Language learning in infancy

Tuomas Teinonen

Cognitive Brain Research Unit
Department of Psychology
Faculty of Behavioural Sciences
University of Helsinki
Finland

Department of Pediatrics
Institute of Clinical Medicine
Faculty of Medicine
University of Helsinki
Finland

Language learning in infancy

Tuomas Teinonen

Abstract

Although immensely complex, speech is also a very efficient means of communication between humans. Understanding how we acquire the skills necessary for perceiving and producing speech remains an intriguing goal for research. However, while learning is likely to begin as soon as we start hearing speech, the tools for studying the language acquisition strategies in the earliest stages of development remain scarce. One prospective strategy is statistical learning. In order to investigate its role in language development, we designed a new research method. The method was tested in adults using magnetoencephalography (MEG) as a measure of cortical activity. Neonatal brain activity was measured with electroencephalography (EEG). Additionally, we developed a method for assessing the integration of seen and heard syllables in the developing brain as well as a method for assessing the role of visual speech when learning phoneme categories.

The MEG study showed that adults learn statistical properties of speech during passive listening of syllables. The amplitude of the N400m component of the event-related magnetic fields (ERFs) reflected the location of syllables within pseudowords. The amplitude was also enhanced for syllables in a statistically unexpected position. The results suggest a role for the N400m component in statistical learning studies in adults.

Using the same research design with sleeping newborn infants, the auditory event-related potentials (ERPs) measured with EEG reflected the location of syllables within pseudowords. The results were successfully replicated in another group of infants. The results show that even newborn infants have a powerful mechanism for automatic extraction of statistical characteristics from speech.

We also found that 5-month-old infants integrate some auditory and visual syllables into a fused percept, whereas other syllable combinations are not fully integrated. Auditory syllables were paired with visual syllables possessing a different phonetic identity, and the ERPs for these artificial syllable combinations were compared with the ERPs for normal syllables. For congruent auditory-visual syllable combinations, the ERPs did not differ from those for normal syllables. However, for incongruent auditory-visual syllable combinations, we observed a mismatch response in the ERPs. The results show an early ability to perceive speech cross-modally.

Finally, we exposed two groups of 6-month-old infants to artificially created auditory syllables located between two stereotypical English syllables in the formant space. The auditory syllables followed, equally for both groups, a unimodal statistical distribution, suggestive of a single phoneme category. The visual syllables combined with the auditory syllables, however, were different for the two groups, one group receiving visual stimuli suggestive of two separate phoneme categories, the other receiving visual stimuli suggestive of only one phoneme category. After a short exposure, we observed different learning outcomes for the two groups of infants. The results thus show that visual speech can influence learning of phoneme categories.

Altogether, the results demonstrate that complex language learning skills exist from birth. They also suggest a role for the visual component of speech in the learning of phoneme categories.

Tiivistelmä

Puhe on monimuotoinen signaali, joka välittää ihmistenvälistä kommunikaatiota erityisen tehokkaasti. On osin vielä hämärän peitossa, miten opimme puhumaan ja havaitsemaan puhetta syntymän jälkeen. Oppiminen alkanee heti, kun alamme kuulla puhetta. Puheen sisältämät lukuisat tilastolliset säännönmukaisuudet saattavat auttaa oppimista. Niiden hyödyntämistä kutsutaan tilastolliseksi oppimiseksi. Tilastollista oppimista ajatellaan kielenoppimisessa voitavan käyttää jatkuvan puheen jakamiseksi erillisiksi sanoiksi.

Tässä tutkimuksessa kehitettiin uusi aivomittauksiin perustuva tutkimusmenetelmä tilastollisen kielenoppimisen mittaamiseksi. Menetelmää testattiin aikuisilla mittaamalla aivokuoriaktivaatiota magnetoenkefalografiaa käyttäen. Tulokset osoittivat, että aikuiset oppivat tilastollisia ominaisuuksia tavuvirrasta silloinkin, kun he eivät kiinnitä siihen tietoisesti huomiota. Tilastollista oppimista pystyttiin mittaamaan aivojen magneettisissa herätevasteissa näkyvän N400m-vasteen yhteydessä.

Samaa koeasetelmaa käytettiin mittamaan tilastollisen kielenoppimisen kykyjä vastasyntyneiltä käyttäen elektroenkefalografiaa aivotoiminnan mittarina. Tulokset osoittivat, että vastasyntyneet oppivat tilastollisia riippuvuuksia tavujen välillä unen aikana kuulemastaan tavuvirrasta. Tutkimustulos toistui myös toisella ryhmällä vastasyntyneitä. Jo vastasyntyneillä on siis hyvä puheen tilastollisten ominaisuuksien oppimiskyky.

Tehokkaassa puheenhavaitsemisessa olennaisena pidetään kykyä yhdistää eli integroida kuultu puhe ja nähty artikulaatio yhdeksi havainnoksi. Ajatellaan, että aikuisilla aivot käsittelevät integroidun puheen tehokkaammin kuin erilliset kuulo- ja näköhavainnot. Kehitimme koeasetelman kuullun ja nähdyn puheen integroinnin mittaamiseksi viiden kuukauden ikäisillä vauvoilla. Tulokset osoittivat, että tässä kehitysvaiheessa vauvat muodostavat yhdistyneen havainnon tietyistä kuulluista ja nähdyistä tavuista, kun taas toisia tavuyhdistelmiä ei integroida. Koska integrointi ei onnistu tilanteessa, jossa muodostuva havainto ei ole äidinkielen sääntöjen mukainen, tulokset myös viittaavat siihen, että viiden kuukauden iässä vauvat ovat jo omaksuneet tietoa äidinkielen tyypillisistä rakenteista.

Lopuksi arvioitiin myös sitä, millainen rooli nähdyllä puheella saattaisi olla kielenoppimisessa. Kaksi ryhmää kuuden kuukauden ikäisiä vauvoja osallistui kokeeseen. Vauvoille näytettiin ruudulta puhetta, jossa videoon tavallisesta artikulaatiosta oli liitetty keinotekoisesti muokattuja tavuääniä. Osalle vauvoista nähty puhe sisälsi vihjeen kahdesta erilaisesta tavuryhmästä, kun taas toiselle ryhmälle nähty puhe pysyi koko ajan samanlaisena. Kuultu puhe oli kaikille vauvoille täysin samanlaista. Lyhyen katselu- ja kuunteluajan jälkeen vauvojen havaintoa kyseisistä tavuäänistä testattiin. Eri ryhmillä havaittiin erilaiset oppimistulokset, joka viittaa siihen, että nähdyillä artikulaatioilla oli vaikutusta oppimiseen. Nähty puhe siis voi vaikuttaa puheäänten oppimiseen tässä kehitysvaiheessa.

Kokonaisuudessaan tässä väitöskirjassa esitetyt tulokset korostavat varhaisimpien vaiheiden merkitystä kielenoppimisessa. Pystymme oppimaan puheen ominaisuuksia monipuolisesti heti syntymästä lähtien. Tulokset myös viittaavat siihen, että nähdyllä puheella saattaa olla tärkeä rooli puheäänien oppimisessa ensimmäisen vuoden aikana.

Contents

List of original publications

This thesis is based on the following publications:

I. **Teinonen, T**. & Huotilainen, M. Implicit segmentation of continuous speech based on transitional probabilities: an MEG study. Under review.

II. **Teinonen, T**., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical language learning in neonates revealed by event-related brain potentials. *BMC Neuroscience*, 10:21.

III. Kushnerenko, E., **Teinonen, T**., Volein, A., & Csibra, G. (2008). Electrophysiological evidence of illusory audiovisual speech percept in human infants. *Proceedings of the National Academy of Sciences of the USA*, 105:11442-11445.

IV. **Teinonen, T**., Aslin, R.N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*, 108:850-855.

The publications are referred to in the text by their roman numerals.

Abbreviations

ANOVA   Analysis of variance
ECD   Equivalent current dipole
EEG   Electroencephalography
EOG   Electro-oculogram
ERF   Event-related magnetic field
ERP   Event-related potential
GA   Gestational age
HAS   High-amplitude sucking
MEG   Magnetoencephalography
MMN   Mismatch negativity
NBAS   Neonatal behavioural assessment scale
SAPP   Stimulus-alternation preference procedure
SQUID   Superconducting quantum interference device

1 Introduction

As any carer can attest, the language acquisition process is easy to identify as an infant's vocal production turns slowly from babbling into intelligible words; even more so later as the child's active vocabulary increases. What happens before these stages, however, remains an intriguing question. How could we, as researchers, best examine the language learning processes in early infancy?

Due to recent advancements in brain imaging methods, it is now possible to gain a greater view into the infant mind than ever before. While new methods based on brain recordings must be viewed through a critical lens, targeted usage has proved particularly useful in studying infant auditory perception, reflected in cortical activation. However, the best approach for studying various aspects of language learning will be comprehensive, taking into account not only these new methods but also those drawn from more classic behavioural research, which can access different kinds of information and help build a more complete picture.

The new view on language learning includes an emphasis on the statistical properties of language – an immensely rich source of information that can provide cues to disentangling many of the variabilities between human languages. Additionally, the fact that we not only hear but also see speech is starting to gain credibility. The multimodal perception of speech may not only be important to individuals engaged in a discussion, but it may also provide complementary information to the learning infant.

This study addresses language learning processes taking place during the first months after birth. While the emphasis of the work is on statistical language learning, i.e., extracting and utilising statistical regularities of speech, as well as on the effects of visual speech on the learning process, these results may also be applicable to other, non-linguistic domains. Additionally, the developed research methods may upon further refinement provide an early measure of cognitive development with clear clinical importance.

Language development has been extensively studied for decades resulting in hundreds of important publications. In the literature review presented in the next chapter a focus is given exclusively on the studies of language development that have served as a basis for my own research. Consequently, an emphasis is put on learning of phoneme categories, speech segmentation, statistical learning, and auditory-visual speech perception. In the subsequent chapter, some of the current infant research methods are presented, again emphasising those used in my own work. The reviews are followed by the particular aims of this study, description of the research methods used, the acquired results, and finally discussion of the results and their implications.

2 Review of the literature

2.1 Language development in early infancy

The discussion of language development has long focused on two competing theories. The nativist theory assumes the existence of innate and highly specialised mental structures enabling the infant brain to determine which of all possible linguistic rules the infant's native language employs (Pinker, 1994). However, more recently an empiricist view questioning the necessity of the language-specific structures and emphasising empirical evidence inconsistent with the assumptions of the nativist view has gained momentum (Tomasello, 2003). Despite many unresolved issues between the two theories and regardless of the biological bases of language, there is now a general agreement that powerful learning mechanisms are needed to learn a language (Hauser, Chomsky, & Fitch, 2002).

