitoring of deviants (consistent with the results of Pugh et al., 1996; Zatorre, Evans, & Meyer, 1994; Zatorre et al., 1992), and differences in spectral content between stimuli of interest (square > sine) as possible explanations. Another interesting possibility is that this region is involved in processing non-phonetic ('voice') aspects of natural speech, but not in processing words themselves (but see Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Stevens & Schwartzreich, 2000).

Is there a ventral specialisation for speech processing?

It has been proposed that the temporal lobe is structured around a functional dichotomy, in which the dorsal and lateral surfaces of the STG are involved in unimodal auditory processing (Galaburda & Sanides, 1980), whereas the ventral aspect receives and integrates input from many modalities (Baylis, Rolls, & Leonard, 1987; Mesulam, 2000; Rauschecker, 1998). Since language processing often integrates information from multiple modalities (McGurk & MacDonald, 1976), language functions are likely to be carried out in areas where different streams converge (Geschwind, 1965). The results of some neuroimaging studies comparing speech and non-speech processing support this hypothesis. Studies comparing speech to white noise (Binder et al., 2000; Binder, Rao, Hammeke, Yetkin et al., 1994; Zatorre et al., 1992), tones (Binder et al., 2000), and signal correlated noise (Mummery et al., 1999) reported non-specific activation in the dorsal aspect of the STG to both speech and non-speech, whereas activity in the ventral STG, and in the STS, was more closely correlated with speech alone.

Our study does not provide the level of anatomical resolution required to address this issue conclusively. However, within our limited resolution, our results seem to be consistent with these studies in suggesting that the detection of speech is more closely associated with ventral processing streams. Compared with complex non-speech, speech activated a large region of the MTG, and areas of the STG along the STS (see Figure 2). A comparison of speech to simple tones revealed a similar, but more extensive, pattern of activation including massive recruitment of the MTG and the ventral aspect of the STG (see Figure 4a). Thus, our results seem to be consistent with a more ventral specialisation

for speech processing. Clearly, further converging research is required to confirm this ventral precedence.

Is differential activation a matter of complexity?

A possible explanation for the differential activation to speech may be that speech stimuli are acoustically more complex than the sine-wave analogues (despite matching on coarse spectral content, speech contains broadband frequency information and harmonic spectra that are lacking in the non-speech analogues), and therefore recruit more of the same cortical resources during processing. Indeed, the overlapping regions of maximal activation observed in the MTG during processing of speech and complex non-speech would point to at least some common processing mechanisms (compare Figures 4a and 4b). The differences in topography of STG activation (particularly the lateralised locus in the left posterior STG) argue against this. Moreover, recent evidence from a study comparing processing of speech and musical non-speech stimuli of systematically varying complexity indicates that non-speech complexity is reflected in corresponding activation of Heschl's gyrus and the areas immediately adjacent, and not in regions posterior to secondary auditory cortex (Benson et al., 2001). For these reasons, it is unlikely that stimulus complexity can fully account for the differential pattern of activation.

Is lateralisation due to the rapid transitions of speech?

Temporal processing theorists suggest that left hemisphere lateralisation for speech processing results from a specialisation for processing all rapidly changing

acoustic information, not for speech per se (e.g., Tallal, Miller, & Fitch, 1993). Although the left hemisphere does possess an advantage for processing rapidly changing temporal information (e.g., Belin et al., 1998; Johnsrude, Zatorre, Milner, & Evans, 1997; Schwartz & Tallal, 1980), the temporal processing hypothesis is unlikely to account for the lateralisation we observe during speech detection. The sine-wave non-speech counterparts in this study maintain the peak frequency changes of the three formants and the fundamental frequency of speech across time, thus preserving the main rapid temporal changes present in our speech stimuli. As a result, rapidly changing temporal information was preserved in both speech and non-speech stimuli, yet a left hemisphere lateralisation was only observed during speech detection. The left hemisphere advantage for speech processing in this detection task is not related to the temporal attributes that characterise speech, but rather to the fact that the stimulus is speech by nature.

Conclusions

This study corroborates evidence from previous neuroimaging and neuropsychological studies indicating that human language is processed by unique neural substrates. Moreover, our results suggest that this specialisation for speech is present from the early stages of processing. Our study is unique in contributing to this growing body of work in five important ways. (A) We compare speech to carefully controlled complex non-speech stimuli (sine-wave analogue) that track key spectral and temporal aspects of speech. (B) We use an event-related fMRI design which allows us to model the haemodynamic response to individual auditory events. This, in comparison with block designs, allows us to isolate more precisely the neural events associated with

speech detection. (C) Our use of a simple oddball detection paradigm embeds speech and complex non-speech within a uniform background, and equates attentional demands for processing the stimuli of interest. (D) Our testing of a larger sample than most neuroimaging studies, and the corroboration of our results using an exploratory random-effects analysis, strengthen the generalisability of the findings. (E) The differential activation we observe cannot be fully accounted for by the differential recruitment of attention during speech processing, by the acoustic-level characteristics of the speech stimuli, such as complexity or rapid transitions, or by higher-order linguistic processing (e.g., semantic or syntactic) of the stimuli. Instead, these neural substrates appear to be specifically activated by properties intrinsic to speech.

Clearly, detected speech input generally requires further linguistic analysis, and that analysis, be it semantic, syntactic or phonological, will likely activate additional brain mechanisms. However, the pattern of activation we observe suggests that some distinct neural mechanisms are involved in the initial processing of speech, and elucidates how, in the cacophony of sounds in the environment, spoken language stands out as an exceptionally salient signal for the human brain.

Chapter 3. Privileged status of speech in early infancy

3.1 Preface

As demonstrated in Experiment One, detecting speech in a series of sounds engages specific neural substrates in the left hemisphere of adults. The neural specificity for speech appears to have its roots in early development: infants show a similar asymmetrical neural signature when listening to speech compared with backward speech (Dehaene-Lambertz et al., 2002; Peña et al., 2003). The neural specificity for speech in infancy could exist in tandem with behavioural differences in how infants listen to speech compared with other sounds. Thus, speech could have a special status for young infants, preferentially engaging their attention compared with other sounds. Experiment Two investigates whether young infants from 2 to 7 months of age demonstrate a bias for listening to speech compared with non-linguistic sounds. A bias for speech would direct infants towards spoken language in their environment and allow them to identify speech as an important sound. This study is reprinted, with permission from Blackwell Publishers, from Developmental Science (in press).

3.2 Introduction

Is speech a privileged signal for infants? At birth, the newborn's perceptual system is already tuned to some of the dimensions of human speech that are exploited by the phonological and syntactic systems of language (Bertoncini et al., 1987; Christophe et al., 1994; Jusczyk, Bertoncini, Bijeljac Babic, Kennedy, & Mehler, 1990; Mehler et al.,

1988; Nazzi et al., 1998; Ramus et al., 2000; Sansavini et al., 1997; Shi et al., 1999). These initial sensitivities become increasingly tuned to the properties of the native language during the infant's first year (Werker & Tees, 1992), a refinement that is reflected both in the decline of infants' discrimination of contrasts or properties that are not informative for processing their native language (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Tees, 1984), and in the gain of sensitivities that are pertinent to native-language processing (Jusczyk, Cutler et al., 1993; Jusczyk, Hohne, & Bauman, 1999; Myers et al., 1996). Moreover, towards the end of their first year, infants become able to integrate multiple cues, and thus perform more sophisticated analyses on linguistic input (Morgan & Saffran, 1995). These observations suggest that infants' processing of the speech signal undergoes radical reshaping in the first year of life, becoming more specific and more sophisticated as infants approach their first birthday. Though the perception of speech compared with non-speech has been shown to engage specific neural substrates in the left hemisphere of adults (e.g., Binder & Price, 2001; Scott, Blank, Rosen, & Wise, 2000; Vouloumanos, Kiehl, Werker, & Liddle, 2001) and infants (Dehaene-Lambertz et al., 2002; Peña et al., 2003), the mechanisms underlying the changes in speech perception evident during infancy are as yet unknown. The late Peter Jusczyk (1997) suggested that one factor that may facilitate infants' enhanced language processing might be a bias for listening to speech compared with other sounds. Such a bias might shape the character of learning processes and guide the action of perceptual mechanisms (Bolhuis & Honey, 1998; Johnson, 1999; Jusczyk, 1997; Marler, 1990).

In this study, we investigate whether young infants demonstrate a bias for listening to speech. Specifically, we ask if infants listen preferentially to speech compared with other non-linguistic sounds. To date, no study has addressed this question directly, though two previous studies provide relevant data.[4] Columbo and Bundy (1981) found that 4.5-month-old infants fixated longer on a visual target when it was associated with continuous female speech compared with continuous unfiltered white noise. However, 2-month-olds failed to show a similar fixation bias when continuous female speech was contrasted with silence (Colombo & Bundy, 1981). Glenn, Cunningham, and Joyce (1981) found that 9-month-olds pulled a lever more frequently to listen to a female voice singing *a cappella* compared with three solo musical instruments playing the same tune. These results suggest that infants favour speech over some sounds. However, our understanding of a potential listening bias remains incomplete. The non-speech counterparts (white noise and musical instruments) used in these studies differ greatly from speech in frequency and timing characteristics; shedding little light on how discriminating a potential listening bias for speech might be (Jusczyk, 1997). Moreover, the listening biases of younger infants, who have significantly less experience with speech, are as yet unknown.

To test the specificity of infants' listening preference for speech and control for infants' sensitivity to superficial acoustic dimensions that are characteristic of the speech signal, we sought to contrast speech with closely matched complex non-speech sounds that preserved many aspects of the spectral and timing dimensions of speech without actually sounding like speech to naïve listeners. To this end, we used complex non-

speech analogues that are modelled on sine-wave analogues of speech (Remez et al., 1981; Vouloumanos et al., 2001). These complex analogues consist of time-varying sinusoidal waves that track the resonant center frequencies of natural speech and reproduce the changes in these frequency peaks across time (see Figure 5). To ensure that the two different signals were equally attractive to infant ears, we created analogues that retain information about the pitch contour of the speech counterparts, since pitch contour has been shown to underlie infants' preference for infant-directed speech (Fernald & Kuhl, 1987) and their ability to discriminate their native language (Mehler et al., 1988). The similarities between the acoustic properties of the complex non-speech analogues and natural speech allow us to investigate whether infants are attracted to signals having a particular acoustic form, and to better delineate the range of signals that engage a potential listening bias.

To track the emergence of a listening bias for speech, we focused on the first half-year of life. Previous studies had shown that that infants listen selectively to speech compared with acoustically dissimilar unpatterned sounds by 4.5 months (Colombo & Bundy, 1981). Since the perceptual similarities between speech and our non-speech analogues may render the task more difficult, we began by testing older infants of 6.5 months. To determine whether speech has a special status during early infancy, we tested younger infants of 4.5 months and 2.5 months.

3.3 Method

3.3.1 Participants

Infants were recruited at birth from the British Columbia Women's and Children's Hospital, in Vancouver, Canada, or through advertisements placed in the community section of various local newspapers. Parents were subsequently contacted by phone to participate. All infants were full term deliveries and heard at least 20% English in their home environment.

Forty-eight infants of three ages were included in this study: sixteen 6.5-month-olds (M = 6;14), sixteen 4.5-month-olds (M = 4;19), and sixteen 2.5-month-olds (M = 2;16). An additional thirty-eight infants were tested but were excluded from the analysis.[5]

3.3.2 Materials

Auditory stimuli were of two types: a "speech" set composed of nonsense words, and a "non-speech" set composed of complex non-speech analogues (Figure 5). These stimuli have been used in previous studies with adults (Vouloumanos et al., 2001)

Speech. Speech stimuli consisted of twelve tokens of two monosyllabic nonsense words (six "lif" tokens and six "neem" tokens)[6] spoken by a female native English speaker. Tokens varied in intonational contour (average minimum and maximum pitch: 197 Hz and 350 Hz, respectively), and in duration (525 - 1155 ms).

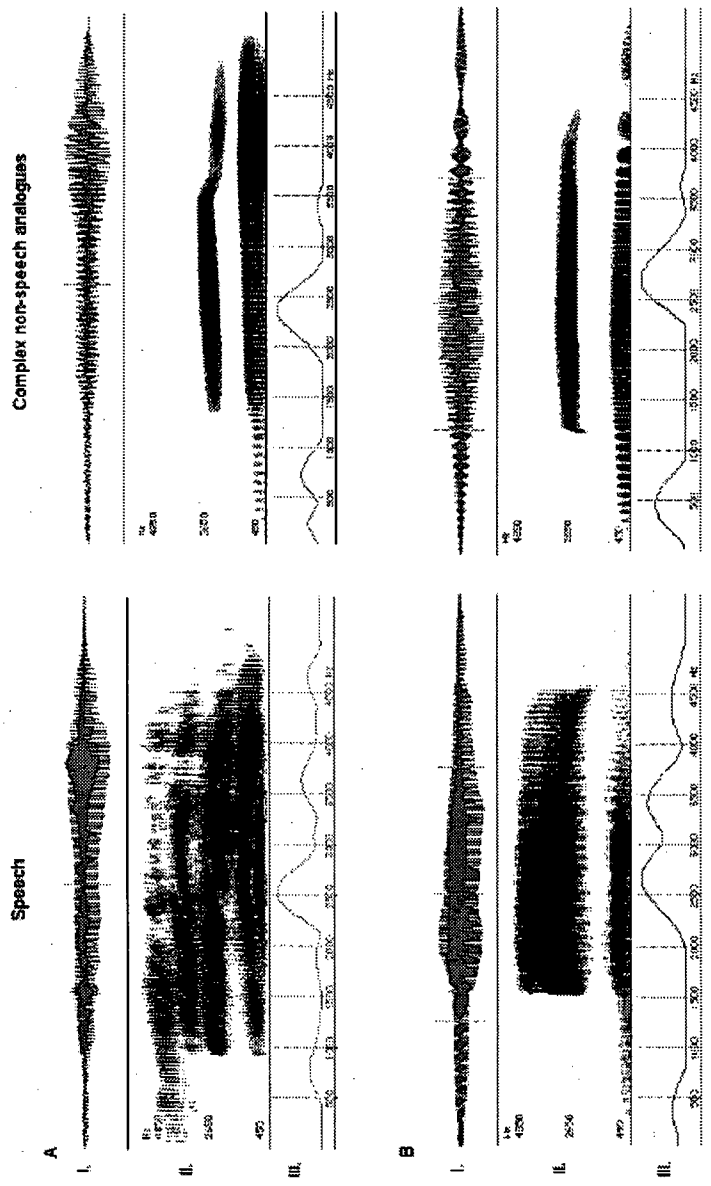

Figure 5. Speech and complex non-speech stimuli for 'lif' (A), and 'neem' (B).

Similarities between the two types of stimuli are illustrated as: (I) waveform diagrams, (II) spectrograms showing changes in frequency across time for the first three formants, as well as for the fundamental frequency (F0), the correlate of pitch, and (III) spectra depicting the relative amplitudes of different frequencies.

Complex non-speech. Non-speech analogues were created by Sonya Bird and Guy Carden (Department of Linguistics, University of British Columbia, Canada). Non-speech stimuli consisted of time varying sinusoidal waves that tracked the main regions of significant energy in natural speech (namely the fundamental frequency and the first three formants). Sinusoidal waves tracking these energy peaks were created individually using Mathcad 3.1 (Mathsoft Inc., Cambridge, MA). Fundamental frequency (corresponding to pitch) was also tracked individually for each of the twelve speech tokens. Because the first three formants were virtually identical across the multiple natural repetitions of the two word types, one representative set of formants from a token of each word type was tracked. For the "lif" tokens, this representative formant set was composed of the first formant of the initial consonant segment ("l"), and the first three formants of the vocalic segment ("i"). The analogue for the fricative "f" was created using a white noise generator and passed through a Butterworth filter (Pass Band cut-offs of 1700 and 4380 Hz, filter order 9) in Signalyze 3.12 (Agora Language Marketplace, Charlestown, MA). This representative set was then added onto a sinusoidal wave tracking the pitch contour of each of the six "lif" segments using Signalyze 3.12 to create six different non-speech analogue "lif" stimuli. For the "neem" tokens, the representative formant set was composed of the first formant of the initial ("n") and final ("m") consonants, and the first three formants of the vocalic segment ("ee"). This representative set was then added onto a sinusoidal wave tracking the pitch contour of each of the six "neem" segments using Signalyze 3.12 to create six different non-speech analogue "neem" stimuli. Tokens were identical to speech foils in intonational contour

(average minimum and maximum pitch: 197 Hz and 350 Hz respectively), and in duration (525 - 1155 ms). Moreover, analogues retained the amplitude envelope, relative formant amplitude, and relative intensity of their speech counterparts. Crucially, the pitch contour of natural speech was preserved in the complex non-speech analogues.

3.3.3 Design and Procedure

Testing was conducted in a 7-ft by 9.5-ft sound-attenuated room. The walls were covered by matte black curtains, and the sole lighting source was a 60 W floor lamp. Infants were seated on the lap of a parent or guardian, 4 ft away from a 27 in. Mitsubishi CS-27205C television monitor (640 x 480 line vertical resolution), that protruded through a hole in the front curtain. Sounds were played at an average amplitude of 68 dB (± 3 dB) using a BOSE 101 speaker placed directly above the television monitor. Infants were recorded with a Panasonic AG 180 video camera placed behind the front curtain with its lens positioned 10 in. below the television monitor. Parents were told that we were investigating "how infants listen to different sounds", and wore Koss TD/65 or Peltor workstyle HT7A headphones playing music in order to mask the experimental sounds. Experimenters were blind to the specific condition being tested for any particular infant, and controlled the presentation of the stimuli from a separate room while monitoring the infants over a closed circuit using a Panasonic CT-13R12CT colour television. Stimulus presentation was controlled from a Power Mac 8500/1200 computer interfaced with a Sony LDP-1550 laser disc player using Habit 7.6 (Leslie Cohen, University of Texas at Austin).

Infants were tested using an infant-controlled sequential looking preference (SLP) procedure (Cooper & Aslin, 1990; Cooper & Aslin, 1994; Shi & Werker, 2001; Pegg, Werker, & McLeod, 1992). In our version of this procedure, the program initially presents a red flashing light on the monitor to attract the infant's attention.[7] Once the infant fixates on the screen, testing begins. A stationary black and white checkerboard is displayed on the monitor at the same time as one set of experimental sounds is played from a hidden speaker placed on top of the monitor. In this SLP procedure, stimulus presentation is contingent on the infant's behaviour: infant fixation determines the onset and offset of every trial. The sound and checkerboard show continues for as long as the baby looks at the monitor. When the infant looks away continuously for longer than 1 s, stimulus presentation ceases. When the infant looks back at the monitor, the next trial begins: the checkerboard is displayed once again, but this time in tandem with the other set of experimental sounds. In any one experiment, the infant is presented with a total of 10 trials, 5 speech trials alternated with 5 non-speech trials. A full trial consists of 14 tokens chosen randomly from the set of 12 tokens, separated by 300-500 ms silence, for a maximum trial length of 20 s with the two older groups and 40 s with the youngest group.[8] For any given trial, speech or non-speech, tokens were ordered in a semi-random fashion so that every fixed window of four tokens included at least 2 "lif"s and 2 "neem"s. For half the infants, trial order was reversed.

3.4 Results

As is standard with the SLP procedure, the first trial was excluded from the analysis (Cooper, Abraham, Berman, & Staska, 1997; Cooper & Aslin, 1994; Shi & Werker, 2001). Since order of presentation was counterbalanced, an equal number of speech and non-speech trials was thus excluded. Using the remaining 9 trials, we calculated each infants' total looking time for each type of sound, speech or non-speech. This was based on 15 frame-per-second, frame-by-frame coding of infant looks towards the screen during each sound trial. Because maximum trial length varied for different age groups and individual infants varied in their total looking time, we normalised the data by transforming total looking time to proportion looking time ($Prop_{speech} = M_{speech}/[M_{speech} + M_{nonspeech}]$; for $Prop_{nonspeech} = 1 - Prop_{speech}$). A 3 (age: 6.5, 4.5, 2.5 month-olds) X 2 (sound type: speech vs. non-speech) X 2 (sex: female vs. male) X 2 (trial order) mixed analysis of variance indicated a main effect of sound type, $F(1,36) = 15.048$, $p = .000$, with proportionally longer looking times during speech trials ($M = .558$, $SE = .015$) than during non-speech trials ($M = .442$, $SE = .015$). No other main effects and no interactions were significant. We undertook a series of planned comparisons to determine if this effect was present for each age group. Individual 2-tailed paired-sample t-tests on the average looking times revealed that a significant difference was present at each age: 2.5 month-olds: $t(15) = 2.143$, $p = .049$; 4.5 month olds: $t(15) = 2.174$, $p = .046$; and 6.5 month-olds: $t(15) = 2.556$, $p = .022$. The average looking times of infants in different age groups are illustrated in Figure 6.

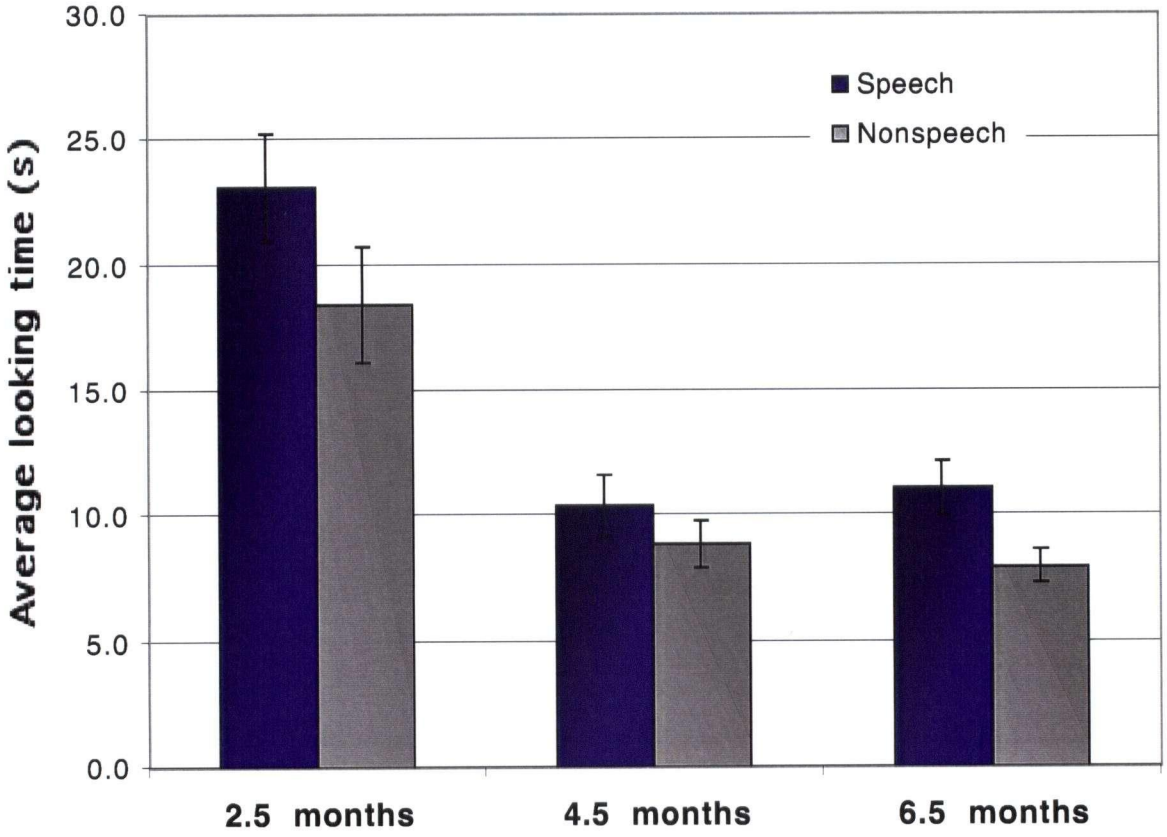

Figure 7. Average looking times for each age group: 2.5 months, 4.5 months and 6.5 months.

Comparisons are significant at each age (see text).

A binomial test revealed that significantly more infants showed longer looking times for speech (n = 34) than for non-speech (n = 14), $p = .006$.

3.5 Discussion

Infants between 2 and 7 months of age listened longer to speech compared with structurally similar complex non-speech sounds. Earlier results showed that infants listen preferentially to speech in comparison to white noise and musical instruments (Colombo & Bundy, 1981; Glenn et al., 1981). Our results demonstrate that infants prefer speech even when it is contrasted with acoustically similar non-speech sounds. Moreover, we demonstrate that a bias for listening to speech is shown by infants as young as 2.5 months of age, 2 months earlier than the youngest age at which a bias for speech had been previously demonstrated (Colombo & Bundy, 1981).

Spectrally and temporally matched complex non-speech analogues fail to capture infant interest as effectively as speech despite the structural similarities between the two types of sounds and despite the inclusion of the pitch contour in the non-speech analogues. The range of signals engaging a speech listening bias thus appears to be quite narrow. The non-speech stimuli used in this study preserve the time-varying frequency intervals that are characteristic of speech, as well as the relative formant amplitudes, the amplitude envelope and, importantly, the pitch contour of speech through the inclusion of the sinusoidal wave tracking the fundamental frequency. The fundamental frequency was included to preserve the pitch component known to be attractive to infants (Fernald, 1985), to add variability to the analogues, and to make the non-speech sounds more

acoustically similar to their speech counterparts. The complex analogues did not, however, retain the characteristics of the voicing source, the broader band formant information and parts of the harmonic spectrum, nor did they preserve the voice quality, biological quality or unified source characteristics of speech (Remez et al., 1981). Thus, analogues lacked some of the qualities unique to speech as a biological sound produced by a human vocal tract. It remains to be seen which of these dimensions engages the bias that we observe.

In previous studies, the presentation of speech in the form of sentences was shown to be important for investigating certain aspects of infant speech perception. Sentences, but not isolated words, elicited a preference for the mother's voice (Mehler et al., 1978). Similarly, infants' early discrimination of their native language is based on rhythmical information present in sentence form (Mehler et al., 1988), and only later is evident for word-level segmental information (Jusczyk, Cutler et al., 1993). In order to create a stringent test for investigating infants' preference for speech, we presented speech as isolated words, rather than sentences. The bias that infants show for speech is present when speech is presented as isolated words, indicating that this level of information is sufficient for eliciting a preference for speech.

The young age at which infants listen preferentially to speech suggests that this bias is present very early in development. Such a listening bias could either be a *reflection* of the more extensive processing that speech undergoes, or the *cause* of subsequently more sophisticated processing, or both. Since the youngest infants tested in this study were 2.5 months old, it is not yet clear whether infants' bias for speech derives

from exposure to language or whether it precedes it. We are currently investigating whether a listening bias for speech is present at birth (Vouloumanos & Werker, under review).