Babbling is a clear reflection of ongoing language development. Moreover, its characteristics evolve according to the native language environment. Infants produce vowel-like sounds from approximately 3 months, and start to babble around the age of 6 months (Kuhl, 2004). The characteristics of the vocal production develop over time, growing gradually to resemble those of the native language (Boysson-Bardies & Vihman, 1991; Petitto & Marentette, 1991).

Among all the other sounds within their environment, speech seems to be very special to infants. From very early on, infants have a bias to attend to speech over similar nonspeech sounds (Vouloumanos & Werker, 2004). Also, carers automatically speak to infants in a specific fashion drawing infants' attention (Cooper & Aslin, 1990). The repetitive character and exaggerated pitch contours of this *infant-directed speech* also aid vowel category acquisition (Trainor & Desjardins, 2002). From birth, the infants can differentiate languages from different rhythmic classes (Nazzi, Bertoncini, & Mehler, 1998), and at the age of 4–5 months, they can differentiate their native language from an unfamiliar one (Bosch & Sebastián Gallés, 1997; Nazzi, Jusczyk, & Johnson, 2000).

2.1.1 Learning of phoneme categories

Eimas and others (1971) conducted a pioneering study in which they showed that the 1- to 4-month-old infants were able to discriminate a voiced consonant /b/ from a voiceless consonant /p/, even though the two consonants are acoustically quite similar. Moreover, the infants failed to discriminate different tokens of /b/ that were acoustically distinct, providing evidence of speech sound categorisation even at this early age. In another study, it was found that the five-month-old infants could also group tokens of the same phonetic category together despite notable acoustic variation due to, e.g., variation of speaker gender (Kuhl, 1979).

Later, it was noticed that infants discriminate not only the phonetic contrasts present in their native language, but also all possible phonetic contrasts from unfamiliar languages.

Werker and Tees (1984) demonstrated that the 6-month-old infants from English-speaking families and environments were able to discriminate a consonant contrast present in Hindi, whereas English-speaking adults showed no such discrimination. However, infants gradually lose this universal perception of phonemes by the age of 12 months, while the ability to discriminate native-language phonemes improves (Cheour et al., 1998; Rivera-Gaxiola, Silva-Pereyra, & Kuhl, 2005).

Kuhl and others (1992) tested 6-month-old infants from English- and Swedish-speaking families. Playing vowel variants from partially overlapping Swedish /y/ and English /i/ categories, they measured how the infants would determine the correct category for these speech sounds. The infants perceived more variants as identical to the prototype vowel of their native language than that of the unfamiliar language. This effect, named the *perceptual magnet effect*, suggested that the 6-month-old infants had already learned the location of the native-language prototype vowels in the formant space, and to categorise the surrounding variants to this category. Later results confirmed that foreign vowel contrasts are discriminated by 4-month-old infants, but some tuning for native-language vowels occurs by 6 months of age (Polka & Werker, 1994). Thus, specificity for native-language phonemes can be observed somewhat earlier for vowels than for consonants.

The perceptual magnet effect is an example of a computational strategy available for infants: by learning the distribution of phonetic variants, they learn to categorise new speech sounds to the most likely category. For other examples of statistical learning related to language acquisition, see Chapter 2.2.

2.1.2 Extracting words from fluent speech

Hearing foreign speech, we often have difficulties determining where one word ends and the next one begins. Similarly, infants are born without a lexicon of their mother tongue. To make it even more demanding, infants do not typically hear most words in isolation. Consequently, infants are faced with a challenge of extracting single words out of fluent, continuous speech. The methods used to segment speech may vary depending on speech characteristics. Additionally, the use of different methods at different stages of development may be due to different types of cues embedded in the infant-directed speech as well as the information available in an individual infant's memory.

Jusczyk and Aslin (1995) familiarised infants with words both in isolation and in a sentence context. They found that the 7.5-month-old infants, but not the 6-month-old infants, recognised these target words from fluent speech in a subsequent test. Furthermore, the 7.5-month-old infants did not recognise phonetically similar words that only differed by their initial phonetic segment, e.g., when the first phoneme was similar but different, suggesting that at this age a detailed representation of the target words was memorised, rather than some salient acoustic feature, such as a vowel identity. Thus, the ability to recognise previously learned words from fluent speech seems to be dramatically enhanced between 6 and 7.5 months of age.

One stress pattern typically dominates in a language. In English and Finnish, the stress for the onset of most words is strong while the following stress is weak, as in the word

kingdom. In many other languages, such as Spanish, a different pattern is dominant. Jusczyk and others (1999) found that 7.5-month-old infants from English-speaking environments correctly segmented bisyllabic words with the dominant (strong/weak) pattern, such as candle or hamlet, but failed to segment words with a less frequent stress pattern (weak/strong), such as guitar or surprise. However, 10.5-month-old infants from a similar background were able to detect both of the patterns. Thus, the importance of the stress as a cue for word segmentation seems to diminish between these two developmental stages.

Phonotactics are also a possible cue to word boundaries: for instance in Finnish, gemination (CC) is a relatively common phonotactic sequence, and it can never occur in the beginning or at the end of a word. At 9 months of age, infants can distinguish between a phonotactic sequence that occurs frequently across word boundaries and a sequence that rarely occurs at a word boundary (Mattys, Jusczyk, Luce, & Morgan, 1999). Similarly, infants are able to distinguish phonotactic sequences occurring frequently and rarely within words. Thus, at least after enough exposure to the native language, the phonotactic patterns serve as prospective cues to word boundaries.

Statistical cues are also likely to bear importance in segmenting fluent speech. The phonotactic patterns mentioned above provide one example: the statistical distributions of phonotactic patterns within words, close to word boundaries, or across word boundaries provide probabilistic information of word boundary locations. Additional examples are presented in the next section.

All cues are not available at all times. For instance, many morphological cues may be available only when an infant has already learned the use of certain morphological patterns in the native language (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993). Also, a noisy environment or an emotional utterance may emphasise some cues over others (Newman, 2003). When one cue provides many alternative estimates, another cue may help to form an accurate estimate of the location of the word boundary. Similarly if two cues would provide a conflicting estimate, a third cue might help to resolve the conflict. Indeed, it is likely that at least more experienced listeners combine multiple cues when segmenting words from fluent speech (Jusczyk, 1999; Morgan & Saffran, 1995; Myers et al., 1996).

2.2 Statistical learning

For computer programmers performing data mining, i.e., extracting hidden patterns from large sets of data, the search for statistical regularities is a given. It is only reasonable that the "computer in our heads", the brain, is also performing data mining, as the sensory input keeps constantly outrunning the processing capacity available and, perhaps more importantly, only a tiny part of the input is relevant. The search of the brain for useful statistical regularities is called statistical learning. Kuhl (2004) defines statistical learning in the following way:

Statistical learning is acquisition of knowledge through the computation of information about the distributional frequency with which certain items occur in relation to others, or probabilistic information in sequences of stimuli, such as the odds (transitional probabilities) that one unit will follow another in a given language.

The major interest in statistical learning in relation to language development started with a pioneering study by Saffran and others (1996). They presented 8-month-old infants with a continuous speech stream consisting of four trisyllabic pseudowords repeated in a pseudorandom order. The stream contained no morphological cues to word boundaries, i.e., the only way that the infants could detect the word boundaries and memorise the words was to learn either the transitional probabilities or frequencies of co-occurrence between successive syllables. The infants showed recognition of the pseudowords in a subsequent test. It was later confirmed that the infants were, indeed, able to learn transitional probabilities, rather than frequencies of co-occurrence, between subsequent syllables (Aslin, Saffran, & Newport, 1998).

In this thesis, the focus is on speech. However, statistical regularities are present in various domains. Often the statistics of everyday life are all too evident to be noticed: the members of a family are more often perceived together than with outsiders, and the opening of a door is frequently followed by a person entering the room. On the other hand, some statistics of everyday life are too complex to determine without the help of mathematics: for instance, statistical characteristics of illumination patterns are immensely complex, but have an important role in how we perceive objects and recognise materials (Dror, Willsky, & Adelson, 2004). The research to date has shown that infants are also able to segment sequences of tones (Saffran, Johnson, Aslin, & Newport, 1999) and organise visual objects (Fiser & Aslin, 2002; Kirkham, Slemmer, & Johnson, 2002) based on transitional probabilities or frequencies of co-occurrence.

At the age of 12 months, infants are able to extract multilevel information from speech input using statistical learning. Saffran and Wilson (2003) exposed infants to speech that contained two overlying structures: words were embedded into continuous speech, and the word order followed a simple grammar. Thus, the infants could discover the grammatical patterns only after segmenting the words. In a post-exposure test the infants distinguished grammatical from ungrammatical sentences, demonstrating that they had first segmented the words and then acquired the grammar.

In addition to syllables, phoneme variants heard by an infant also contain distributional information useful for language learners. The perceptual magnet effect discussed earlier (Kuhl et al., 1992) demonstrated how the distribution of allophones in the formant space is typically centred around a prototype phoneme and also how learning of the distributions drives the perception of rare phonetic variants. Maye and others (2002) tested the effect of phonetic distributions on phonetic category learning in 6-month-old infants. They exposed two groups of infants to different distributions of speech sounds from a continuum between /da/ and /ta/. One group was exposed to a unimodal distribution, the peak of which lay in the middle of the continuum (i.e., the middle syllables were the most common). The other group was exposed to a bimodal distribution, the peaks of which lay close to the two endpoints of the continuum (i.e., the middle syllables were the rarest). The

learning outcomes of the two groups differed notably: the infants exposed to the unimodal distribution did not distinguish the two test items /da/ and /ta/, whereas the infants in the bimodal group did. Thus, the bimodal distribution of speech sounds, typically occurring between neighbouring phonemes in formant space, can facilitate discrimination of difficult phoneme contrasts (Maye, Weiss, & Aslin, 2008).

The statistical learning studies are often confined to using overly simplified artificial languages. Therefore, it is important to validate the applicability of the results for natural language. The corpus analyses of Swingley (2005) suggest that transitional probabilities between syllables are a potentially useful cue to word boundaries also in natural speech. Saffran (2001) showed that 8-month-old infants adopt segmented units as possible words in their native language. Furthermore, Pelucchi and others (2009) showed that 8-month-old infants are sensitive to transitional probabilities in an unfamiliar natural language. Thus, there is evidence suggesting that statistical learning indeed is useful for unscrambling the code of natural language along language development.

Although many of the studies presented above used morphologically poor stimuli and demonstrated learning based on statistics only, it does not render prosody redundant or unimportant. In fact, studies combining prosodic and statistical cues have acquired more efficient learning results when both cues were available than in the presence of only statistical cues (Cunillera, Toro, Sebastián Gallés, & Rodriguez-Fornells, 2006; Pena, Bonatti, Nespor, & Mehler, 2002). Some studies also highlight the existence of rules that cannot be learned based on their statistical properties only (Marcus, Vijayan, Bandi Rao, & Vishton, 1999; Toro, Nespor, Mehler, & Bonatti, 2008). Indeed, statistical properties of linguistic items can only be learned after the items have been detected as such. Perhaps the most important role of prosody in the learning context lies in this very task: intonation, stress, intensity, and other morphological cues may highlight essential structures of speech to enable statistical learning mechanisms to parse these structures further.

2.3 Auditory-visual integration of speech

We easily understand speech that we hear on the radio, but most of us cannot make sense of a speaker on a muted television. Indeed, speech is often seen as taking place in the auditory domain only. However, multiple studies have shown that the visual component of speech can also affect speech perception. For instance, in a noisy environment seeing the speaker has been shown to increase the intelligibility of speech (Sumby & Pollack, 1954).

McGurk and MacDonald (1976) discovered a phenomenon that later proved important for understanding auditory-visual integration of speech. When shown a video of the articulation /ga/ while /ba/ is simultaneously played from the loudspeakers, both adults and children perceive a third syllable /da/. This finding was the first display of *auditory-visual integration* in the brain: both auditory and visual syllables are processed in the brain, and an integrated percept is created according to both inputs. It is worth noting that the place of articulation for /da/ is approximately halfway between those for /ba/ and /ga/. However, in addition to the place of articulation, the integration process may also

influence at least the perception of the voice-onset-time of a plosive (Green & Kuhl, 1989).

Massaro (1984) studied further how visual information can affect the perception of phonemes. He presented adults and children with syllables from an acoustic continuum between /ba/ and /da/ combined with a visual articulation of either /ba/ or /da/. For both age groups, the syllables were more often categorised as /ba/ with a visual stimulus /ba/ and as /da/ with a visual stimulus /da/. Thus, the visual information drew the perception of the ambiguous auditory syllables toward the phonetic identity of the visual syllable.

Infants are able to match auditorially and visually presented speech early in development. They can match auditory vowels to the articulation of the same vowel on the screen already at the age of 2 months (Patterson & Werker, 2003) and this ability remains stable throughout the development (Kuhl & Meltzoff, 1982; Patterson & Werker, 1999; Patterson & Werker, 2002). This suggests that infants associate the visual and auditory inputs. Interestingly, Pons and others (2009) found that 6-month-old but not 11-month-old Spanish infants were able to perform auditory-visual matching of syllables /ba/ and /va/ – syllables that adult speakers of Spanish do not discriminate. However, 11-month-old infants from English-speaking environments retained the ability, as the phonemes /b/ and /v/ belong to separate categories in English language. Thus, perceptual narrowing for speech is not only observed in the auditory domain, but also in the auditory-visual perception of speech. However, by studying auditory-visual matching only, we cannot address the question of prospective integration of the two modalities because the successful matching may be due merely to sensitivity to frequently co-occurring stimuli.

Bristow and others (2009) recently studied cross-modal processing of phonemes in 2.5-month-old infants. They recorded ERPs to auditory-only test vowels after two repetitions of either a silent visual articulation of a vowel or an auditory vowel combined with a masked face. The phoneme identity of a test vowel was either the same or different from the preceding stimuli. A mismatch response was observed in the ERPs for test vowels that differed from the preceding auditory vowels. More importantly, a mismatch response was also observed for test vowels, the phoneme identity of which differed from that of the preceding silent visual articulations. In other words, the brain responses observed for an auditory vowel differed according to whether or not the preceding silent articulations were of the same phoneme identity. While this does not directly assess integration of speech sounds, it convincingly demonstrates that infants have acquired a cross-modal representation of vowels early in development. If the infants had failed to identify the silent articulations as speech, no mismatch response would have been elicited, as the test stimuli were acoustically identical. Additionally, if the infants had failed to detect the phonetic identity of the silent articulations, no mismatch response would have been elicited. Thus, the infants were able detect the visual stimuli as a speaking face, and more importantly to compare the visual phoneme identity to that of the acoustic test stimulus.

Rosenblum and others (1997) studied auditory-visual integration in 5-month-old infants. They habituated infants to an audiovisual /va/, after which they used two different dishabituation stimuli: an auditory /ba/ combined with a visual /va/ (perceived by adults as /va/) and an auditory /da/ with a visual /va/ (perceived by adults as /da/). The results

suggested that the infants distinguished the latter but not the former test case from the habituation stimuli suggesting that they successfully integrated the auditory /ba/ and the visual /va/ into a percept /va/. Burnham and Dodd (2004) tested 4-month-old infants in a similar study. However, they habituated infants with the syllables used by McGurk and MacDonald (1976), i.e., an auditory /ga/ combined with a visual /ba/ (perceived by adults as /da/). In an auditory-only post-test, the infants showed different responses for the not-previously-heard test syllables /da/ and /ba/, suggesting that they had indeed integrated the habituation syllables into a percept /da/. Furthermore, control group infants treated the test syllables equally, thus not perceiving the distinction. These results suggest that the infants are able to integrate auditory and visual syllables. It seems, however, that the integration is not always mandatory, depending also on the developmental stage and sex of the infant as well as the task (Desjardins & Werker, 2004).

3 Review of the methods

3.1 Event-related potentials

The brain's electrical activity can be recorded from the scalp with electroencephalography (EEG). An event-related potential (ERP) is acquired by averaging electrical potentials measured time-locked to a certain type of stimulus. ERPs are typically recorded simultaneously from multiple electrodes at various locations on the scalp. The electrodes are referenced to a specific reference electrode, typically placed either on the nose or on the mastoid processes. Alternatively, the average signal value of the scalp electrodes can be used as a reference. The scalp potentials reflect the activity of the cortical pyramidal cells. A single unaveraged potential consists mostly of electric noise, i.e., electrical activity unrelated to the stimulus. Therefore, an average of multiple epochs, often further band-pass filtered, is needed to observe the stimulus-related activity (Picton et al., 2000).

The ERPs contain positive (P) and negative (N) deflections from zero, which are named components. The components are defined either by their polarity and latency, their anatomical source in the brain, or in terms of the functional process with which they are associated (Otten & Rugg, 2005). The ERP components are named according to the number of order of the ERP (e.g., N1, the first negative component), the approximate latency of the peak of the component in milliseconds (e.g., N100, a negative component occurring 100 ms after the stimulus onset), or their function (e.g., mismatch negativity).

ERPs can be obtained for stimuli in a single modality or for cross-modal stimuli. The following text focuses on the auditory ERPs relevant for Studies I–III.

3.1.1 N1

The N1 in the auditory domain, also named the N100, is an ERP component elicited as a response to the onset of any sound presented in isolation. It is an exogenous component, i.e., it reflects the activation of the auditory cortex in response to a sound. Multiple processes contribute to the elicitation of the N1 (Näätänen & Picton, 1987). The N1 is strongly modulated by stimulus features as well as by the subject's state of arousal.

Recently, the N1 was found to reflect natural speech segmentation. Word-initial sounds elicited a larger N1 than sounds starting with a word-medial syllable (Sanders & Neville, 2003). Similarly, an enhanced N1 amplitude has been found to reflect the initial item of coherent units of three stimuli, or a low-probability transition in a stream of syllables (Sanders, Newport, & Neville, 2002) or tones (Abla, Katahira, & Okanoya, 2008). The modulation of the N1 according to the unexpectedness of a stimulus is in concordance with the previous findings that the N1 is enhanced by infrequent changes in repetitive acoustic stimulation (Korzyukov et al., 1999), in a fashion similar to the mismatch negativity (MMN) (Näätänen, 1992).

3.1.2 N400

The N400 is an auditory ERP component traditionally associated with semantic processing of words. It was first described in response to semantically anomalous words within sentences heard in isolation or written text (Kutas & Hillyard, 1980; Kutas & Hillyard, 1983). However, it was later discovered that the N400 was elicited also by congruent words. Interestingly, the amplitude of the N400 was found to be directly proportional to the goodness-of-fit between the sentence frame and the eliciting word (Kutas & Hillyard, 1984). Indeed, the current view is that the N400 is not a response elicited for anomalous words; rather it is a default response, which becomes reduced as the sentence context builds up aiding in the interpretation of a potentially meaningful stimulus (Kutas, van Petten, & Kluender, 2006).

Low-frequency words, i.e., words that occur relatively rarely in natural language, have been found to elicit a larger N400 than high-frequency words. However, this effect is strongest for sentence-initial words and diminishes towards the end of the sentence, when the N400 is suppressed by a meaningful sentence context (van Petten, 1995). In statistical learning studies, the N400 has been found to be largest for syllable or tone transitions with a low probability (Abla et al., 2008; Cunillera et al., 2006; Sanders et al., 2002). These results suggest N400 a general role not only in semantic processing of words, but also in probabilistic parsing of auditory streams consisting of similar items with differing statistical characteristics.

3.1.3 Neonatal ERPs

The newborn ERP to sounds is dominated by a single large slow positive wave unlike the more rapid positive and negative adult ERP components (Kushnerenko, 2001). In addition to this fronto-central positivity, negativity can be seen at mastoid and temporal sites. The positive component of the neonates shows similar characteristics for various auditory stimuli, such as tones (Huotilainen et al., 2003; Kushnerenko, Ceponiene, Balan, Fellman, Huotilainen, & Näätänen, 2002a) and speech sounds (Kurtzberg, 1984; Molfese, 1985). The response begins about 100 ms after stimulus onset, peaking around 250 to 300 ms.

Newborn infants also show a mismatch response similar to MMN observed in adults (Näätänen, Gaillard, & Mantysalo, 1978). The MMN can be observed in neonates for rare deviant sounds occurring among frequent standard sounds using the traditional Oddball paradigm (Alho, Sainio, Sajaniemi, Reinikainen, & Näätänen, 1990). Furthermore, it is also elicited for deviants that break a simple abstract rule in the auditory stream (Carral et al., 2005; Ruusuvirta, Huotilainen, Fellman, & Näätänen, 2003). However, the newborn MMN shows large variability between individual infants as well as between research designs, which is typical of newborn ERP studies in general. Consequently, the current research is focused on group-level effects rather than on the assessment of individual neonates. The inter-individual differences may be due to variability in the task difficulty for the individual infants or the varying neurological conditions or developmental states of the infants, even though neither can explain all variabilities (Trainor, 2008).