Regardless of the origin of a listening bias for speech, such a bias could benefit young language learners by allowing infants to separate and select speech out of their auditory environment in order to analyse the signal more completely (e.g., Jusczyk, 1997; Jusczyk & Bertoncini, 1988). Similar biases have been proposed in other domains, and may be pervasive in early development. For example, Johnson, Umilta and their colleagues have proposed that an initial bias for orienting towards face-like stimuli plays a role in the development of face recognition (Johnson et al., 1991; Valenza et al., 1996).

The preference infants show for speech is reminiscent of the differential processing that some birds and primates demonstrate for calls of their own species (Gottlieb, 1997; Hauser, 1996; Marler, 1990). In some non-human primates, conspecific (same species) vocalisations are preferentially processed by neurons along the superior temporal sulcus (Wang & Kadia, 2001; Wang, Merzenich, Beitel, & Schreiner, 1995). This area may also be preferentially recruited in humans for processing human vocalisations (Belin et al., 2000; Binder & Price, 2001; Scott et al., 2000). Studies using behavioural, electrophysiological and neuroimaging methods with newborns (Bertoncini et al., 1989; Ecklund-Flores & Turkewitz, 1996; Peña et al., 2003) and older infants (Dehaene-Lambertz et al., 2002) suggest that young infants show a left hemisphere bias for processing speech. It is not yet known which parts of the human auditory system are tuned to conspecific sounds early in development, or whether neurons that respond

specifically to conspecific *communication* sounds are present in the human cortex as they are in other species. The existence of conspecific biases implemented by specific neural substrates across different species suggests that these preferences may serve an adaptive function by helping organisms orient to conspecific information.

The results reported in this study provide the best evidence to date that very young infants listen selectively to speech compared with other sounds, in this case, sounds that mimic many of the physical characteristics of speech. This listening selectivity indicates that early-functioning biases direct infants' attention to speech, granting speech a special status in relation to other sounds. Such a bias might provide the basis for more sophisticated analyses of the speech signal which help propel the infant into rapid language acquisition.

Chapter 4. The origin of the speech bias

4.1 Preface

The results of Experiment Two demonstrate that infants as young as 2.5 months show a bias for listening to speech. However as 2.5 month-olds have already had experience with spoken language, the origin of the speech bias is not yet known: does the speech bias derive from infants' exposure to language or does it precede language exposure? To begin to address this question, Experiment Three ("Preference") investigates whether human neonates demonstrate a preference for speech.

Though neonates are relatively naïve, studies have shown that information from the auditory environment percolates into the womb before birth and can influence listening preferences. For example, neonates' attentional biases for particular sounds—such as the mother's voice (DeCasper & Fifer, 1980) and the native language (Mehler et al., 1988)—are clearly based on the infants' specific prenatal experience. At the same time, however, the preference infants show for speech is reminiscent of the differential processing that some birds and primates demonstrate for calls of their own species (e.g., Gottlieb, 1997; Marler & Peters, 1977), which, in some cases, cannot be accounted for by experience (Gottlieb, 1997; Marler & Peters, 1989; Park & Balaban, 1991). The comparative work showing experience-independent biases for conspecific vocalisations in other species suggests that the human bias for speech (demonstrated in young infants in Experiment Two) may be a species-typical behaviour that is independent

of experience. Experiment Four ("Discrimination") investigates whether prenatal experience can account for the human speech bias. This study is under review.

4.2 Introduction

At birth, infants have a remarkable facility for discriminating and categorising many aspects of human language (Mehler et al., 1988; Nazzi et al., 1998; Sansavini et al., 1997; Shi et al., 1999). Moreover, newborn infants show perceptual preferences for specific speech stimuli, such as their mother's voice (DeCasper & Fifer, 1980), and their native language (Mehler et al., 1988; Moon et al., 1993), suggesting that certain stimuli are subject to privileged processing. According to some theories of language acquisition, infants are born with mechanisms that tune them to the properties of language even in advance of experience (Chomsky, 1975). Though the precocious abilities demonstrated in these studies are consistent with this hypothesis, infants' behaviour in these studies could be the product of prenatal listening experience. Is there any aspect of neonatal language processing that could not plausibly be attributed to specific listening experience?

Findings from across the animal kingdom suggest that many organisms may be predisposed to attend to species-specific communication. Birds raised in acoustic isolation prefer listening to their own species' song (Braaten & Reynolds, 1999; Whaling et al., 1997). When given a choice, isolated birds learn to sing their own species' songs more rapidly and accurately than the songs of other species (Marler, 1997). Similarly, female frogs and naïve ducklings prefer to approach conspecific songs compared with

those of other species (Gerhardt, 1988; Gottlieb, 1997). This behavioural selectivity is implemented by specificity at the cellular level. For example, neurons along the superior temporal sulcus of primates respond preferentially and specifically to species-specific vocalisations (Wang et al., 1995). The bias towards conspecific communication signals that organisms show in the absence of prior exposure suggests that these biases may have a genetic basis (Marler, 1990). In this paper, we investigate whether human infants begin language learning with a bias for listening to speech.

It is sometimes claimed that human infants prefer listening to speech, but no study to date has provided convincing evidence of such a bias in young infants. Indeed, Doupe and Kuhl (1999) reach a similar conclusion in a recent review, noting that "[i]n humans, there is no convincing experimental evidence that infants have an innate description of speech". Two often-cited methodological studies compared speech, or speech-like, stimuli to broadband white noise, a non-speech condition deliberately chosen to be unappealing to infants (broadband white noise—the unmodulated sound of radio static, Butterfield & Siperstein, 1970; Colombo & Bundy, 1981). Nonetheless, on the basis of these studies, it was widely reported that infants prefer speech. Because the spectral and temporal parameters of speech and white noise are so different, this interpretation is premature; white noise is a *constant* signal whose physical properties are invariant in time, while speech is a *modulated* signal with individual frequencies that are invariant only relative to each other (Eisenberg, 1976). Any modulated sound, including a tone with changing frequency, could be more interesting than temporally unpatterned white noise. Moreover, in these studies, the youngest infants who favoured legitimate speech

stimuli over white noise were 4.5 months old, leaving open the possibility that the preference observed was based on post-natal exposure to speech.[9]

Electrophysiological studies suggest that the best predictor of newborn and foetal responsiveness to sounds, including speech, is the complexity of an acoustic stimulus, with stimuli that are rich in spectral characteristics (representing many frequencies) or patterned in temporal properties (variable rather than constant tones) eliciting greater changes in EMG (Hutt, Hutt, Lenard, van Bernuth, & Muntjewerff, 1968), EEG (Lenard, von Bernuth, & Hutt, 1969), and heart rate in human infants (Clarkson & Berg, 1983; Groome et al., 2000; Turkewitz, Birch, & Cooper, 1972). Timing and spectral parameters also contribute to the preference for approaching and listening to conspecifics in other species (Gerhardt, 1988) and to the selective responses in primate neurons for conspecific vocalisations (Wang et al., 1995). Differences in spectral and temporal parameters between speech and the non-speech conditions could account for the pattern of infant responsiveness demonstrated in previous studies.

4.3 Methods

4.3.1 Subjects

Neonates (0 – 4 days old) were recruited from a local hospital and tested in a high-amplitude sucking (HAS) procedure.

4.3.2 Stimuli

Speech. Speech stimuli were taken from a set of four tokens of a monosyllabic nonsense word ("lif") spoken by a female native English speaker. Tokens varied in intonational contour (average minimum and maximum pitch: 203 Hz and 325 Hz respectively), and in duration (average: 665 ms).

Complex non-speech analogues. Non-speech stimuli consisted of time varying sinusoidal waves that tracked the main regions of significant energy, namely the fundamental frequency and the first three formants. Sinusoidal waves tracking these energy peaks were created individually using Mathcad 3.1 (Mathsoft Inc., Cambridge, MA). Fundamental frequency (corresponding to pitch) was also tracked individually for each of the four speech tokens. Because the first three formants were virtually identical across the multiple natural repetitions of the word, one representative set of formants was tracked. This representative formant set was composed of the first formant of the initial consonant segment ("l"), and the first three formants of the vocalic segment ("i"). The analogue for the fricative "f" was created using a white noise generator and filtered in Signalyze 3.12 (Agora Language Marketplace, Charlestown, MA) using a Butterworth filter (cut-offs = 1700 and 4380 Hz, filter order = 9). This representative set was then added onto a sinusoidal wave tracking the pitch contour of each of the four "lif" segments to create four different non-speech stimuli. Non-speech analogues retained the duration, pitch contour, amplitude envelope, relative formant amplitude, and relative intensity of their speech counterparts. Though these non-speech stimuli are modelled on sine-wave analogues used in adult-studies (Remez et al., 1981) they are not traditional sine-wave

analogues because they include the fundamental frequency that carries information about pitch contour (Fernald & Kuhl, 1987).

Low-pass filtered sounds. Speech and complex non-speech sounds were low-pass filtered in Signalyze using a Butterworth filter (low pass cut off = 400 Hz, filter order = 8) to approximate the signals audible to foetuses in utero. This manipulation retains frequencies below 400 Hz, to preserve prosody while reducing segmental information (Mehler et al., 1988).

4.3.3 HAS Procedure

Approximately 2 hours after feeding, infants were given a sterilised pacifier that was coupled to a pressure transducer, which measured changes in air pressure caused by sucking. Every session began with one silent minute during which each infant's sucking amplitude was monitored. Following this baseline minute, sound stimuli were presented every time the infant delivered a suck that was in the top 80% of his or her sucking amplitude range. Newborns' HA sucks elicited the presentation of either speech or non-speech tokens. The number of HA sucks for each minute was recorded.

4.3.3.1 Preference experiment

Speech minutes and non-speech minutes were alternated. To ensure that infants had enough exposure to hear the different sounds and learn the contingency, it was necessary to implement criteria for the minimum number of HA sucks. First, in order to hear at least one stimulus per minute and thus be given an opportunity to modulate sucking behaviour, infants were excluded from the analysis if they did not deliver at least

one HA suck in each experimental minute. Second, in order to ensure that enough sounds were heard at the beginning of the study for the infants to demonstrate a potential preference, we began with an arbitrary criterion of a minimum of five HA sucks in each of the first five minutes of the experiment. Upon examination of infant data patterns, it appeared that this low criterion may not have allowed the infants to become familiar with the contingency or the sounds. The criterion was revised to exclude infants who delivered fewer than ten HA sucks in each of the first five minutes ensuring that infants heard enough sounds to be able to display their listening preferences. Twenty-two neonates (mean age = 44.9 hr) met this criterion. An additional 24 infants were not included for the following reasons: falling asleep (1), failing to meet the sucking criterion (8), equipment failure (1), experimenter interference (5), crying or fussing (2), rejection of pacifier (3), sucking weakly (2) and hospital fire alarm ringing during experiment (2).

4.3.3.2 Discrimination experiment

Forty-eight neonates (mean age = 38.6 hr) participated in this study and were in one of three conditions: alternating unfiltered speech and unfiltered non-speech (16 infants), repeating speech (8) or repeating non-speech (8) control conditions, and alternating LPF speech and LPF non-speech (16). According to sucking criteria used in other discrimination studies using the HAS technique (Sansavini et al., 1997), we required that infants undergo no more than 2 minutes without any sucks, and that these "zero" minutes not be consecutive. An additional 52 infants participated but were excluded for the following reasons: falling asleep (3), changing their state after emptying bowels (1), crying or fussing (18), not meeting the sucking criterion (8), experimenter

interference (2), other interference (1), hiccups (1), sucking weakly (2) and rejecting the pacifier (15).

4.4 Preference experiment

4.4.1 Preference experiment: Introduction

To investigate whether infants genuinely demonstrate a bias for species-specific vocalisations, we contrasted human speech with non-speech stimuli that were crafted to control for infants' sensitivity to critical spectral and temporal parameters of speech stimuli. In particular, we contrasted speech with structurally similar complex modulated non-speech analogues that preserve many spectral and timing dimensions of speech without actually sounding like speech to naïve listeners (Vouloumanos et al., 2001). The speech signal is composed of concentrations of energy at multiple frequencies that change over time (Figure 7c). The non-speech analogues were modelled on sine-wave analogues of speech (Remez et al., 1981), and consisted of time-varying sinusoidal waves that track the resonant center frequencies (formants) of natural speech to reproduce the changes in these frequency peaks across time (Figure 7b). These two types of stimuli differ in voice quality (non-speech analogues have none), in naturalness or biological quality (non-speech analogues are artefacts), and in the characteristics of the source (speech has one source, the vocal tract, while non-speech analogues have four, one per sinusoidal tone).

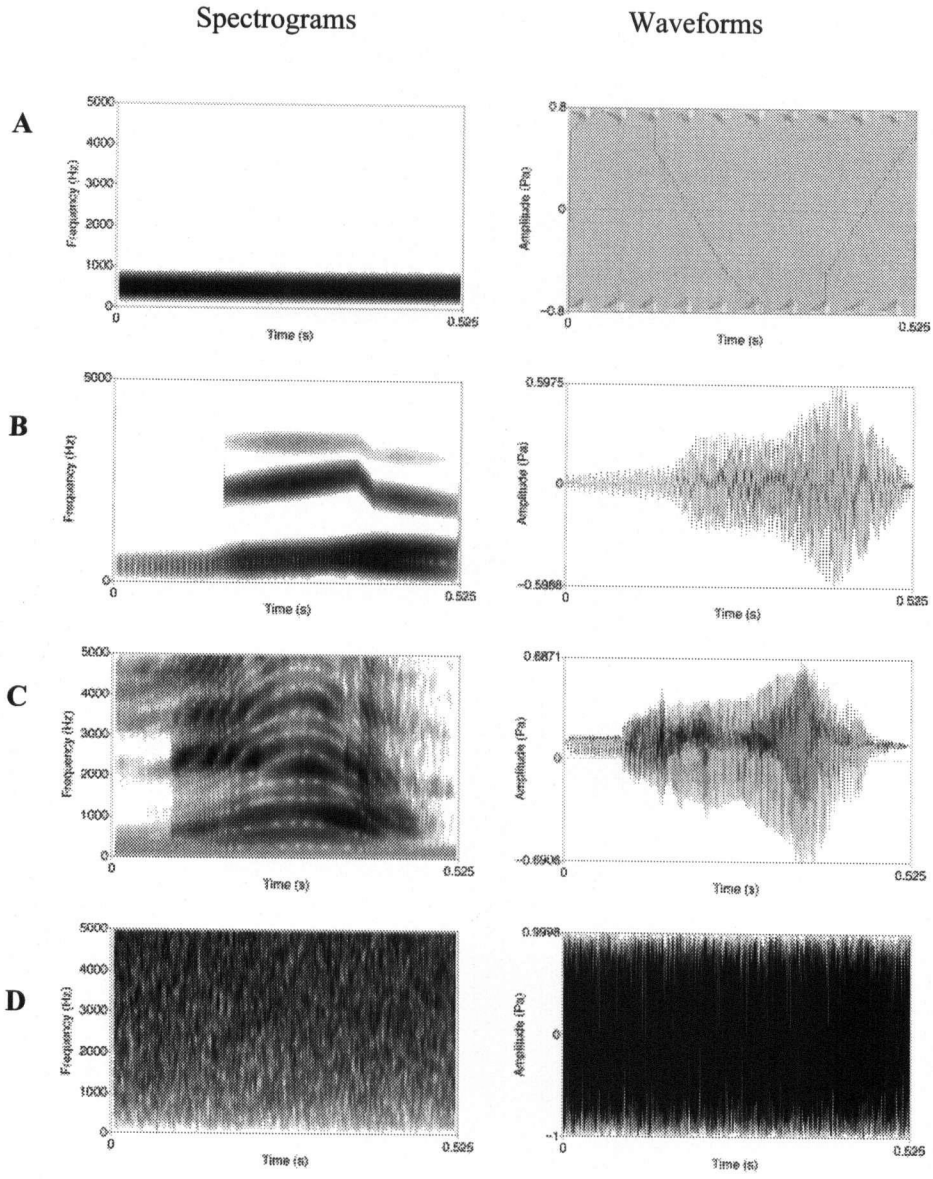

Figure 8. Comparison of acoustic stimuli differing in complexity.

Wide-band spectrograms (left) depict the change in frequency across time, and waveform diagrams illustrate the amplitude changes across time in pressure units (right); (A) single frequency tones, (B) sine-wave analogues, (C) speech, and (D) white noise. The changes in frequency across time seen in the speech signal are also observed in the sine-wave analogues. This time-varying property is absent from the other sounds.

The speech qualities are unique to speech as a biological sound produced by a human (species-specific) vocal tract. However, and crucially for the question asked in the current study, the non-speech analogues were designed to track changes across time for the peak frequencies of their speech counterparts, and in so doing, follow very closely the spectral and timing changes of natural speech.

Preserving the approximate changes in frequency and timing in our non-speech analogues allows us to test whether stimulus complexity can account fully for newborn infants' responsiveness, as suggested by previous studies (Clarkson & Berg, 1983; Groome et al., 2000; Hutt et al., 1968; Lenard et al., 1969; Turkewitz et al., 1972). For a stringent comparison, we further equated the two different signals for infant ears by creating analogues which retain the pitch contour of the speech counterparts, since pitch contour contributes importantly to infants' preference for infant-directed speech (Fernald & Kuhl, 1987), and discrimination of their native language (Mehler et al., 1988). Omitting pitch information from the non-speech analogues would render the comparison trivial, since this dimension alone would predispose the infants towards speech.

4.4.2 Preference experiment: Results

To assess whether infants show a listening bias for human speech, we used a contingent procedure that allows infants to become active participants in determining their auditory environment by controlling the presentation of auditory stimuli (Cooper & Aslin, 1990; Eimas et al., 1971). This measures the intrinsic reinforcing power of a stimulus. Using the HAS procedure (Eimas et al., 1971; Sansavini et al., 1997), in

Experiment Three we tested neonates by presenting them with alternating minutes of speech and non-speech stimuli, and measured the number of high amplitude sucks that infants delivered to elicit speech compared with non-speech stimuli (Sansavini et al., 1997). A 2 (sound type: speech vs. non-speech) x 2 (stimulus block: first 2 minute block vs. second 2 minute block) x 2 (sex: female vs. male) x 2 (trial order) mixed analysis of variance (ANOVA) indicated no main effect of sound type. Although there was no main effect of stimulus type, there was a significant interaction between stimulus type and experimental block ($F(1,20) = 9.087$, $p = .007$); a follow-up analysis of simple main effects showed that neonates sucked significantly more to listen to speech ($M = 72.4$, $SE = 3.5$) than to complex non-speech analogues ($M = 63.3$, $SE = 4.3$) in the second experimental block ($F(1,42) = 5.70$, $p = .022$; Table 3). Means in the first block were not significantly different from each other. These results show that over the course of the experiment, neonates adjusted their sucking rate to listen preferentially to speech, providing the first evidence that human infants are born with a bias for speech.

Table 3. Neonates' high amplitude sucking for speech and complex non-speech sounds.

Although there was no main effect of stimulus type, there was a significant interaction between stimulus type and experimental block ($p = .007$); a follow-up analysis of simple main effects showed that neonates sucked significantly more to listen to speech than to complex non-speech analogues in the second experimental block ($p = .022$). Means in the first block are not significantly different.

	Speech	Non-speech
Block 1	70.0 (5.0)	77.1 (4.6)
Block 2	72.4 (3.5)*	63.3 (4.3)*

4.5 Discrimination experiment

4.5.1 Discrimination experiment: Introduction

A fundamental question is whether the origin of the bias for listening to speech lies in a biologically specified perceptual predisposition or whether it reflects the effects of prior intra-uterine experience with human speech. Foetuses begin to respond to extra-uterine sounds by the 24th week of gestation, with consistent responses emerging at 28 weeks (Birnholz & Benacerraf, 1983). The developing foetus may thus be exposed to extra-uterine sounds for 12-16 weeks before birth. The mammalian amniotic sac and fluids act as low-pass filters, attenuating sounds above 400-500 Hz by 40-50 dB (Gerhardt & Abrams, 2000; Lecanuet, 1998). However the fundamental frequency of human speech ranges between 100-400 Hz, allowing the foetus exposure to the prosodic information in human speech. Newborn infants have been shown to use this prosodic information to discriminate their native language from rhythmically distinct languages (Mehler et al., 1988). Thus it is possible that neonates show an increasing interest in speech in this experiment based on their prenatal exposure to spoken language. Indeed prenatal exposure is thought to underlie other perceptual preferences in the newborn period, such as a preference for the mother's voice (DeCasper & Fifer, 1980), for stories and songs heard pre-natally (DeCasper & Spence, 1986), and even preference for the native language (Mehler et al., 1988; Moon et al., 1993). Similarly, pre-hatching

experience plays a critical role in some aspects of mallard ducklings' preference for their species-specific calls (Gottlieb, 1997).

Can prenatal exposure account for the preference for speech shown by newborn infants in the current study? To determine whether human infants' bias for speech is based on prenatal learning, in the Discrimination experiment (Experiment Four) we low-pass filtered (LPF) the speech and non-speech stimuli to approximate the frequency range audible in the intra-uterine environment (Gerhardt & Abrams, 2000; Lecanuet, 1998), and tested newborns on their ability to discriminate between these two sounds (Figure 8). For prenatal experience to be responsible for the preference for speech over complex non-speech, newborns should distinguish between (familiar) low-pass filtered speech sounds, and (unfamiliar) low-pass filtered complex non-speech, and therefore should succeed in a discrimination test. If newborns are unable to distinguish between the low-pass filtered versions of speech and non-speech, then they could not be recognising speech as a familiar prenatal stimulus compared to our complex non-speech control, and thus the preference we observe in the Preference experiment (Experiment Three) could not be based on prenatal experience. An explanation that is *not* based on prenatal exposure, then, predicts that newborns will fail to discriminate.

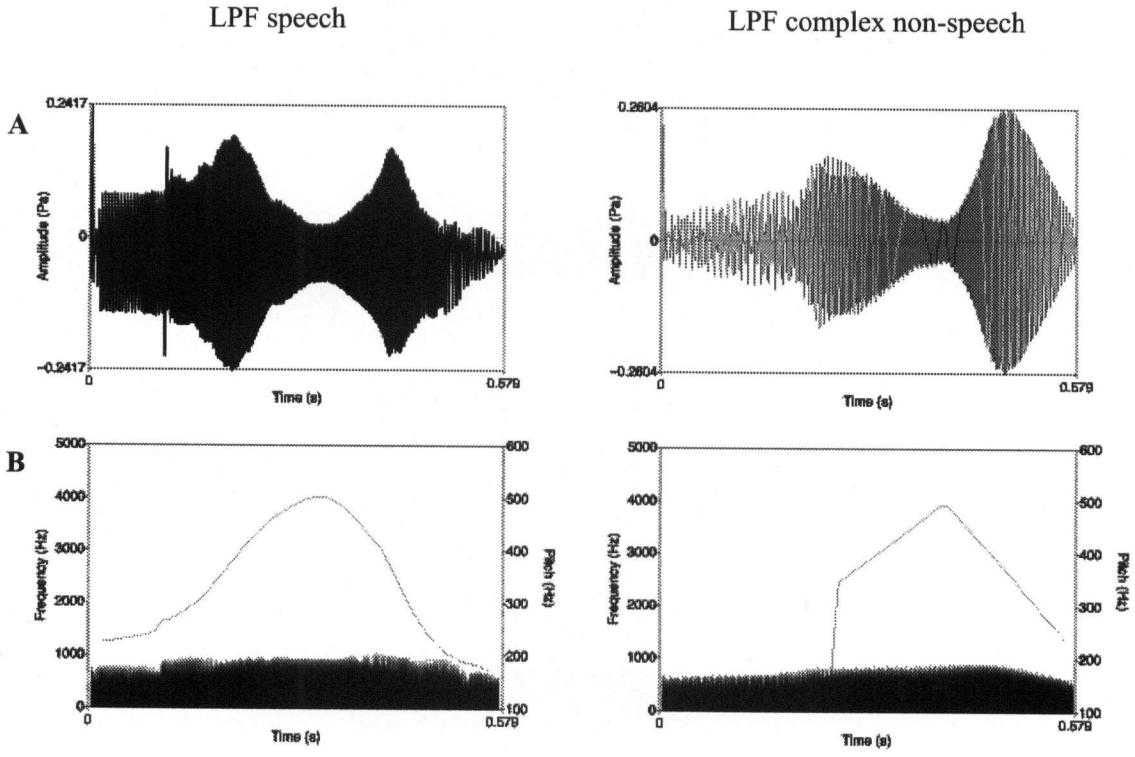

Figure 9. Low-pass filtered speech and non-speech sounds.

LPF speech (left) and LPF complex non-speech (right) illustrated as waveform diagrams (A) and spectrograms (B). LPF sounds retain only information below 400 Hz including the pitch contour (thin lines in B, pitch values in Hz on right axis).

Our working hypothesis that newborns may show a predisposition to prefer speech independent of prenatal experience would predict a failure to discriminate. A discrimination failure in a traditional habituation-dishabituation design would ground our claim on an awkward and unsatisfying foundation of negative evidence. A more suitable test of discrimination abilities is provided by a recently developed procedure which assesses discrimination by comparing infants' habituation slopes for different types of stimuli (Floccia, Christophe, & Bertoncini, 1997; Sansavini et al., 1997). When discriminable stimuli are presented in alternating minutes, newborns will maintain their sucking rate (a negative result), whereas when a single stimulus is repeated, newborns' sucking rates decrease significantly (a positive result, Sansavini et al., 1997). We tested newborns on their ability to discriminate between LPF stimuli by presenting them with alternating minutes of LPF speech and LPF non-speech. If newborns treat LPF speech and LPF non-speech as discriminable stimuli, they should maintain their sucking rate. If, however, LPF speech and LPF non-speech are not discriminable, newborns should show a *significant* decrease in their sucking rate. A failure to discriminate would then result in a *positive* finding. To ensure that this procedure could indeed index discrimination abilities, two control conditions were run. In one control condition, a new group of neonates heard stimuli of a single type, either speech or complex non-speech, repeated for the entire experiment. Infants in a condition with repeating stimuli are expected to become bored and decrease their sucking rates. In the other control condition, newborns were tested on alternating trials of (unfiltered) speech and complex non-speech. Because newborns had already demonstrated a preference for speech compared to non-speech

stimuli in the Preference experiment, they must be able to discriminate between the two types of stimuli. Thus, infants in this condition were expected to maintain their sucking rates.