Additionally, the arousal state of the infant has a significant effect on the recorded responses. The newborn recordings are typically performed when the infants are asleep for practical reasons and to avoid movement artefacts. There are two major sleep states, active sleep and quiet sleep, which are differentiated according to the behavioural state of the infant and, when available, EEG, EOG, and EMG (Grigg-Damberger et al., 2007). As the EEG characteristics between these two sleep states differ, the part of the data acquired when the infants are in active sleep is usually chosen for analyses (DeBoer, Scott, & Nelson, 2005). It is also possible to average the data from both active and quiet sleep states, or even from all arousal states, when the data characteristics between the different arousal states are sufficiently similar.

3.1.4 Maturation of ERPs

Characteristics of a typical auditory ERP change dramatically during the first year of development. From birth to 3 months of age, the negativity at temporal sites becomes more positive. At the age of 3 months, almost all infants show a positive response across frontal, central, and temporal regions. Following the positive response, a slow negative wave appears between about 400 and 800 ms from stimulus onset (Trainor, 2008). Around 3 months of age, the positive peak can also be divided by a negative deflection around 200 to 300 ms from stimulus onset (e.g., Dehaene-Lambertz, 2000; Kushnerenko, Ceponiene, Balan, Fellman, Huotilainen, & Näätänen, 2002b).

By 6 months of age, the early negative deflection has matured into a component, and also other fast components can be observed (Kushnerenko, Ceponiene, Balan, Fellman, Huotilainen, & Näätänen, 2002b; Trainor, Samuel, Desjardins, & Sonnadara, 2001). From 6 months to 12 months of age, the peak identities remain stable, while the peak-to-peak amplitude differences keep changing (Kushnerenko, Ceponiene, Balan, Fellman, Huotilainen, & Näätänen, 2002b). It is difficult to estimate, whether the observed peaks are analogous to those of adult ERPs, but they seem to match those recorded from 3- to 9-year-old children (Ceponiene, Cheour, & Näätänen, 1998; Paetau, Ahonen, Salonen, & Sams, 1995). Thus, the waveforms keep changing notably even after the age of 1 year, and they reach the adult form only several years later.

Fetal brain responses can be recorded with MEG (see next chapter) at least from the 28th gestational week. The fetal responses contain large inter-individual differences. Typically, they may include two peaks: one at around 200 ms and another at around 400 ms after stimulus onset. Furthermore, a mismatch response to a frequency change can be recorded from the 28th gestational week (Draganova et al., 2005; Draganova, Eswaran, Murphy, Lowery, & Preissl, 2007; Huotilainen et al., 2005). This response is considered analogous to the later mismatch response, or MMN, with a latency of around 300 ms after the stimulus onset. Additionally, a late discriminative response can be seen around 500 ms after the stimulus onset from the 30th gestational week (Draganova et al., 2007).

The change-detection mechanisms undergo major changes during the first 4 post-natal months. As its characteristics vary notably between individual studies, the neonatal MMN has been widely discussed (see, e.g., He, Hotson, & Trainor, 2007). Indeed, the polarity of

MMN within the first post-natal months can be either negative (Alho et al., 1990) or positive (Leppänen et al., 2004) due to individual maturational factors as well as differences in research design. The stimulus domain may also affect the source location of the MMN (Dehaene-Lambertz & Baillet, 1998). Recent results suggest that the mismatch response develops from a slow, often positive wave in neonates to a rapid MMN-type negativity in 4- to 8-month-old infants (He, Hotson, & Trainor, 2009). Along the maturation, these two types of infant mismatch responses may even co-occur in the same infant for the same stimulus (He et al., 2007).

3.5 Magnetoencephalography

MEG is a non-invasive method for recording changes in magnetic field outside the scalp. The measured magnetic field fluctuations reflect changes in cortical electric currents (Otsubo & Snead, 2001). Due to the weakness of the measured magnetic fields, the technical demands are high. Similarly to EEG, MEG provides millisecond-scale information of cortical activation patterns. However, MEG is thought to provide more precise information concerning the location and orientation of the underlying sources (Cohen et al., 1990; Ko, Kufta, Scaffidi, & Sato, 1998; but see also Liu, Dale, & Belliveau, 2002). MEG and EEG can also be recorded simultaneously for complementary information (Huotilainen et al., 1998).

The magnetic fields are measured outside the scalp with SQUID (superconducting quantum interference device) sensors within a measurement helmet surrounding the head of the participant. The SQUID sensors are superconducting loops stored in liquid helium (Zimmerman, Thiene, & Harding, 1970). The sensors are typically of two types. A magnetometer is a sensor with a single loop, whereas a gradiometer includes two loops next to each other (the formation resembling the number "8"). Magnetometers provide a more sensitive measure than gradiometers, but they are also more sensitive to noise. Consequently, gradiometers are often used to provide a better signal-to-noise ratio. In a typical MEG recording, magnetometer and gradiometer sensor arrays are used simultaneously, and the sensor array providing the best measure is used for the final analyses.

The recordings are conducted in a magnetically shielded room to minimise the effect of external magnetic fields. Concomitantly, the auditory stimuli are not delivered through normal loudspeakers to avoid magnetic disturbances. Either plastic tubes connected to earpieces or piezoelectric speakers are typically used.

An event-related magnetic field (ERF) can be acquired in response to a certain stimulus in a fashion similar to an ERP: an average of several magnetic responses to the same stimulus type measured by an individual SQUID is calculated. The averaging reduces both cortical activity unrelated to stimulus processing as well as external magnetic noise. Unlike EEG, which detects brain activity from currents in any orientation and location, MEG is sensitive mainly to tangential components of the neural currents close to the sensors. Consequently, signals measured with MEG are principally generated within the fissural cortex.

A measure acquired with MEG that is comparable to ERPs is an *areal mean signal*. An areal mean signal is acquired by averaging ERFs recorded from SQUIDs over a specific region, such as above the temporal lobe (Service, Helenius, Maury, & Salmelin, 2007). To increase the quality of the measured signal, an areal mean signal can be calculated using vector sums of gradiometer pairs as an estimate of the signal for an individual channel.

Estimating the cranial source of the measured signal is called the *inverse problem*. The most common way to characterise the sources is by performing an equivalent current dipole (ECD) analysis (Williamson & Kaufman, 1981). The ECD analysis is performed by minimising the difference between the calculated and measured magnetic fields. An ECD represents the mean location, strength of activation, and the orientation of the current flow in the determined brain region.

A magnetic counterpart has been determined for the ERP components mentioned earlier. As the signal sources measured by MEG are not identical to those of EEG, the characteristics of the components can differ in relative amplitude, latency, and source location. The magnetic components in the ERFs are denoted with the letter m to differentiate them from their ERP counterparts (e.g., N1m, N400m).

3.6 Behavioural research methods

Behavioural research methods provided an important window into the perceptual and learning processes of infants decades before the brain research methods became available. Even today the results provided by the behavioural methods are adequate and often easier to interpret than those from brain imaging techniques presented in the earlier chapters.

The neonatal behavioural assessment scale (NBAS) includes a series of tests assessing the behaviour of the neonate (Brazelton & Nugent, 1995). The scale includes 28 behavioural and 18 reflex items that assess the infant's capabilities across different developmental areas, such as maintaining the autonomic systems, managing the motor behaviour, controlling the arousal states, and social interaction. In addition to providing a health screen, NBAS provides information of an individual infant's strengths, adaptive responses, and possible vulnerabilities.

To assess the function of a neonate's central nervous system, the Amiel-Tison neurologic assessment provides a comprehensive result in a relatively short time (Amiel-Tison, 1995). Because brain damage in the neonate is most typically located in cerebral hemispheres, the best predictive value should be found in responses depending on the upper control system and not in responses depending mainly on brainstem activity (Amiel-Tison, 2002). Consequently, higher neurologic functions such as adaptive capacity, primary reflexes and overall vision are evaluated. The result is not defined as the sum of individual scores, but rather as a gradation system based on clusters of signs and symptoms.

Various methods have been developed to specifically probe more cognitive aspects of an infant's central nervous system. The behavioural method available earliest in infancy to investigate learning is the high-amplitude sucking (HAS) paradigm developed by Siqueland and DeLucia (1969) and further modified by Eimas and others (1971). In HAS,

an infant is given a comforter to suck on, and the sucking rate is measured. The rate is used as an index of the infant's attention: changes in stimulation produce notable difference in the rate of high-amplitude sucking. Thus HAS can be used when measuring, e.g., discrimination, stimulus variability, or early representation of speech sounds. The paradigm is applicable with infants from birth to about 4 months of age.

Most of behavioural methods assessing learning are based in various ways on infants' headturn responses. In the operant headturn procedure, an infant is typically conditioned in a preliminary phase to turn his/her head in response to a stimulus change. After the conditioning, the discrimination of a wanted contrast can then be tested (Kuhl, 1979). In the headturn preference procedure, the infant learns to associate a headturn to a certain direction with a certain type of stimulus. Consequently, a preference for one stimulus over another can be tested (Kemler Nelson et al., 1995). The procedure has been used, for instance, to study infants' preference for infant-directed versus adult-directed speech (Fernald, 1985) and more recently to study learning of rules (Gomez, 2002) and phonotactic regularities (Chambers, Onishi, & Fisher, 2003) in speech.

Finally, the visual-fixation procedure uses infants' visual fixation times as an index of attention towards the stimuli. In this procedure, an infant sits while facing a computer screen. An attractive fixation image lures the infant's attention, after which auditory stimulation is played until the infant turns away. Upon this headturn, the auditory stimulus stops and the visual attractor disappears. The duration of the gaze during the stimulation is measured, and an average of multiple test trials is used to minimise stimulus-independent variation in the looking times (Polka & Werker, 1994). The visual-fixation procedure can be used across a wide age range, i.e., from 2-month-olds to 14-month-olds (Jusczyk, 1997).

The stimulus-alternation preference procedure (SAPP) is a variation of the visual-fixation procedure. In SAPP, test trials of two types are used: in half of the test trials, an auditory stimulus is repeated, and in the other half, two auditory stimuli are alternating. The rationale is that if the infants detect the difference between the two alternating stimuli, their looking times for those trials differ from the looking times to the trials in which the same stimulus is repeated. However, if the infants do not differentiate the two alternating stimuli, the looking times for both trial types are expected to be equal (Best & Jones, 1998). As infants sometimes halt to stare at the screen for a longer time with no apparent correspondence to the stimuli, a maximum duration for one test trial is often set beforehand. Depending on the design, the infants may show a preference, i.e., a longer looking time for the alternating trials (Best & Jones, 1998) or for the repeating trials (Maye et al., 2002).

4 Aims of the study

This study aimed at investigating the language learning processes during the first months of post-natal life. The focus was on statistical learning, i.e., how the infants might utilise statistical regularities embedded in speech to learn specific aspects of language. An equally important aim was to develop novel research methods for investigating the learning.