4.5.2 Discrimination experiment: Results

As predicted, infants exposed to repeating minutes of either speech or non-speech habituated to the repeating stimuli and showed a significant decline in sucking rate (Pearson's Correlation, $r = -.173$, $p = .029$), while infants hearing the unfiltered speech and non-speech sounds in alternating trials maintained high sucking rates ($r = -.024$, $p > .7$), as expected for two discriminable sounds. In the critical LPF condition, infants hearing alternating trials of LPF speech and LPF complex non-speech showed a significant decrement in sucking ($r = -.223$, $p = .005$), behaving like the infants who heard a single repeating stimulus (Figure 9). The significant decline in sucking reveals an inability to discriminate the two types of sounds. Newborn infants treated LPF speech and LPF complex non-speech as equivalent.[10]

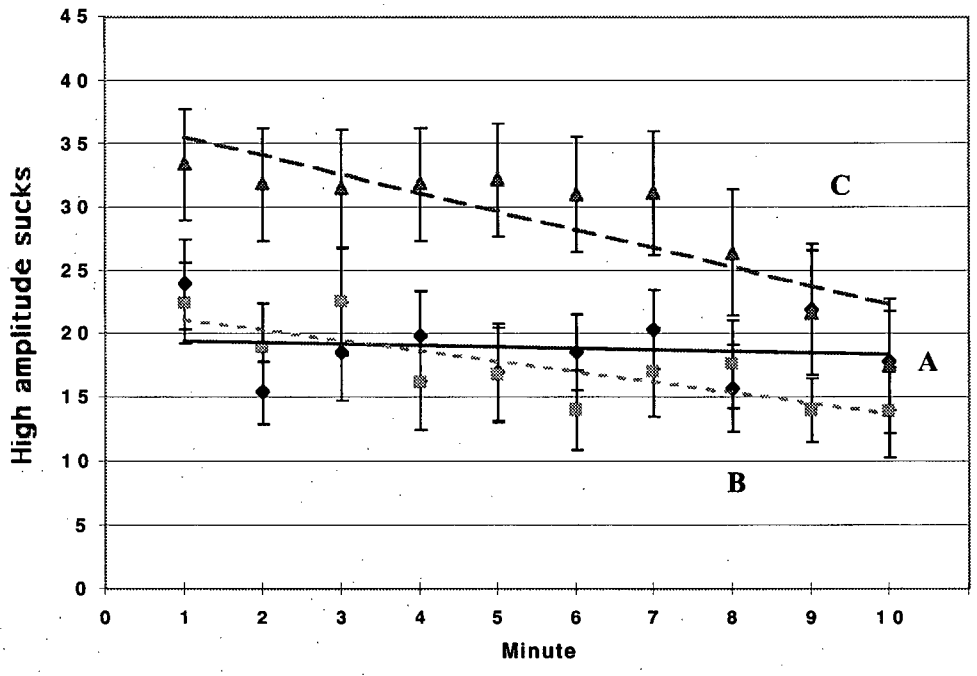

Figure 10. Habituation slopes of infants in different auditory conditions.

(A) Alternating trials of speech and non-speech (diamonds and solid line), (B) repeating trials of either unfiltered speech or non-speech (squares and dotted line), and (C) alternating trials of LPF speech and LPF non-speech (triangles and dashed line). As predicted, infants exposed to repeating minutes of either speech or non-speech habituated to the repeating stimuli. In contrast, infants hearing the unfiltered speech and non-speech sounds in alternating trials maintained high sucking rates, as expected for two discriminable sounds. In the critical LPF condition, infants hearing alternating trials of LPF speech and LPF complex non-speech showed a significant decrement in sucking, behaving like the infants who heard a single repeating stimulus.

A cursory examination of Figure 9 suggests possible differences in the overall number of HAS between infants in the three conditions in the discrimination experiment. A one-way ANOVA of overall HAS confirmed a difference between groups ($F(2,46) = 3.528, p = .038$). Post-hoc tests of Least Squares Differences revealed that newborns in the condition with LPF sounds sucked more overall than the newborns in either the repeating control group ($t(46) = 2.136, p = .038$) or the unfiltered alternating control group ($t(46) = 2.445, p = .018$). This would suggest a preference for the LFP sounds. An alternative explanation is that infants in the LPF group were differentially aroused prior to beginning the study. However, a one-way ANOVA of HAS delivered during a silent baseline minute that precedes the delivery of sounds revealed no differences in the number of HAS between the three groups ($F(2,45) = .282, p > .7$) suggesting that there were no differences in the HAS behaviour of infants prior to the presentation of sounds. In light of other perceptual preferences that newborn infants show for familiar sounds such as the mother's voice and the native language, newborns' higher sucking rates for filtered sounds that mimic prenatally audible sounds is not surprising. Indeed, the HAS patterns elicited in Experiment Four simultaneously confirm the influence of prenatal experience as seen in the overall higher sucking rate to LPF sounds, and substantiate the hypothesis that newborns' preference for (unfiltered) speech over non-speech is independent of prenatal experience.

An alternative explanation might suggest that the high rate of the sucking in the low pass filtered condition would be difficult to maintain for the duration of the experimental session. According to this hypothesis, the infants in the low-pass filtered

condition did not significantly decrease their sucking because they could not discriminate between the two different low-pass filtered sounds, but because they were unable to maintain a high sucking rate. This explanation is extremely unlikely as newborns presented with variable stimuli in high amplitude sucking studies typically maintain sucking rates of more than 40 sucks/min for at least 10-11 minutes (Floccia et al., 1997; Sansavini et al., 1997). The most likely explanation for neonates' sucking decrement is therefore their failure to discriminate between the sounds. Neonates' inability to discriminate speech from complex non-speech when the stimuli have been filtered to approximate intra-uterine sounds confirms that the infants could not have recognised (unfiltered) speech as a familiar stimulus compared with the non-speech analogues. The preference that newborn infants show for unfiltered speech in the Preference experiment, therefore, cannot be explained by their familiarity with speech stimuli through prenatal exposure, suggesting that the bias for speech reflects an inborn predisposition.

4.6 Discussion

This study provides the first demonstration that human neonates are biased to listen to speech, and that this preference is based on a biologically-specified predisposition. Like other animals, human newborns show initial proclivities which direct them towards particular types of information (Marler, 1990). Animals from which experience has been withheld show patterns of behaviour that suggest some degree of innate specification. For example, ducklings deprived of auditory stimulation before hatching (hearing not even their own vocalisations) prefer to approach their own species'

call compared with the calls of other birds, and are fooled only when heterospecific calls share the frequencies and repetition rate of their own species' call (Gottlieb, 1997). Similarly, chicks raised in the dark prefer to approach stuffed hens rather than less naturalistic objects (Johnson et al., 1985). Experience-independent visual preferences have been characterised in humans as well, with newborns showing preferences for face-like stimuli (Johnson et al., 1991) and for direct eye gaze compared with averted eye gaze (Farroni, Csibra, Simion, & Johnson, 2002). However, in the auditory domain, previous studies with human infants have demonstrated only experience-based preferences such as the attraction to the mother's voice and to their native language. Our results are the first to show an auditory preference in human newborns that is not based on prenatal experience.

Newborn infants could be drawn to speech because of its biological quality, and/or because of its source in the human vocal tract. The attraction to spoken language persists into the first few months of life (Vouloumanos & Werker, in press), and may extend beyond spoken language to communicative gestures in other modalities (Hildebrandt, 2003). Initial biases may be elaborated by experience to refine the perceptual preferences of developing organisms (Gottlieb, 1997). An initial orientation to speech could facilitate the pick-up of distributional information in the speech signal that specifies, for example, native language phonetic categories (Maye, Werker, & Gerken, 2002) and native phonotactics (Jusczyk, Friederici et al., 1993).

The ways in which initial predispositions are specified in biological substrates is under current investigation. Neuroimaging studies suggest that cortical substrates in the

temporal lobes, specifically along the superior temporal sulcus, differentiate between speech and non-speech sounds (Belin et al., 2000; Benson et al., 2001; Binder et al., 2000; Scott et al., 2000; Vouloumanos et al., 2001). Specific areas in the temporal lobe are recruited during the processing of speech compared with non-speech even in young infants (Dehaene-Lambertz, 2000; Dehaene-Lambertz et al., 2002) and in newborns (Peña et al., 2003). Single-cell recordings in other primates suggest that neurons in these areas are activated specifically by species-specific vocalisations (Wang et al., 1995). The origins of a speech bias in humans may lie in the precocious tuning of neurons in the superior temporal sulcus to respond to signals used for conspecific communication.

Mastering a language is arguably one of the most complex feats of human learning requiring the interplay of innate biases and sophisticated learning machinery. The present study demonstrates that human infants begin language acquisition with a bias for listening to speech at birth, helping them to focus on speech as the grist for linguistic discovery.

Chapter 5. General discussion

This thesis provides two kinds of evidence for considering speech a special sound for humans. In Experiment One, event-related functional MRI that models cortical activation for individual sounds demonstrated that specific areas of the left hemisphere of adults are involved in detecting speech compared with non-speech sounds matched in frequency and timing characteristics. That this differential activation is evident even when participants are not required to perform linguistic analyses on the stimuli suggests that specialised substrates are recruited automatically by the very nature of speech.

Such results do not speak, however, to the developmental origins of the specialisation for speech. As a means towards better understanding the roots of the human specialisation for speech, a different approach was taken in Experiments Two through Four. Inspired by findings from the ethological literature reporting predispositions in young animals for the visual and vocal characteristics of biologically important stimuli, speech and non-speech stimuli were used to investigate the biases of young infants for speech. Experiment Two demonstrated that 2-7 month olds selectively listen longer to speech than non-speech sounds. Young humans show the same kinds of biases that other animals have for listening to the vocalisations of conspecifics. In Experiment Three the preference of newborns was tested. Neonates demonstrated a preference for speech, suggesting that the origins of the bias may be prenatal.

Though neonates have had little prenatal visual experience, they have already experienced auditory input in the womb, inviting the question of whether the speech bias is experience-based or innate. The prenatal environment acts as a low-pass filter,

allowing human foetuses to hear mainly the low frequency information in extra-uterine sounds, including speech (reviewed in Abrams & Gerhardt, 2000; Gerhardt & Abrams, 2000). The similarities between the speech and non-speech stimuli, particularly in the low frequency range, suggest that they may not be discriminable on the basis of the information available to foetuses. If prenatal experience were responsible for the preference for speech over non-speech, newborns should be to able discriminate between (familiar) low-pass filtered speech sounds, and (unfamiliar) low-pass filtered complex non-speech. Instead, Experiment Four showed that newborns failed to discriminate between speech and non-speech sounds when the sounds were low-pass filtered to mimic conditions in the womb. This suggests that foetuses do not have access to the information that would differentiate between speech and non-speech stimuli. The results of these experiments demonstrate that neonates have a bias for listening to speech that appears independent of specific experience.

These findings directly inform the traditional debate about the special status of speech for humans. Human infants are prewired to preferentially attend to speech, granting it a special status in relation to other sounds. Much of the earlier work posited specialised neural substrates and specialised processing mechanisms (reviewed in Trout, 2001). The specialised neural substrates recruited by speech compared with matched non-speech sounds in Experiment One provide solid evidence to buttress the traditional argument. The developmental studies add a new perspective to this traditional argument, demonstrating that young infants have specialised and experience-independent biases for listening to speech.

The conclusion that the speech bias is independent of experience does not suggest that a hypothetical human developing in a vacuum would demonstrate such a bias. As suggested in Chapter One, the claim for experience-independence should not be taken to mean that *no* experience is necessary, simply that no *specific* experience is necessary. This general point about the nature of "experience" has been investigated in the case of the chick predisposition for the head and shoulder configuration of fowl. If chicks are raised in conditions of complete sensory deprivation (in which tactile, auditory, and visual stimulation are withheld), they fail to demonstrate the preference for the stuffed jungle fowl configuration (Bolhuis et al., 1985). It appears that some stimulation in any modality is sufficient, and necessary, for this innate predisposition to emerge. But the need for stimulation does not mean that specific experience is required. Being in an environment that provides general dynamic non-specific environmental input may allow internally-generated patterns of neural activity (Katz & Shatz, 1996) to express the organism's genetic program. As such, general stimulation that is provided by the species-typical environment may permit the expression of innate biases; in this case, the environment is a permissive, rather than an instructive, influence (a similar point is made in Morton & Johnson, 1991). An alternative possibility is that non-specific but *related* experience may be required to refine the preferences observed at birth (Gottlieb, 1980; Werker & Tees, 1992); for example, it might be necessary for the foetus to be stimulated by auditory input at specific frequencies in order for an innate bias to emerge post-natally.

Some open questions remain concerning the nature of the speech bias. The bias may be narrow and specific to human communication sounds. Alternatively, infants' bias(es) could be more broadly specified. For example, infants may be biased to listen preferentially and indiscriminately to all sounds of biological origin, including the calls of other species. Alternatively, sounds of human origin, including sighs and snores, could be interesting for infants. This bias could be further characterised by contrasting infants' listening preferences for human speech with non-human vocalisations as well as human non-speech sounds. Determining the scope of this bias is an important endeavour for future research.

Just as there are parallels between the biases for head and shoulder configurations in chicks and faces in humans, there may also be similarities between the vocal biases of humans and other species. If so, we may expect the bias for speech to be quite narrowly construed. An apparent difference between face (or configuration) perception and the perception of vocalisations is that configural information appears more broadly prespecified, allowing young infants to orient to structural information available in many species, while the biases for conspecific vocalisations appear to be more finely described. For example, ducklings have a narrowly specified range for preferred vocalisations, and only become fooled when the heterospecific sounds are very similar in timing and frequency to conspecific vocalisations (Gottlieb, 1997). If the principles of analogy for face (or configuration) perception biases hold true, *mutatis mutandis*, for vocal perception, the speech bias may conceivably be a genuine domain-specific adaptation for speech.

References

Abrams, R. M., & Gerhardt, K. J. (2000). The acoustic environment and physiological responses of the fetus. *Journal of Perinatology, 20*(8 Pt 2), S31-36.

Alegria, J., & Noirot, E. (1978). Neonate orientation behaviour towards human voice. *International Journal of Behavioral Development, 1*(4), 291-312.

Annett, M. (1967). The binomial distribution of right, mixed, and left handedness. *Quarterly Journal of Experimental Psychology, 19*, 327-333.

Baylis, G. C., Rolls, E. T., & Leonard, C. M. (1987). Functional subdivisions of the temporal lobe neocortex. *Journal of Neuroscience, 7*(2), 330-342.

Bednar, J. A., & Miikkulainen, R. (2000). Self-organization of innate face preferences: could genetics be expressed through learning? *Proceedings of the 17th National Conference on Artificial Intelligence*, 117-122.

Belin, P., Zatorre, R. J., Hoge, R., Evans, A. C., & Pike, B. (1999). Event-related fMRI of the auditory cortex. *Neuroimage, 10*(4), 417-429.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*(6767), 309-312.

Belin, P., Zilbovicius, M., Crozier, S., Thivard, L., Fontaine, A., Masure, M. C., & Samson, Y. (1998). Lateralization of speech and auditory temporal processing. *Journal of Cognitive Neuroscience, 10*(4), 536-540.

Benson, R. R., Whalen, D. H., Richardson, M., Swainson, B., Clark, V. P., Lai, S., & Liberman, A. M. (2001). Parametrically dissociating speech and nonspeech perception in the brain using fMRI. *Brain and Language, 78*(3), 364-396.

Bertoncini, J., Bijeljac-Babic, R., Blumstein, S. E., & Mehler, J. (1987). Discrimination in neonates of very short CVs. *Journal of the Acoustical Society of America, 82*(1), 31-37.

Bertoncini, J., & Mehler, J. (1981). Syllables as units in infant speech perception. *Infant Behavior and Development, 4*(3), 247-260.

Bertoncini, J., Morais, J., Bijeljac-Babic, R., McAdams, S., Peretz, I., & Mehler, J. (1989). Dichotic perception and laterality in neonates. *Brain and Language, 37*(4), 591-605.

Best, C. T., Hoffman, H., & Glanville, B. B. (1982). Development of infant ear asymmetries for speech and music. *Perception and Psychophysics, 31*(1), 75-85.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., & Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex, 10*(5), 512-528.

Binder, J. R., Frost, J. A., Hammeke, T. A., Rao, S. M., & Cox, R. W. (1996). Function of the left planum temporale in auditory and linguistic processing. *Brain, 119*(Pt 4), 1239-1247.

Binder, J. R., & Price, C. (2001). Functional neuroimaging of language. In R. Cabeza & A. Kingstone (Eds.), *Handbook of Functional Neuroimaging of Cognition* (pp. 187-251). Cambridge, MA: MIT Press.

Binder, J. R., Rao, S. M., Hammeke, T. A., Frost, J. A., Bandettini, P. A., & Hyde, J. S. (1994). Effects of stimulus rate on signal response during functional magnetic resonance imaging of auditory cortex. *Brain Research. Cognitive Brain Research, 2*(1), 31-38.

Binder, J. R., Rao, S. M., Hammeke, T. A., Yetkin, F. Z., Jesmanowicz, A., Bandettini, P. A., Wong, E. C., Estkowski, L. D., Goldstein, M. D., Haughton, V. M., & al., e. (1994). Functional magnetic resonance imaging of human auditory cortex. *Annals of Neurology, 35*(6), 662-672.

Birnholz, J. C., & Benacerraf, B. R. (1983). The development of human fetal hearing. *Science, 222*(4623), 516-518.

Blumstein, S. E. (1994). Impairments of speech production and speech perception in aphasia. *Philosophical transactions of the Royal Society of London. Series B: Biological sciences, 346*(1315), 29-36.

Blumstein, S. E., & Milberg, W. P. (2000). Neural systems and language processing: toward a synthetic approach. *Brain and Language, 71*(1), 26-29.

Boehner, J. (1990). Early acquisition of song in the zebra finch, Taeniopygia guttata. *Animal Behaviour, 39*(2), 369-374.

Bolhuis, J. J., & Honey, R. C. (1998). Imprinting, learning and development: from behaviour to brain and back. *Trends in Neurosciences, 21*(7), 306-311.

Bolhuis, J. J., Johnson, M. H., & Horn, G. (1985). Effects of early experience on the development of filial preferences in the domestic chick. *Developmental Psychobiology, 18*(4), 299-308.

Braaten, R. F., & Reynolds, K. (1999). Auditory preference for conspecific song in isolation-reared zebra finches. *Animal Behaviour, 58*(1), 105-111.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound.* Cambridge, MA, US: The MIT Press.

Bushnell, I. W., Sai, F., & Mullin, J. T. (1989). Neonatal recognition of the mother's face. *British Journal of Developmental Psychology, 7*(1), 3-15.

Butterfield, E. C., & Siperstein, G. N. (1970). Influence of contingent auditory stimulation upon non-nutritional suckle. In J. F. Bosma (Ed.), *Third Symposium on Oral Sensation and Perception: The Mouth of the Infant* (pp. 313-334). Springfield, IL: Charles C. Thomas.

Cassia, V. M., Simion, F., & Umilta, C. (2001). Face preference at birth: The role of an orienting mechanism. *Developmental Science, 4*(1), 101-108.

Celsis, P., Boulanouar, K., Doyon, B., Ranjeva, J. P., Berry, I., Nespoulous, J. L., & Chollet, F. (1999). Differential fMRI responses in the left posterior superior temporal gyrus and left supramarginal gyrus to habituation and change detection in syllables and tones. *Neuroimage, 9*(1), 135-144.

Cheour, M., Alho, K., Ceponiene, R., Reinikainen, K., Sainio, K., Pohjavuori, M., Aaltonen, O., & Naatanen, R. (1998). Maturation of mismatch negativity in infants. *International Journal of Psychophysiology, 29*, 217-226.

Cheour, M., Alho, K., Sainio, K., Reinikainen, K., Renlund, M., Aaltonen, O., Eerola, O., & Naatanen, R. (1997). The mismatch negativity to changes in speech sounds at the age of three months. *Developmental Neuropsychology, 13*(2), 167-174.

Chomsky, N. (1975). *Reflections on language*. New York: Pantheon Books.

Chomsky, N. (1986). Knowledge of language as a focus of inquiry, *Knowledge of Language: Its Nature, Origin, and Use* (pp. 1-14).

Chomsky, N. (2000). *New horizons in the study of language and mind*. Cambridge, UK; New York, NY: Cambridge University Press.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *Journal of the Acoustical Society of America, 95*(3), 1570-1580.

Clarkson, M. G., & Berg, W. K. (1983). Cardiac orienting and vowel discrimination in newborns: crucial stimulus parameters. *Child Development, 54*(1), 162-171.

Cohen, L. B. (1972). Attention-getting and attention-holding processes of infant visual preferences. *Child Development, 43*(3), 869-879.

Cohen, L. B., DeLoache, J. S., & Rissman, M. W. (1975). The effect of stimulus complexity on infant visual attention and habituation. *Child Development, 46*(3), 611-617.

Cole, R. A., & Jakimik, J. (1980). How are syllables used to recognize words? *Journal of the Acoustical Society of America, 67*(3), 965-970.

Colombo, J. A., & Bundy, R. S. (1981). A method for the measurement of infant auditory selectivity. *Infant Behavior and Development, 4*(2), 219-223.

Condon, W. S., & Sander, L. W. (1974). Neonate movement is synchronized with adult speech: interactional participation and language acquisition. *Science, 183*(120), 99-101.

Cooper, R. P., Abraham, J., Berman, S., & Staska, M. (1997). The development of infants' preference for motherese. *Infant Behavior and Development, 20*(4), 477-488.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development, 61*, 1584-1595.

Cooper, R. P., & Aslin, R. N. (1994). Developmental differences in infant attention to the spectral properties of infant-directed speech. *Child Development, 65*(6), 1663-1677.

D'Esposito, M., Zarahn, E., & Aguirre, G. K. (1999). Event-related functional MRI: implications for cognitive psychology. *Psychological Bulletin, 125*(1), 155-164.

Dale, A. M. (1999). Optimal experimental design for event-related fMRI. *Human Brain Mapping, 8*(2-3), 109-114.

Damasio, A. R., & Damasio, H. (2000). Aphasia and the neural basis of language. In M. M. Mesulam (Ed.), *Principles of behavioral and cognitive neurology* (2nd ed., pp. 294-315). New York, NY, US: Oxford University Press.

Damasio, A. R., & Geschwind, N. (1984). The neural basis of language. *Annual Review of Neuroscience 7*, 127-147.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: newborns prefer their mothers' voices. *Science, 208*(4448), 1174-1176.

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development, 9*, 133-150.

Dehaene-Lambertz, G. (2000). Cerebral specialization for speech and non-speech stimuli in infants. *Journal of Cognitive Neuroscience, 12*(3), 449-460.

Dehaene-Lambertz, G., & Baillet, S. (1998). A phonological representation in the infant brain. *NeuroReport, 9*(8), 1885-1888.

Dehaene-Lambertz, G., & Dehaene, S. (1994). Speed and cerebral correlates of syllable discrimination in infants. *Nature, 370*(6487), 292-295.

Dehaene-Lambertz, G., Dehaene, S., & Hertz-Pannier, L. (2002). Functional neuroimaging of speech perception in infants. *Science, 298*(5600), 2013-2015.

Demonet, J. F., Chollet, F., Ramsay, S., Cardebat, D., Nespoulous, J. L., Wise, R., Rascol, A., & Frackowiak, R. (1992). The anatomy of phonological and semantic processing in normal subjects. *Brain, 115*(6), 1753-1768.

Demonet, J. F., Fiez, J. A., Paulesu, E., Petersen, S. E., & Zatorre, R. J. (1996). PET studies of phonological processing: A critical reply to Poeppel. *Brain and Language, 55*(3), 352-379.

Diehl, R. L., & Kluender, K. R. (1986). On the objects of speech perception. *Ecological Psychology, 1*(2), 121.

Diehl, R. L., & Kluender, K. R. (1987). On the categorization of speech sounds. In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition* (pp. 226-253). New York, NY, USA: Cambridge University Press.

Diehl, R. L., & Kluender, K. R. (1989a). On the objects of speech perception. *Ecological Psychology, 1*(2), 121-144.

Diehl, R. L., & Kluender, K. R. (1989b). Reply to commentators. *Ecological Psychology, 1*(2), 195-225.

Dooling, R. J., Okanoya, K., & Brown, S. D. (1989). Speech perception by budgerigars (Melopsittacus undulatus): The voiced/voiceless distinction. *Perception and Psychophysics, 46*(1), 65-71.

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience, 22*, 567-631.

Dowd, J. M., & Tronick, E. Z. (1986). Temporal coordination of arm movements in early infancy: do infants move in synchrony with adult speech? *Child Development, 57*(3), 762-776.

Easterbrook, M. A., Kisilevsky, B. S., Hains, S. M. J., & Muir, D. W. (1999). Faceness or complexity: Evidence from newborn visual tracking of facelike stimuli. *Infant Behavior and Development, 22*(1), 17-35.

Ecklund-Flores, L., & Turkewitz, G. (1996). Asymmetric headturning to speech and nonspeech in human newborns. *Developmental Psychobiology, 29*(3), 205-217.

Eimas, P. D., & Miller, J. L. (1980). Contextual effects in infant speech perception. *Science, 209*(4461), 1140-1141.

Eimas, P. D., & Miller, J. L. (1992). Organization in the perception of speech by young infants. *Psychological Science, 3*(6), 340-345.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science, 171*(968), 303-306.

Eisenberg, R. B. (1976). *Auditory competence in early life*. Baltimore, MD: University Park Press.

Entus, A. K. (1977). Hemispheric asymmetry in processing of dichotically presented speech and nonspeech by infants. In S. Segalowitz & F. Gruber (Eds.), *Language development and neurological theory* (pp. 63-73). New York: Academic Press.

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences (USA), 99*(14), 9602-9605.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development, 8*(2), 181-195.

Fernald, A., & Kuhl, P. K. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development, 10*(3), 279-293.

Fiez, J. A., Raichle, M. E., Miezin, F. M., Petersen, S. E., Tallal, P., & Katz, W. F. (1995). PET studies of auditory and phonological processing: effects of stimulus characteristics and task demands. *Journal of Cognitive Neuroscience, 7*(3), 357-375.

Floccia, C., Christophe, A., & Bertoncini, J. (1997). High-amplitude sucking and newborns: the quest for underlying mechanisms. *Journal of Experimental Child Psychology, 64*, 175-198.

Fodor, J. A. (1983). *The modularity of mind: an essay on faculty psychology*. Cambridge, MA: MIT Press.