The aims of the individual studies were:
1. To develop a paradigm for measuring statistical speech segmentation with event-related brain measures and to test the reliability of the paradigm with adults in MEG (Study I).
2. To assess statistical speech segmentation skills of neonates using the paradigm developed for Study I (Study IIa).
3. To test the reliability of the statistical speech segmentation results in another group of neonates (Study IIb).
4. To assess auditory-visual integration of syllables in 5-month-old infants with ERPs (Study III).
5. To test effects of visual speech on learning of phoneme categories in 6-month-old infants (Study IV).

5 Methods

5.1 Participants

The participant groups in the four studies differed in their developmental stage, as the studies assessed different aspects of language learning. All participants were healthy volunteers with no diagnosed neurological or hearing disorders or medications, and all the infant participants were born full-term. The major characteristics of the participants are presented in Table 1.

The adult participants were recruited through an internet newsgroup for students (Study I). A written informed consent to participate in the study was acquired from the participants. The adult studies were performed at the BioMag Laboratory at Helsinki University Central Hospital.

The infant participants were recruited from the maternity ward (Studies IIa and IIb), and through announcements in baby activity groups and a magazine (Studies III and IV). In Studies IIa and IIb, the procedure was explained in detail to the parents of the neonates. In Studies III and IV, a parent was present throughout the experiment. A written consent was obtained from one or both of the parents of the infants. The infant studies were performed at the Women's hospital (Studies IIa and IIb) at Helsinki University Central Hospital, as well as at the Centre for Brain and Cognitive Development, Birkbeck, University of London (Studies III and IV).

Table 1   *Number, age, and gender of participants as well as the method used are presented for all studies. The number does not include rejected participants. Study II contained two datasets, called IIa and IIb. All of the neonatal subjects were tested 0.5–2 days after birth. Thus, the gestational age (GA) is given in the table.*

Study	Number	Mean age (range)	Gender (m/f)	Method
Study I	13	23.2 years (19 – 25)	8/5	MEG
Study IIa	15	GA 40 weeks 1 day (38+1 – 42+2)	8/7	ERP
Study IIb	15	GA 40 weeks 1 day (37+5 – 41+4)	7/8	ERP
Study III	17	21 weeks 3 days (20+4 – 23)	7/10	ERP
Study IV	48	6 months 11 days (5+20 – 6+20)	20/28	SAPP

5.2 Research design

5.2.1 Statistical speech segmentation

Studies I, IIa, and IIb used a similar design, first tested in adults with MEG (Study I) and then applied to neonates with ERPs (Studies IIa and IIb). Ten trisyllabic pseudowords were created out of a pool of thirty 300-ms-long Finnish syllables. The syllables were chosen from four subgroups according to their phonetic characteristics: a diphthong, a long vowel, /k/+vowel, and /s/+vowel. The pseudowords were balanced according to these subgroups, i.e., syllables were divided equally to all three positions of the pseudowords and next to syllables from all other subgroups to avoid any unwanted regularities in the stimuli. The resulting pseudowords are presented in Table 2.

The pseudowords were played for 15 minutes in a pseudorandom order, i.e., in an otherwise randomised order but keeping transitional probabilities from each pseudoword to all other pseudowords equal (1/9). There was a silent interval of 200 ms between individual syllables, equally within pseudowords and between pseudowords. As the syllables contained no morphological cues relevant to their location in the pseudowords, the only cue to the word boundaries in the syllable stream were the differences in the transitional probabilities between the syllables.
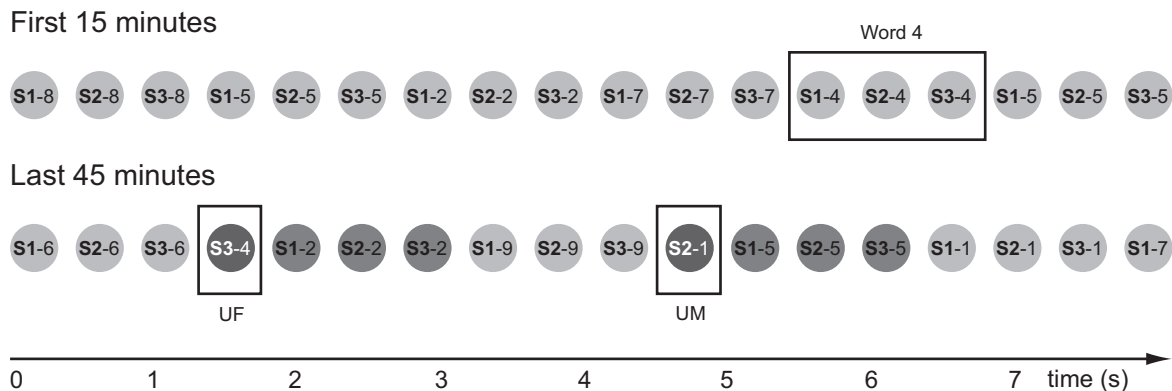
Figure 1 *Schematic of the experimental procedure. During the first 15 minutes, pseudowords were repeated in a pseudorandom order. Thereafter, medial (UM), final (UF), and novel unexpected syllables were added between every 2–4 pseudowords. Rectangles mark Word 4 as an example of one pseudoword as well as two unexpected syllables. Responses to pseudowords immediately following unexpected syllables (dark grey) were not included in the response average of standard pseudowords.*

After 15 minutes of exposure to the stream of pseudowords, i.e., 60 repetitions for each word, additional syllables were occasionally added between pseudowords (see Figure 1). One fourth of these added syllables were previously unheard novel syllables (/su/, /au/, /ua/, /ae/, /ei/, /ue/; acoustically similar to the previously heard syllables), and the rest

were medial and final syllables of the pseudowords. A syllable was added after every 2–4 (on average 3) pseudowords. The duration of the whole experiment was 60 minutes.

ERFs (with adults in Study I) and ERPs (with newborn infants in Studies IIa and IIb) were recorded while the syllable stream was played to the participants. The adult participants were instructed to ignore the sounds and to attend to a silent movie played before them. For the newborn participants, only the data recorded during active sleep were included in the analyses.

The responses for the pseudowords immediately following an unexpected syllable were not included in the averaged response to the standard pseudowords. Additionally, the responses to unexpected syllables, which were also present in the preceding word, were omitted from the averaged responses to unexpected syllables. Such cases were not present in Study IIb. Additionally, the novel unexpected syllables were omitted from Study IIb.

Table 2 *The pseudowords used in Studies I and IIa (left) and Study IIb (right). The pseudowords in both sets contained the same syllables, but the order of syllables within the words was different.*

Pseudowords in Studies I and IIa	Pseudowords in Study IIb
/öö ai ka/	/ka öö ai/
/ee ky sä/	/sä ee ky/
/yy sö ki/	/ki yy sö/
/ea ke sa/	/sa ea ke/
/ui si oo/	/oo ui si/
/ie ää kä/	/kä ie ää/
/sy kö eu/	/eu sy kö/
/so ia uu/	/uu so ia/
/ku ii se/	/se ku ii/
/ko aa iu/	/iu ko aa/

5.2.2 Auditory-visual integration of speech

A stream of auditory-visual syllables was played on a computer screen to 5-month-old infants while they were sitting on their parents' laps. ERPs were recorded simultaneously. The stimuli included all possible combinations of auditory and visual syllables /ba/ and /da/, i.e., VbAg (visual /ba/ and auditory /ga/), VbAb, VgAg, and VgAb. The syllables in which the auditory syllable did not match the visual articulation are typically perceived by adults as /da/ (VgAb), see Chapter 2.3, or /bga/ (VdAg). The stimuli were created with three different speakers.

The four stimuli were played in a pseudorandom order, keeping the probability of each combination equal, i.e., 0.25. The stimuli were stopped whenever the infants turned their gaze away from the screen. When necessary, attracting musical sounds were played to lure the infants to look back toward the screen. Also, the speaker was changed approximately every 40 s to keep the infants interested. The recording was continued as long as the

infants were happy and attentive, the entire duration thus becoming 4–9 min depending on the infant.

### 5.2.3 Visual influence on phoneme categorisation

Two groups of 6-month-old infants were exposed to auditory-visual syllables so that the auditory exposure was identical for the two groups, but the visual exposure differed between them. The auditory syllables were synthesised by Prof. Paavo Alku using a variation of Semi-Synthetic Speech Generation (Alku, Tiitinen, & Näätänen, 1999) to form a continuum of eight sounds from /ba/ to /da/ (hereafter, /ba1/, /ba2/, /ba3/, /ba4/, /da5/, /da6/, /da7/, /da8/). The phoneme boundary for adult English speakers was tested to lie between syllables /ba4/ and /da5/.

The presentation of the auditory syllables followed a unimodal distribution (see Figure 2), shown earlier to result in a one-category representation of speech sounds for another continuum (Maye et al., 2002). The speech sounds were dubbed onto video clips of visual articulations of /ba/ and /da/. The resulting auditory-visual exposures for the two groups were as follows:

The infants in the one-category visual group always saw the same visual articulation (either /ba/ or /da/, counterbalanced across the infants) with all the auditory syllables. The infants in the two-category visual group saw a visual articulation that best suited the auditory syllable, i.e., a visual articulation of /ba/ with the auditory syllables /ba1/–/ba4/ and a visual articulation of /da/ with the auditory syllables /da5/–/da8/.

The stimuli were always stopped when the infants looked away from the screen. When necessary, attractive musical sounds were used to reattract the infants' attention. The total duration of the exposure, excluding these prospective pauses, was 2 min 5 s.

In an auditory-only post-test using SAPP, the infants' discrimination of auditory syllables /ba3/ and /da6/ was tested. This test was designed to reflect the infants' category formation: if the syllables were perceived as equal, they were likely to be categorised in the same phonetic category. However, if the syllables were perceived as different, they were possibly seen as belonging to different phonetic categories. The order of tokens in the alternating test trials was modified to provide more balanced results, as suggested by pilot studies. Instead of simple "ABABABAB" alternation, we used the order "ABBAABBA", but counterbalancing the order between trials and infants.
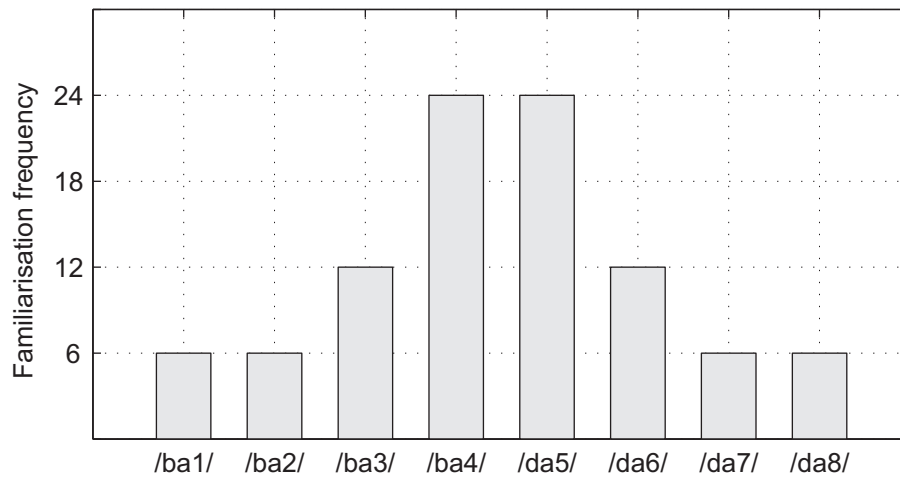
Figure 2   *The unimodal distribution of presentation of the auditory syllables. The syllables close to the typical adult phoneme boundary were presented more often than those close to the endpoints /ba1/ and /da8/.*

5.3 Data acquisition

In Study I, MEG was recorded from adult participants with a Neuromag Vector-view whole-head system containing 102 triple-sensor elements (two gradiometers and one magnetometer). Simultaneously, EEG was measured from electrodes Fz and Cz according to the 10-20 system with a nose reference. Additionally, horizontal and vertical electro-oculograms (EOG) were measured. The signals were bandpass filtered from 0.1 to 200 Hz, digitised at 603 Hz, and averaged on-line across trials over a time interval of −150–500 ms relative to the auditory stimulus onset. Trials with MEG or EEG/EOG signals exceeding 3000 fT/cm or ±150 µV, respectively, were discarded. The averaged ERFs and ERPs were baseline corrected to the 50-ms interval preceding stimulus onset. The Fz and Cz signals were averaged together to increase the signal-to-noise ratio.

In Studies IIa and IIb, EEG was recorded from newborn infants with Neuroscan II amplifiers at the electrodes F3, F4, C3, C4, T3, T4, P3, and P4 according to the 10-20 system with a linked mastoid reference. Additionally, horizontal and vertical EOGs were measured. The sampling rate was 250 Hz (Study IIa) or 500 Hz (Study IIb). The signals were bandpass filtered from 0.2 to 30 Hz and averaged off-line across trials over a time interval of −100–500 ms relative to the auditory stimulus onset. Epochs with EEG/EOG signals exceeding ±150 µV were discarded.

In Study III, EEG was recorded from 5-month-old infants using a 128-channel Geodesic Sensor Net with a vertex reference. The sampling rate was 500 Hz. The signals were bandpass filtered from 0.1 to 200 Hz and averaged off-line across trials over a time interval of −100–1100 ms relative to the auditory-visual stimulus onset. Individual channels contaminated by eye or motion artefacts were manually rejected, and epochs with over 20 bad channels were discarded. Additionally, trials during which the infant did

not attend the screen were excluded from further analyses. The trials were re-referenced to the average reference and then baseline corrected to the 200-ms interval preceding sound onset.

In Study IV, the 6-month-old infants were videotaped during the whole experiment, and the looking times for SAPP were scored off-line from the video recordings. A second coder unaware of both the hypothesis of the study and to which group the infants belonged also coded the looking behaviour of 16 randomly chosen infants. The interrater correlation was .983, and the average absolute difference between the two codings was 0.04 s, confirming the reliability of the scores. Additionally, the average number of times the infants looked away from the screen during the exposure was scored separately for the two groups of infants. The numbers did not differ significantly (t[46]=.536, p=.594; two-tailed t-test) suggesting no differences in the general attention level of the infants between the two conditions.

5.4 Statistical analyses

In Study I, the amplitude comparisons for the components of interest (N1, N1m, N400, and N400m) were performed for areal mean signals and ERPs over time intervals 90–130 ms (N1 and N1m) and 380–420 ms (N400 and N400m) using repeated measures analysis of variance (ANOVA) with Stimulus Category (different syllable locations) and Hemisphere (left or right) as factors for the ERFs and Stimulus Category as the only factor with the ERPs. ANOVA for the ECDs was performed equally to that for the ERFs. The Greenhouse-Geisser correction for sphericity was used for most comparisons (the exceptions are marked in the text). Post-hoc LSD tests were used to compare the relative strengths of activation for different stimulus categories.

In Studies IIa and IIb, a three-way repeated measures ANOVA with Stimulus Category (syllable location: S1, S2, or S3), Hemisphere, and Location (frontal, central, temporal, parietal) as factors was calculated for 50-ms bins that shifted in steps of 10 ms. A significant main effect of Stimulus Category in four consecutive bins was taken as evidence for differences between the responses for different syllables. The most significant bin was selected as the latency of the effect, and post-hoc LSD tests were performed to compare the relative strengths of activation for different stimulus categories at this latency.

In Study III, the ERP responses to the visual stimuli were first compared over the occipital cortex by performing a two-way repeated measures ANOVA with Stimulus Category (/ba/, /ga/) and Hemisphere as factors. Consequently, the auditory-visual stimuli were compared by performing a three-way ANOVA with Stimulus Category (VbAg, VbAb, VgAg, VgAb), Location, and Hemisphere as factors. The comparisons were performed in successive 100-ms bins starting from 90 ms up to 690 ms after sound onset.

In Study IV, a two-way repeated measures ANOVA with Visual Syllable (/ba/, /da/) and Trial Type (repeating, alternating) as factors was performed within the one-category group showing no interaction of Visual Syllable and Trial Type (p<.519) or main effect of Visual Syllable (p<.361). Thus, the data from the infants in this group was pooled in

subsequent analyses. A three-way repeated measures ANOVA was calculated with Condition (one-category, two-category), Trial Type, and Trial Order (ABBAABBA, BAABBAAB) as factors. Planned paired-samples t-tests were performed for each group comparing average looking times for the two trial types. A non-parametric two-tailed Fisher's exact test was additionally performed to compare the number of infants that showed preference for one of the two trial types.

5.5 Ethical considerations

Study I was approved by the Ethics Committee of the Hospital District of Helsinki and Uusimaa. Studies IIa and IIb were approved by the Ethics Committee for Paediatrics, Adolescent Medicine, and Psychiatry of the Hospital District of Helsinki and Uusimaa. Studies III and IV were approved by the Ethics Committee of the Department of Psychology, Birkbeck, University of London. The research methods in all studies were painless, safe, and designed to progress effectively to minimise the total duration of the experiments.

For adults, Study I was completely non-invasive. The only discomfort to the subjects was caused by the instruction to stay as still as possible and to avoid unnecessary blinking, preferable for the MEG method.

In Studies IIa and IIb, neonates recruited from the maternity ward were studied in a specific room within the same hospital building. During the experiment, the infants were apart from their parents for the duration of the experiment and transport, altogether about 75 minutes. Additionally, attaching and detaching the electrode stickers may have caused inconvenience to some of the infants. However, the neonate recordings were performed by a research nurse with wide experience in soothing and comforting infants. Most of the neonates slept through most of the experiment as well as the transport. The stimuli were normal speech sounds played at a comfortable volume. Thus, the stimuli were not likely to cause distress to the infants.

In Study III, ERPs were recorded from 5-month-old infants. The ERP sensor net was placed on the infant's head while the infant was entertained, the procedure typically causing no distress to the infant. During the preparations and the experiment, the infants were always sitting on their parents' laps. On signs of distress, such as crying or complete disinterest toward the stimuli, the experiment was discontinued, and the ERP net was removed from the infant's head.

In Study IV, the 6-month-old infants were sitting on their parents' laps with no recording devices attached to the infants. The duration of the whole experiment was less than 5 minutes. On signs of distress, the experiment was paused, and continued only if the infant seemed happy. In both Study III and Study IV, the stimulation consisted of a neutral female speaker articulating syllables on a video. The infants were generally interested in the speaking face, i.e., the stimuli were pleasant to them.

6 Results

6.1 Statistical speech segmentation

In the adult study (Study I), it was found that the location of the syllable within the pseudowords modulated the N1m amplitude in the areal mean signals with no significant differences between the hemispheres (see Figure 3). The N1m amplitude for the word-final syllables was found to be significantly larger than that for the word-medial syllables and for the unexpected final syllables. Modulation of the ECDs at the N1m latency was also found according to the syllable location.

Additionally, it was found that the N400m amplitude was modulated in a different fashion in the two hemispheres. No differences in the N400m amplitude were seen for syllables in different locations in the left hemisphere. However, in the right hemisphere the N400m amplitude acquired from the areal mean signal was smaller for the word-initial syllables compared with the word-medial syllables and most of the unexpected syllables (see Figure 3). Some modulation of the N400 amplitude of the ERPs was also observed.

The standard word-initial syllables were also compared to the word-initial syllables following an unexpected syllable, because both types of syllables were acoustically identical but present in a different context. The N1m amplitude for the standard word-initial syllable was significantly larger than those for word-initial syllables following unexpected medial and final syllables. However, no differences were found in the N400m amplitudes.
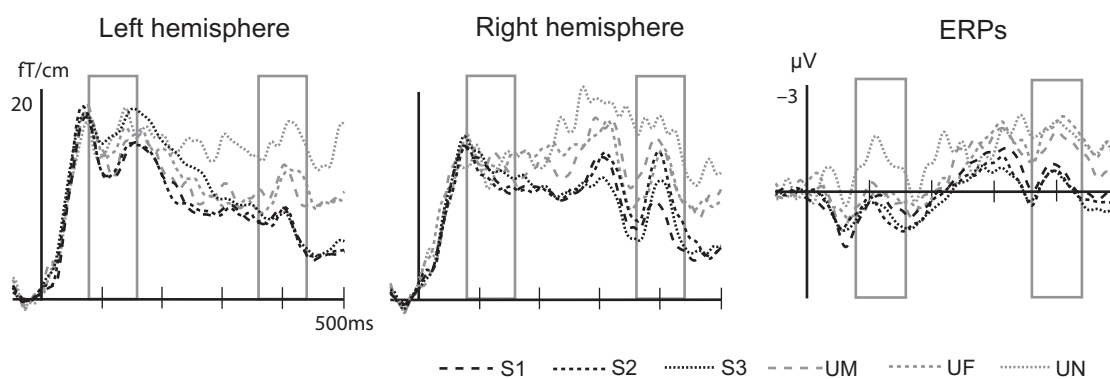

Figure 3   *Grand average areal mean signals of the MEG and ERPs elicited by syllables in different locations of the syllable stream: word-initial (S1), word-medial (S2), and word-final (S3) syllables, as well as unexpected medial (UM), final (UF), and novel (UN) syllables. The rectangles mark the time intervals used for testing the N1(m) and N400(m) amplitudes. The participants were healthy adults.*

In Studies IIa and IIb in neonates, the location of the syllable within the pseudowords was found to modulate the ERP amplitude at a time interval around 350 ms after the stimulus onset, as shown in Figure 4. Equal tests of the last 45 minutes including the unexpected

syllables and the word-initial syllables following the unexpected syllables did not provide significant main effects or interactions. This was likely due to the smaller number of trials included in the signal averages. The averages for the unexpected syllables contained notably fewer epochs than those for the within-pseudoword syllables.