Fowler, C. A., & Rosenblum, L. D. (1990). Duplex perception: a comparison of monosyllables and slamming doors. *Journal of Experimental Psychology: Human Perception and Performance, 16*(4), 742-754.

Galaburda, A., & Sanides, F. (1980). Cytoarchitectonic organization of the human auditory cortex. *Journal of Comparative Neurology, 190*(3), 597-610.

Galaburda, A. M., Sanides, F., & Geschwind, N. (1978). Human brain. Cytoarchitectonic left-right asymmetries in the temporal speech region. *Archives of Neurology, 35*(12), 812-817.

Gerhardt, H. C. (1988). Acoustic properties used in call recognition by frogs and toads. In B. Fritzsch & M. J. Ryan & W. Winlczynski & T. Hetherington & W. Walkowiak (Eds.), *The evolution of the amphibian auditory system* (pp. 253-273). New York: J. Wiley.

Gerhardt, K. J., & Abrams, R. M. (2000). Fetal exposures to sound and vibroacoustic stimulation. *Journal of Perinatology, 20*(8 Pt 2), S21-30.

Geschwind, N. (1965). Disconnexion syndromes in animals and man. I. *Brain, 88*(2), 237-294.

Geschwind, N., & Levitsky, W. (1968). Human brain: left-right asymmetries in temporal speech region. *Science, 161*(837), 186-187.

Glenn, S. M., Cunningham, C. C., & Joyce, P. F. (1981). A study of auditory preferences in nonhandicapped infants and infants with Down's syndrome. *Child Development, 52*(4), 1303-1307.

Goren, C. C., Sarty, M., & Wu, P. Y. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics, 56*(4), 544-549.

Gottlieb, G. (1980). Development of species identification in ducklings: VI. Specific embryonic experience required to maintain species-typical perception in ducklings. *Journal of Comparative Physiology and Psychology, 94*(4), 579-587.

Gottlieb, G. (1997). *Synthesizing nature-nurture: Prenatal roots of instinctive behavior.* Mahwah, NJ, US: Lawrence Erlbaum Associates, Inc., Publishers.

Gould, J. L., & Marler, P. (1987). Learning by instinct. *Scientific American, 256*(1), 74-85.

Grady, C. L., Van Meter, J. W., Maisog, J. M., Pietrini, P., Krasuski, J., & Rauschecker, J. P. (1997). Attention-related modulation of activity in primary and secondary auditory cortex. *NeuroReport, 8*(11), 2511-2516.

Groome, L. J., Mooney, D. M., Holland, S. B., Smith, Y. D., Atterbury, J. L., & Dykman, R. A. (2000). Temporal pattern and spectral complexity as stimulus parameters for eliciting a cardiac orienting reflex in human fetuses. *Perception and Psychophysics, 62*(2), 313-320.

Hashimoto, R., Homae, F., Nakajima, K., Miyashita, Y., & Sakai, K. L. (2000). Functional differentiation in the human auditory and language areas revealed by a dichotic listening task. *Neuroimage, 12*(2), 147-158.

Hauser, M. D. (1996). *The evolution of communication*. Cambridge, MA, US: The MIT Press.

Hickok, G., Love, T., Swinney, D., Wong, E. C., & Buxton, R. B. (1997). Functional MR imaging during auditory word perception: a single-trial presentation paradigm. *Brain and Language, 58*(1), 197-201.

Hildebrandt, U. (2003). *An investigation of infant preferences for American Sign Language and nonlinguistic biological motion.* University of Washington, Seattle, WA.

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in Infancy Research, 5*, 69-95.

Hutt, S. J., Hutt, C., Lenard, H. G., van Bernuth, H., & Muntjewerff, W. J. (1968). Auditory responsivity in the human neonate. *Nature, 218*, 888-890.

Johnson, M. H. (1999). Ontogenetic constraints on neural and behavioral plasticity: evidence from imprinting and face processing. *Canadian Journal of Experimental Psychology, 53*(1), 77-91.

Johnson, M. H., Bolhuis, J. J., & Horn, G. (1985). Interaction between acquired preferences and developing predispositions during imprinting. *Animal Behaviour, 33*(3), 1000-1006.

Johnson, M. H., Bolhuis, J. J., & Horn, G. (1992). Predispositions and learning: Behavioural dissociations in the chick. *Animal Behaviour, 44*(5), 943-948.

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition, 40*(1-2), 1-19.

Johnson, M. H., & Horn, G. (1986). Dissociation of recognition memory and associative learning by a restricted lesion of the chick forebrain. *Neuropsychologia, 24*(3), 329-340.

Johnson, M. H., & Horn, G. (1988). Development of filial preferences in dark-reared chicks. *Animal Behaviour, 36*(3), 675-683.

Johnsrude, I. S., Zatorre, R. J., Milner, B. A., & Evans, A. C. (1997). Left-hemisphere specialization for the processing of acoustic transients. *NeuroReport, 8*(7), 1761-1765.

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA, USA: MIT Press.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology, 29*(1), 1-23.

Jusczyk, P. W., & Bertoncini, J. (1988). Viewing the development of speech perception as an innately guided learning process. *Language and Speech, 31*(Pt 3), 217-238.

Jusczyk, P. W., Bertoncini, J., Bijeljac Babic, R., Kennedy, L. J., & Mehler, J. (1990). The role of attention in speech perception by young infants. *Cognitive Development, 5*(3), 265-286.

Jusczyk, P. W., Cutler, A., & Redanz, N. J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development, 64*(3), 675-687.

Jusczyk, P. W., Friederici, A. D., Wessels, J. M., & Svenkerud, V. Y. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language, 32*(3), 402-420.

Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Perception and Psychophysics, 61*(8), 1465-1476.

Jusczyk, P. W., Pisoni, D. B., Reed, M. A., Fernald, A., & Myers, M. (1983). Infants' discrimination of the duration of a rapid spectrum change in nonspeech signals. *Science, 222*(4620), 175-177.

Jusczyk, P. W., Pisoni, D. B., Walley, A., & Murray, J. (1980). Discimination of relative onset time of two-component tones by infants. *Journal of the Acoustical Society of America, 67*(1), 262-270.

Kaplan, J., & Russell, M. (1974). Olfactory recognition in the infant squirrel monkey. *Developmental Psychobiology, 7*(1), 15-19.

Kato, T., Takahashi, E., Sawada, K., Kobayashi, N., Watanabe, T., & Ishii, T. (1983). A computer analysis of infant movements synchronized with adult speech. *Pediatric Research, 17*(8), 625-628.

Katz, L. C., & Shatz, C. J. (1996). Synaptic activity and the construction of cortical circuits. *Science, 274*(5290), 1133-1138.

Kiehl, K. A., Laurens, K. R., Duty, T. L., Forster, B. B., & Liddle, P. F. (2001). Neural sources involved in auditory target detection and novelty processing: an event-related fMRI study. *Psychophysiology, 38*(1), 133-142.

Kimura, D. (1967). Functional asymmetry of the brain in dichotic listening. *Cortex, 3*, 163-178.

Kleiner, K. A. (1987). Amplitude and phase spectra as indices of infants' pattern preferences. *Infant Behavior and Development, 10*(1), 49-59.

Kluender, K. R. (1994). Speech perception as a tractable problem in cognitive science. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics.* (pp. 173-217). San Diego, CA, US: Academic Press, Inc.

Kluender, K. R., Diehl, R. L., & Killeen, P. R. (1987). Japanese quail can learn phonetic categories. *Science, 237*(4819), 1195-1197.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics, 50*(2), 93-107.

Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. *Science, 190*(4209), 69-72.

Kuhl, P. K., & Padden, D. M. (1982). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. *Perception and Psychophysics, 32*(6), 542-550.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*(5044), 606-608.

Ladefoged, P. (1993). *A course in phonetics* (3rd ed.). Forth Worth, USA: Harcourt Brace Jovanovich College Publishers.

Lecanuet, J. P. (1998). Foetal responses to auditory and speech stimuli. In A. Slater (Ed.), *Perceptual development: visual, auditory, and speech perception in infancy* (pp. 317-355). East Sussex, UK: Psychology Press.

Lenard, H. G., von Bernuth, H., & Hutt, S. J. (1969). Acoustic evoked responses in newborn infants: the influence of pitch and complexity of the stimulus. *Electroencephalography and Clinical Neurophysiology, 27*(2), 121-127.

Lenneberg, E. (1967). *Biological foundations of language*. New York: Wiley.

Liberman, A. M. (1996). *Speech: A special code*. Cambridge, MA, USA: The Mit Press.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*, 358-368.

Liberman, A. M., Isenberg, D., & Rakerd, B. (1981). Duplex perception of cues for stop consonants: evidence for a phonetic mode. *Perception and Psychophysics, 30*(2), 133-143.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*(1), 1-36.

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

Lieberman, P. H. (1990). *Uniquely human: The evolution of speech, thought and selfless behavior*. Cambridge, MA: Harvard University Press.

Lieberman, P. H. (1994). Human language and human uniqueness. *Language and Communication, 14*(1), 87-95.

Long, K. D., Kennedy, G., & Balaban, E. (2001). Transferring an inborn auditory perceptual predisposition with interspecies brain transplants. *Proceedings of the National Academy of Sciences (USA), 98*(10), 5862-5867.

Lorenz, K. (1952). *King Solomon's ring*. New York, NY: Thomas Y Crowell Company.

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science, 283*(5398), 77-80.

Marler, P. (1990). Innate learning preferences: Signals for communication. *Developmental Psychobiology, 23*(7), 557-568.

Marler, P. (1997). Three models of song learning: evidence from behavior. *Journal of Neurobiology, 33*(5), 501-516.

Marler, P., & Peters, S. (1977). Selective vocal learning in a sparrow. *Science, 198*(4316), 519-521.

Marler, P., & Peters, S. (1989). Species differences in auditory responsiveness in early vocal learning. In R. J. Dooling & S. H. Hulse (Eds.), *The comparative psychology of audition: Perceiving complex sounds* (pp. 243-273). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Massaro, D. W. (1994). Psychological aspects of speech perception: Implications for research and theory. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics.* (pp. 219-263). San Diego, CA, US: Academic Press, Inc.

Maurer, D., & Young, R. (1983). Newborns' following of natural and distorted arrangements of facial features. *Infant Behavior and Development, 6*, 127-131.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*(3), B101-111.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746-748.

Mehler, J., Bertoncini, J., & Barriere, M. (1978). Infant recognition of mother's voice. *Perception, 7*(5), 491-497.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertoncini, J., & Amiel-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29*(2), 143-178.

Mesulam, M. M. (2000). Behavioral neuroanatomy: Large-scale networks, association cortex, frontal syndromes, the limbic system, and hemispheric specializations. In M. M. Mesulam (Ed.), *Principles of behavioral and cognitive neurology* (2nd ed., pp. 540). Oxford; New York: Oxford University Press.

Miller, J. L., & Eimas, P. D. (1983). Studies on the categorization of speech by infants. *Cognition, 13*(2), 135-165.

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception and Psychophysics, 25*(6), 457-465.

Molfese, D. L. (1980). Hemispheric specialization for temporal information: implications for the perception of voicing cues during speech perception. *Brain and Language, 11*(2), 285-299.

Molfese, D. L., Buhrke, R. A., & Wang, S. L. (1985). The right hemisphere and temporal processing of consonant transition durations: electrophysiological correlates. *Brain and Language, 26*(1), 49-62.

Molfese, D. L., & Molfese, V. J. (1979a). Hemisphere and stimulus differences as reflected in the cortical responses of newborn infants to speech stimuli. *Developmental Psychology, 15*(5), 505-511.

Molfese, D. L., & Molfese, V. J. (1979b). Infant speech perception: Learned or innate? In H. A. Whitaker & H. Whitaker (Eds.), *Advances in neurolinguistics* (Vol. 4). New York: Academic Press.

Molfese, D. L., & Molfese, V. J. (1980). Cortical response of preterm infants to phonetic and nonphonetic speech stimuli. *Developmental Psychology, 16*(6), 574-581.

Moon, C., Cooper, R. P., & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant Behavior and Development, 16*(4), 495-500.

Morgan, J. L., & Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development, 66*(4), 911-936.

Morton, J., & Johnson, M. H. (1991). CONSPEC and CONLERN: A two-process theory of infant face recognition. *Psychological Review, 98*(2), 164-181.

Muir, D., & Field, J. (1979). Newborn infants orient to sounds. *Child Development, 50*(2), 431-436.

Mummery, C. J., Ashburner, J., Scott, S. K., & Wise, R. J. (1999). Functional neuroimaging of speech perception in six normal and two aphasic subjects. *Journal of the Acoustical Society of America, 106*(1), 449-457.

Myers, J., Jusczyk, P. W., Kemler Nelson, D. G., Charles Luce, J., Woodward, A. L., & Hirsh Pasek, K. (1996). Infants' sensitivity to word boundaries in fluent speech. *Journal of Child Language, 23*(1), 1-30.

Najm-Briscoe, R. G., Thomas, D. G., & Overton, S. (2000). The impact of stimulus "value" in infant novelty preference. *Developmental Psychobiology, 37*(3), 176-185.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance, 24*(3), 756-766.

Nelson, D. A. (2000). A preference for own-subspecies' song guides vocal learning in a song bird. *Proceedings of the National Academy of Sciences (USA), 97*(24), 13348-13353.

Novak, G. P., Kurtzberg, D., Kreuzer, J. A., & Vaughan Jr., H. G. (1989). Cortical responses to speech sounds and their formants in normal infants: Maturational sequence and spatiotemporal analysis. *Electroencephalography and Clinical Neurophysiology, 73*, 295-305.

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences (USA), 87*(24), 9868-9872.

Ojemann, G., Ojemann, J., Lettich, E., & Berger, M. (1989). Cortical language localization in left, dominant hemisphere. An electrical stimulation mapping investigation in 117 patients. *Journal of Neurosurgery, 71*(3), 316-326.

Park, T., & Balaban, E. (1991). Relative salience of species maternal calls in neonatal gallinaceous birds: a direct comparison of Japanese quail (Coturnix coturnix japonica) and domestic chickens (Gallus gallus domesticus). *Journal of Comparative Psychology, 105*(1), 45-54.

Parker, E. M., Diehl, R. L., & Kluender, K. R. (1986). Trading relations in speech and nonspeech. *Perception and Psychophysics, 39*(2), 129-142.

Pascalis, O., & de Haan, M. (2003). Recognition memory and novelty preference: What model? In H. Hayne & J. W. Fagen (Eds.), *Progress in infancy research* (Vol. 3, pp. 95-119). Mahwah, NJ, US: Lawrence Erlbaum Associates.

Pastore, R. E., Schmuckler, M. A., Rosenblum, L., & Szczesiul, R. (1983). Duplex perception with musical stimuli. *Perception and Psychophysics, 33*(5), 469-474.

Pegg, J. E., Werker, J. F., & McLeod, P. J. (1992). Preference for infant-directed over adult-directed speech: Evidence from 7-week-old infants. *Infant Behavior and Development, 15*(3), 325-345.

Peña, M., Maki, A., Kovacic, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., & Mehler, J. (2003). Sounds and silence: An optical topography study of language recognition at birth. *Proceedings of the National Academy of Sciences (USA), 100*(20), 11702-11705.

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences, 13*(4), 707-784.

Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: implications for voicing perception in stops. *Journal of the Acoustical Society of America, 61*(5), 1352-1361.

Pisoni, D. B., Carrell, T. D., & Gans, S. J. (1983). Perception of the duration of rapid spectrum changes in speech and nonspeech signals. *Perception and Psychophysics, 34*(4), 314-322.

Poeppel, D. (1996). A critical review of PET studies of phonological processing. *Brain and Language, 55*(3), 317-351; discussion 352-385.

Pons, F., Trobalon, J. B., Bosch, L., & Sebastián-Gallés, N. (under review). The role of experience on phonetic categorization in rats. *Animal Cognition.*

Porter, R. H. (1998). Olfaction and human kin recognition. *Genetica, 104*(3), 259-263.

Price, C. J., Wise, R. J., Warburton, E. A., Moore, C. J., Howard, D., Patterson, K., Frackowiak, R. S., & Friston, K. J. (1996). Hearing and saying. The functional neuro-anatomy of auditory word processing. *Brain, 119*(Pt 3), 919-931.

Pugh, K. R., Shaywitz, B. A., Shaywitz, S. E., Fulbright, R. K., Byrd, D., Skudlarski, P., Shankweiler, D. P., Katz, L., Constable, R. T., Fletcher, J., Lacadie, C., Marchione, K., & Gore, J. C. (1996). Auditory selective attention: an fMRI investigation. *Neuroimage, 4*(3 Pt 1), 159-173.

Querleu, D., Lefebvre, C., Titran, M., Renard, X., Morillion, M., & Crepin, G. (1984). [Reaction of the newborn infant less than 2 hours after birth to the maternal voice]. *Journal de Gynécologie Obstétrique et Biologie de la Reproduction (Paris), 13*(2), 125-134.

Ramus, F., Hauser, M. D., Miller, C., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science, 288*(5464), 349-351.

Rauschecker, J. P. (1998). Cortical processing of complex sounds. *Current Opinion in Neurobiology, 8*(4), 516-521.

Remez, R. E. (1994). A guide to research on the perception of speech. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics.* (pp. 145-172). San Diego, CA, US: Academic Press, Inc.

Remez, R. E., Pardo, J. S., Piorkowski, R. L., & Rubin, P. E. (2001). On the bistability of sine wave analogues of speech. *Psychological Science, 12*(1), 24-29.

Remez, R. E., Rubin, P. E., & Pisoni, D. B. (1983). Coding of the speech spectrum in three time-varying sinusoids. *Annals of the New York Academy of Sciences, 405*, 485-489.

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 212*(4497), 947-949.

Repp, B. H. (1982). Phonetic trading relations and context effects: new experimental evidence for a speech mode of perception. *Psychological Bulletin, 92*(1), 81-110.

Roder, B. J., Bushnell, E. W., & Sasseville, A. M. (2000). Infants' preferences for familiarity and novelty during the course of visual processing. *Infancy, 1*(4), 491-507.

Rose, S. A., Gottfried, A. W., Melloy-Carminar, P., & Bridger, W. H. (1982). Familiarity and novelty preferences in infant recognition memory: Implications for information processing. *Developmental Psychology, 18*(5), 704-713.

Rosen, B. R., Buckner, R. L., & Dale, A. M. (1998). Event-related functional MRI: past, present, and future. *Proceedings of the National Academy of Sciences (USA), 95*(3), 773-780.

Russell, M. J. (1976). Human olfactory communication. *Nature, 260*(5551), 520-522.

Ryan, M. J., Phelps, S. M., & Rand, A. S. (2001). How evolutionary history shapes recognition mechanisms. *Trends in Cognitive Sciences, 5*(4), 143-148.

Ryan, M. J., & Rand, A. S. (1995). Female responses to ancestral advertisement calls in Túngara frogs. *Science, 269*(5222), 390-392.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926-1928.

Samuel, A. G., & Tartter, V. C. (1986). Acoustic-phonetic issues in speech perception. *Annual Review of Anthropology, 15*, 247-273.

Sansavini, A., Bertoncini, J., & Giovanelli, G. (1997). Newborns discriminate the rhythm of multisyllabic stressed words. *Developmental Psychology, 33*(1), 3-11.

Schwartz, J., & Tallal, P. (1980). Rate of acoustic change may underlie hemispheric specialization for speech perception. *Science, 207*(4437), 1380-1381.

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain, 123*(12), 2400-2406.

Shi, R., & Werker, J. F. (2001). Six-month-old infants' preference for lexical words. *Psychological Science, 12*(1), 70-75.

Shi, R., Werker, J. F., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition, 72*(2), B11-21.

Simion, F., Cassia, V. M., Turati, C., & Valenza, E. (2001). The origins of face perception: Specific versus non-specific mechanisms. *Infant and Child Development, 10*(1-2), 59-65.

Simos, P. G., & Molfese, D. L. (1997). Electrophysiological responses from a temporal order continuum in the newborn infant. *Neuropsychologia, 35*(1), 89-98.

Stevens, A. A., & Schwartzreich, E. (2000). *Dissociating message and messenger: fMRI of word and voice recognition.* Paper presented at the Cognitive Neuroscience Meeting, San Francisco, CA, USA.

Streeter, L. A. (1976). Language perception of 2-mo-old infants shows effects of both innate mechanisms and experience. *Nature, 259*(5538), 39-41.

Tallal, P., Miller, S., & Fitch, R. H. (1993). Neurobiological basis of speech: a case for the preeminence of temporal processing. *Annals of the New York Academy of Sciences, 682*, 27-47.

Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child Development, 47*(2), 466-472.

Trout, J. D. (2001). The biological basis of speech: what to infer from talking to the animals. *Psychological Review, 108*(3), 523-549.

Turkewitz, G., Birch, H. G., & Cooper, K. K. (1972). Responsiveness to simple and complex auditory stimuli in the human newborn. *Developmental Psychobiology, 5*(1), 7-19.

Tzourio, N., Massioui, F. E., Crivello, F., Joliot, M., Renault, B., & Mazoyer, B. (1997). Functional anatomy of human auditory attention studied with PET. *Neuroimage, 5*(1), 63-77.

Underbakke, M., Polka, L., Gottfried, T. L., & Strange, W. (1988). Trading relations in the perception of /r/-/l/ by Japanese learners of English. *Journal of the Acoustical Society of America, 84*(1), 90-100.

Valenza, E., Simion, F., Cassia, V. M., & Umilta, C. (1996). Face preference at birth. *Journal of Experimental Psychology: Human Perception and Performance, 22*(4), 892-903.

Vouloumanos, A., Kiehl, K. A., Werker, J. F., & Liddle, P. F. (2001). Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. *Journal of Cognitive Neuroscience, 13*(7), 994-1005.

Vouloumanos, A., & Werker, J. F. (in press). Tuned to the signal: the privileged status of speech for young infants. *Developmental Science.*

Vouloumanos, A., & Werker, J. F. (under review). A bias for speech in human newborns.

Walton, G. E., Bower, N. J., & Bower, T. G. (1992). Recognition of familiar faces by newborns. *Infant Behavior and Development, 15*(2), 265-269.

Wang, X., & Kadia, S. C. (2001). Differential representation of species-specific primate vocalizations in the auditory cortices of marmoset and cat. *Journal of Neurophysiology, 86*(5), 2616-2620.

Wang, X., Merzenich, M. M., Beitel, R., & Schreiner, C. E. (1995). Representation of a species-specific vocalization in the primary auditory cortex of the common marmoset: temporal and spectral characteristics. *Journal of Neurophysiology, 74*(6), 2685-2706.

Ward, C. D., & Cooper, R. P. (1999). A lack of evidence in 4-month-old human infants for paternal voice preference. *Developmental Psychobiology, 35*(1), 49-59.

Werker, J. F., & McLeod, P. J. (1989). Infant preference for both male and female infant-directed talk: A developmental study of attentional and affective responsiveness. *Canadian Journal of Psychology, 43*(2), 230-246.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*(1), 49-63.

Werker, J. F., & Tees, R. C. (1992). The organization and reorganization of human speech perception. *Annual Review of Neuroscience, 15*, 377-402.

Werker, J. F., & Vouloumanos, A. (2001). Speech and language processing in infancy: a neurocognitive approach. In C. A. Nelson & M. Luciana (Eds.), *Handbook of Developmental Cognitive Neuroscience* (pp. 269-280). Cambridge, MA: MIT Press.

Wetherford, M. J., & Cohen, L. B. (1973). Developmental changes in infant visual preferences for novelty and familiarity. *Child Development, 44*(3), 416-424.

Whalen, D. H., & Liberman, A. M. (1987). Speech perception takes precedence over nonspeech perception. *Science, 237*(4811), 169-171.

Whaling, C. S., Solis, M. M., Doupe, A. J., Soha, J. A., & Marler, P. (1997). Acoustic and neural bases for innate recognition of song. *Proceedings of the National Academy of Sciences (USA), 94*(23), 12694-12698.

Wong, D., Miyamoto, R. T., Pisoni, D. B., Sehgal, M., & Hutchins, G. D. (1999). PET imaging of cochlear-implant and normal-hearing subjects listening to speech and nonspeech. *Hearing Research, 132*(1-2), 34-42.

Zatorre, R. J., Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *Journal of Neuroscience, 14*(4), 1908-1919.

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science, 256*(5058), 846-849.

Endnotes

[1] Note, however, that in conditions of complete sensory deprivation, chicks show no visual preferences. This experience is so removed from the chick's normal experience that many abnormalities are likely to arise and conclusions about normal development are difficult to draw (Bolhuis, Johnson, & Horn, 1985).

[2] Perceptual tests in our laboratory with 8 monolingual English speakers confirmed that all 8 listeners identified the nonsense speech sounds as human vocalisations, while none identified the sine-wave analogues as such (unpublished data from our laboratory).

[3] For ease of readability, the speech stimuli are described using "gloss". The International Phonetic Alphabet (IPA) equivalent symbol is /lIf/.

[4] The methodological studies of Butterfield and Siperstein (1970) are sometimes cited as providing relevant evidence. However, the "speech" condition in their experiments consisted of folk music, and as such doesn't directly bear on infants' preference for speech.

[5] Twelve 6.5-month-olds were excluded because of excessive fussiness (2), technical errors (3), experimenter error (5), failure to learn the contingency (1) or parental interference in the experiment (1). Thirteen 4.5-month-olds were excluded because of excessive fussiness (3), technical errors (2), parental interference during the experiment (1), failure to learn the contingency (3), experimenter error (3) or hiccups (1). Thirteen 2.5-month-olds were excluded because of fussiness or sleepiness (9), technical errors (1), failure to learn the contingency (1), and experimenter error (2).

[6] For ease of readability, we describe the speech stimuli using "gloss". The equivalent in IPA symbols is /lIf/ and /nim/.

[7] In testing 2.5-month-olds, the red flashing light was also used between experimental trials to elicit infant attention.

[8] Pilot testing revealed that 20 s trials were too short for 2.5 month-olds to reliably learn the contingency therefore 2.5-month-olds were tested using a maximum trial length of

40 s. Forty-second trials were created by simply repeating the 14 tokens presented during the first 20 s of that trial.

[9] One study with newborns used folk music as the "speech" condition and compared it to white noise (Butterfield & Siperstein, 1970). Folk music has both vocal and musical components, which were not dissociated in this study, and thus the basis of infants' preference for folk music over white noise is unknown.

[10] Adults who were tested on their ability to discriminate LPF speech from LPF complex non-speech performed at chance, while showing no difficulty when discriminating unfiltered speech from unfiltered complex non-speech, echoing the results of newborns (unpublished data from our laboratory).