Study IIb was designed to replicate the results acquired in Study IIa with an equal number of new participants. Differences in the ERP amplitude according to the syllable location were found at a similar latency as in Study IIa (see Figure 4). In a pooled analysis of Studies IIa and IIb, the word-initial syllable was found to elicit a more negative deflection compared to the word-medial and word-final syllables. The effect was robustly visible in the ERPs recorded from the frontal, central, and parietal electrode cites.
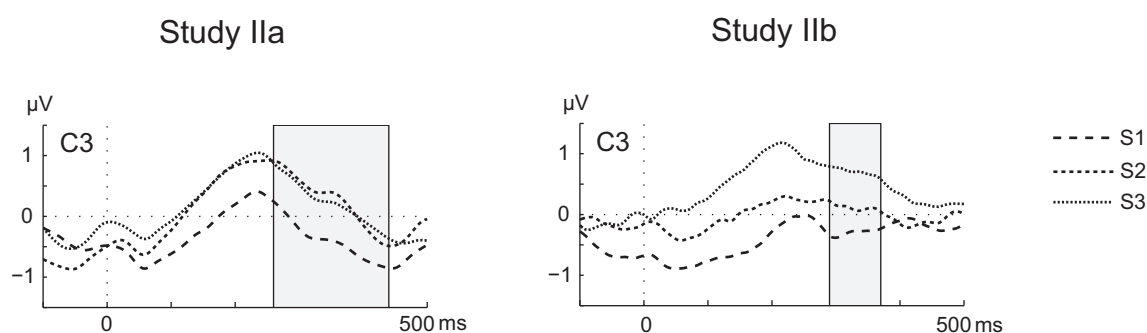
Figure 4   *Grand average ERPs of neonates at C3 electrode elicited by word-initial (S1), word-medial (S2), and word-final (S3) syllables in a continuous stream of syllables. The rectangles mark the time interval of significant differences between responses to different syllable types. Study IIb was a replication of Study IIa with minor modifications to the research design.*

6.2 Auditory-visual integration of syllables

In Study III, when the ERPs were analysed according to the visual stimulus (VbAb and VbAg vs. VgAg and VgAb), an early difference within the time interval 190–290 ms relative to the sound onset was observed in the occipital ERPs between the responses to stimuli containing a visual /ba/ and those to stimuli containing a visual /ga/. No other differences were observed associated with only the visual stimulus. Thus, the occipital cortex processed the two visual stimuli differently, which is natural as the articulations of /ba/ and /ga/ differ notably. Indeed, without these notable differences, the stimuli would have been unsuitable for the chosen research design. However, as the only observed differences were in the early occipital responses, the visual stimulus did not influence the later ERP characteristics *per se*. The lack of differences over temporal and frontal regions was especially important, as these regions typically show auditory-visual integration effects in adults.

Next, the ERPs for all four auditory-visual syllable combinations were compared. The ERPs to the coherent stimuli VbAb and VgAg did not differ significantly either from each other or from the combined stimulus VgAb (corresponding to the adult percept /da/).

However, the combination VbAg (adult percept /bga/) was more positive over frontal areas and more negative over temporal areas compared to the other three stimulus types (see Figure 5). The differences occurred at time range 290–590 ms after sound onset.
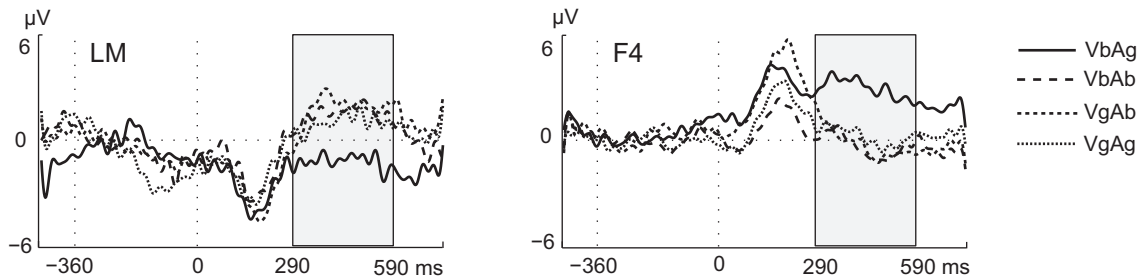
Figure 5   *Grand average ERPs of 5-month-old infants at left mastoid and F4 electrodes elicited by all auditoryvisual /ba/ and /ga/ syllable combinations. The rectangle marks the time interval of significant difference between the VbAg stimulus and the other stimuli.*

6.3 Visual effects on phoneme categorisation

In Study IV, the looking times (see Table 3) for the alternating and repeating test trials were compared between the infants in the one-category visual group versus the two-category visual group. There was a marginally significant interaction of Condition and Trial Type ($F[1,44]=3.525$, $p=.067$), suggesting that the two groups processed the two trial types in a different fashion. The pairwise t-tests confirmed that the looking times for the two trial types differed for the infants in the two-category group ($t[23]=-2.477$, $p=.021$), but not for the infants in the one-category group ($t[23]=.336$, $p=.740$).

The infants in the two-category group, exposed to both visual /ba/ and /da/ articulations, showed a looking bias toward the repeating test trials. The direction of the preference was equal to that found earlier for infants exposed to a bimodal auditory continuum in an auditory-only experiment (Maye et al., 2002). The Fisher's exact test confirmed that significantly more infants in the two-category group than in the one-category group showed longer average looking times for repeating test trials ($p=.017$).

Table 3   *Mean looking times (with standard error) of the infants according to group, i.e., which type of visual input the infants received, and test trial type.*

	Alternating test trials	Repeating test trials
One-category group	7.72 s (0.34)	7.64 s (0.36)
Two-category group	7.38 s (0.40)	7.83 s (0.39)

7 Discussion

The present studies probe the very early mechanisms available to perceive speech and learn structures of the ambient language. We found that statistical learning can be used to extract coherent units from a continuous speech stream from birth, and that this preattentive skill is still available in adults, reflected by ERPs. Additionally, infants are able to integrate auditory and visual speech into a coherent percept from early on. However, the integration may be limited to permissible combinations of the ambient language early in infancy. Visual speech can also guide learning of phoneme categories of the native language. These results are further discussed below.

7.1 Statistical speech segmentation reflected by ERPs

7.1.1 Adults

In healthy adults, the N400 ERP component was earlier found to reflect syllable location in a stream lacking morphological cues to word boundaries (Cunillera et al., 2006; Sanders et al., 2002). As the participants learned the pseudowords, the N400 amplitude for the word-initial syllable increased. A similar increase was found for the N1 amplitude, but it was only visible for participants with high learning results. We designed a new procedure in which we measured ERFs to syllables within and between pseudowords under an unattentive condition.

In the ERFs, we observed a smaller N400m amplitude for word-initial syllables compared to word-medial syllables and unexpected syllables. The larger N400m amplitude for the unexpected syllables was expected, as these syllables were in the "wrong sentence context", i.e., in a position in which they did not normally occur within the syllable stream. Indeed, this situation is quite comparable to semantically odd words, for which an enhanced N400 was originally found (Kutas & Hillyard, 1983). Our finding also reflects the recent findings that the N400 is enhanced for syllable or tone transitions with a low probability (Cunillera et al., 2006; Sanders et al., 2002).

However, in contrast to these recent studies we found the N400m to be decreased for the pseudoword-initial syllables. When comparing our results to those of earlier studies, it is important to note that all earlier studies have used a continuous speech stream with no acoustic pauses between the syllables, whereas in our study the syllable transitions were always accompanied with a 200-ms silent interval. The silence was equal between all syllables regardless of the probability of the syllable transition. Thus, even though the syllable triplets in our study are often referred to as pseudowords for convenience, they could as well be referred to as "pseudosentences". This marked difference between the research designs may account for the discrepancy in the N400m modulation.

Typically sentence-initial words produce a larger N400 than semantically context-appropriate later words in the sentence (van Petten & Kutas, 1990). Thus, it would be

reasonable to expect a larger N400 for the initial syllable in the triplet compared to the medial syllable, even though the comparison between sentences of multiple multisyllabic words and two first syllables of our pseudowords may be artificial. Still, it is useful to note that the areal mean signal that we used to measure the N400m amplitude is by its nature a measure of signal power, always giving a positive value regardless of the orientation of the source signal. Thus, it is not very useful for measuring changes in signal orientation. Additionally, the source signal measured by MEG differs from that measured by EEG, and this may also account to some extent for the differences.

Interestingly, we only observed the N400m effect in the right hemisphere. Right-handedness is a possible factor that can cause lateralisation of the N400 toward the right (see, e.g., Kutas, van Petten, & Besson, 1988), but almost a third of our participants were left-handers, so this may be an unlikely explanation. Further research is needed to assess the prospective lateralisation of the N400 for isolated syllables in a statistical learning paradigm.

In the previous brain research studies of statistical learning, the participants were instructed to pay attention to the sounds (Cunillera et al., 2006), or the sounds were even played to them in isolation (Sanders et al., 2002), whereas in our study, the participants were instructed to ignore the sounds and pay attention to a video unrelated to the stimuli. Earlier behavioural results suggest that when attentional resources are depleted, statistical word segmentation is compromised (Toro, Sinnett, & Soto-Faraco, 2005). However, in our study, which diverted attention to a task not especially demanding, the brain responses showed evidence of learning.

Our study included no behavioural test, and thus we were unable to separate the participants into groups of high and low learners. An earlier study that utilised such a separation found word-initial syllable to elicit a larger N1 compared to the other syllables in high learners, but not in low learners (Sanders et al., 2002). We found a general group-level N1m effect: the word-medial syllables elicited a smaller N1m than the word-final syllables. Thus, the location of a syllable within words seems to modulate the N1 amplitude both in normal speech (Sanders & Neville, 2003) and in morphologically poor speech. Similarly to the N400m results, the direction of our N1m modulation was opposite to the earlier results. However, as discussed earlier, our research design was different in many ways, and thus the results are not directly comparable.

The results of Study I thus suggest that adults can perform statistical speech segmentation even unattentively and that the learning is reflected in the N400m and, perhaps less reliably, the N1m ERFs.

7.1.2 Neonates

Using the same research design with newborn infants, in Study IIa we found the newborn ERPs to reflect the location of the syllable in the pseudowords. The ERP for the whole 1500-ms-long pseudoword had an ascending trend, the peak amplitudes in the responses to single syllables becoming more positive toward the end of the word. When analysing responses to single syllables, this word-long trend would have been lost in a

baseline correction performed separately for each syllable. Consequently, we opted not to perform any baseline correction. This is a functional approach within such data in which there are enough trials to average out the transient stimulus-independent fluctuations from the final ERP. We acquired on average over 1000 trials per stimulus type, a very large number compared to typical ERP studies of neonate subjects. On the other hand, the average number of trials for the unexpected syllables was less than 200, and thus the resulting ERPs were not robust enough for further analyses.

We used naturally spoken recorded syllables as stimuli as they were likely to be more efficiently perceived by the infants and adult participants than synthesised speech sounds. Syllables having similar morphological features were chosen from the many alternative utterances. However, there were slight differences left in the pitch, intensity, and intonation of the syllables. With the newborn infants, this was even considered an asset, because it was crucial for the infants to perceive all the syllables as being different from each other. As the syllable locations within the pseudowords were carefully controlled, the slight differences in the morphology of the syllables were unlikely to have any effects in the results. However, to avoid any ambiguity about whether the grand-average ERP characteristics actually resulted from the probabilistic differences in the syllables, the same syllables were used again in another set of ten pseudowords, but changing the syllable order in Study IIb. The results showed that even when the word-final syllables of Study IIa became the word-initial syllables in Study IIb, the same word-initial characteristics, i.e., a more negative ERP peak amplitude compared to the later syllables, were still observed.

In our analyses, the ERP differences between the syllables were significant only when the whole 60-minute experiment was included in the averages, i.e., they were not found significant when only the last 45 minutes were included. This would suggest that already during the first 15 minutes, the syllables were distinguished according to their location in the pseudowords. Concomitantly, it is probable that the learning of the newborn infants is quite rapid. However, with the current paradigm, it is difficult to accurately assess the rapidity of the learning. In a behavioural study, 8-month-old infants learned the transitional probabilities between nine syllables within two minutes (Saffran et al., 1996). Recently, Gervain and others (Gervain, Macagno, Cogoi, Pena, & Mehler, 2008) used near-infrared spectroscopy to study neonatal learning of abstract rule patterns. Their recordings took less than 25 minutes, which is a relatively short period, but also not fast enough to test for the rapidity of learning. Brain oscillation studies may in the future provide a novel way of studying learning processes, but so far the studies of infant brain oscillations have concentrated on social perception (see, e.g., Csibra, Davis, Spratling, & Johnson, 2000; Grossmann, Johnson, Farroni, & Csibra, 2007), and thus linguistic oscillation studies of infants have yet to be conducted.

7.2 Auditory-visual integration of syllables

Past studies of auditory-visual integration in infants have for practical reasons been behavioural. These studies have suggested that even infants are capable of auditory-visual

integration of syllables (Burnham & Dodd, 2004; Rosenblum et al., 1997), at least under specific conditions (Desjardins & Werker, 2004). However, our study was the first to directly probe the underlying brain processes. In our procedure, all four auditory-visual stimuli had an equal probability, and they were played to the infants in a pseudorandom order. If the infants detected a mismatch between the auditory and visual components, we expected to see this in a mismatch response. If they did not detect any mismatch, we did not expect to see any differences between the ERPs apart from minor differences in the primary obligatory responses of the visual cortex to the visual differences between the /ba/ and /ga/ articulations.

Indeed, we observed a robust mismatch response selectively for the combination of a visual /ba/ and an auditory /ga/. This was the expected response pattern, as the VbAg was a syllable combination that, in adults, leads to a combined percept /bga/, a phonotactic sequence that is not permissible in many languages. On the contrary, the opposite stimulus, VgAb, leads in adults to a fused percept /da/. The infants failed to detect the mismatch between the auditory and visual components of this stimulus, and thus it is likely that they successfully integrated the components into a unified percept.

The two visual stimuli, /ba/ and /ga/, had very different characteristics. Articulating /ga/ involves earlier and faster opening of the mouth than articulating /ba/, which explains the early differences in the ERPs to these stimuli. Also, the place of articulation of /ga/ is not visually clear enough to substantially restrict the perceptual outcome of the auditory-visual integration, whereas that of /ba/ (lips pressed together) restricts the perceptual outcome to /b/, /p/, or /m/ only (van Wassenhove, Grant, & Poeppel, 2005). This restriction may even account for the resulting disparity between the processing of the auditory-visual combinations VgAb (fusion effect) and VbAg (combination effect), the visual component /ba/ disallowing its fusion with the auditory component /ga/.

The latency (peaking at about 360 ms after sound onset) and scalp distribution (positive over frontal regions and negative over temporal regions) of the mismatch response observed for the VbAg stimulus resembles the auditory MMN in infants (Dehaene-Lambertz, 1994; Dehaene-Lambertz & Baillet, 1998; Dehaene-Lambertz & Pena, 2001). The adult MMN, typically suggested to reflect auditory sensory memory being elicited to rare stimulus changes in a repetitive acoustic environment, has also been found to reflect long-term memory in adults (Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001; Pulvermuller & Shtyrov, 2006). Similarly, it is possible that in Study III the mismatch response observed for the VbAg stimulus was due to a mismatch with the infants' long-term memory traces for permissible phonotactic sequences or auditory-visual relations learned during the first 5 months after birth.

7.3 Visual effects on learning of phoneme categories

Young infants form cross-modal representations of speech sounds (Bristow et al., 2009) and integrate auditory and visual speech (Burnham & Dodd, 2004; Rosenblum et al., 1997). Thus, it would not be surprising if the visual input had an effect when learning phoneme categories. Our study tested 6-month-old infants. At this age, slight tuning for

native-language vowels has occurred, but infants are only starting to learn the consonants of the ambient language more efficiently (Cheour et al., 1998; Polka & Werker, 1994; Rivera-Gaxiola et al., 2005; Werker & Tees, 1984). 6-month-old infants have also been shown to discriminate their native language from a foreign one watching silent articulations of the two languages, a skill that seems to disappear by 8 months of age (Weikum et al., 2007). We utilised a procedure used earlier to study effects of auditory phonetic distributions on the learning of phoneme categories (Maye et al., 2002). The procedure was modified by introducing a simultaneous visual component of speech, i.e., during exposure the stimuli were auditory-visual syllables. For the sounds, an auditory unimodal distribution was used, providing a strong cue of only one phoneme category. For the visual articulations, the infants received either the same visual articulation with all the speech sounds, i.e., a unary visual distribution, or two clearly different visual articulations combined with the auditory stimuli, i.e., a binary visual distribution. The binary visual distribution was designed so that the visual stimulus always provided a cue to "correct" (adult-like) categorisation of the speech sounds (see Chapter 4.2.3). The unary visual distribution was an additional cue to only one phonetic category, whereas the binary visual distribution was a cue to two different phonetic categories, thus in contradiction with the auditory input.

Indeed, we found that the infants in the group receiving the binary visual input showed evidence for two different phonetic categories in an auditory-only test following the exposure, whereas the infants receiving the unary visual input showed evidence for categorising all the speech sounds to one single category. The results provide strong evidence for the hypothesis that visual input can influence phoneme category learning. Importantly, the results were acquired in an auditory-only test, demonstrating that the visual effects were carried over to a situation, in which the visual cue was no longer available.

It would be premature to make a conclusion that the visual input is weighted more than the auditory input based on the fact that the infants that received the binary visual input and the unimodal auditory input formed a two-category representation of the speech sounds, i.e., one following the visual cue and contradicting with the auditory cue. Such a conclusion cannot be drawn before more is known about the relative strengths of these cues. For instance, a continuous unimodal distribution may be a weaker cue for one category than a clearly discontinuous binary distribution is for two categories. Additionally, the endpoint syllables /ba/ and /da/ used in the study may have a natural tendency to be categorised into different phonetic categories as their place of articulation is quite distinct and no natural language apparently seems to encompass the phonemes /b/ and /d/ into the same phoneme category.

Whether the infants integrated the auditory syllables from the /ba/–/da/ continuum with the visual articulations of /ba/ and /da/ during the exposure is an intriguing question. The results from both Study III and earlier behavioural studies (Burnham & Dodd, 2004; Rosenblum et al., 1997) suggest that the infants at this age are able to integrate visual and auditory speech into a fused percept. Thus, it would be reasonable to assume that they did integrate the auditory syllables with the relatively closely matching visual articulations also in this study. However, within these data it is impossible to assess the degree of

integration. It is important to note that two distinct-enough visual stimuli may *per se* be enough to provide the infants a cue to two distinct categories. In fact, it may be enough if the infants learn to associate the first half of the speech sounds with one visual stimulus and the other half of the speech sounds with another visual stimulus. This way, learning to associate different sets of speech sounds with different visual stimuli may already provide a cue to two distinct categories that is strong enough to overcome the conflicting auditory one-category cue. Additionally, the visual stimulus may provide advance information about the upcoming auditory event, thus contributing to the resulting auditory percept. Indeed, it would be interesting to test whether a picture of a cat displayed with the first half of the speech sounds and a picture of a dog with the second half would work as well as the visual articulations. However, regardless of whether such visual non-speech stimuli would or would not work, it is still possible that in our study the infants were integrating the visual and auditory speech components. The prospective integration may also have facilitated the categorisation or strengthened the visual categorisation cue.

It is also possible that the infants had already learned the /ba/–/da/ contrast from the bimodal auditory distribution present in the ambient language during the first post-natal months. Such an existing category would likely form a natural bias toward perceiving the two-category representation of the speech sounds. According to this interpretation, an auditory-visual one-category cue was necessary to overcome the natural bias toward two categories, whereas the auditory one-category cue combined with the visual two-category cue was too conflicting or weak, and thus the existing two-category representation remained valid.

7.4 Future directions

The results of Study I combined with those of other recent studies (Cunillera et al., 2006; Sanders et al., 2002) suggest that the N400 as well as the corresponding magnetic response N400m reflect transitional probabilities between syllables in adults. This effect has also been reported for non-linguistic tokens (Abla et al., 2008). Our study was the first to assess syllables isolated from each other by short silent intervals, thus providing more robust brain responses. However, our study lacked a behavioural measure of learning. The level of learning has been shown to correlate notably with the ERP characteristics (Abla et al., 2008; Sanders et al., 2002). Consequently, a behavioural test combined with our approach of using silent intervals between successive tokens would be important in order to acquire a more comprehensive view of the N400 in relation to statistical learning. Additionally, our results showed that adults can learn transitional probabilities between syllables even while engaged in a simple task, while an earlier study suggested that the learning is very restricted at least under a demanding task (Toro et al., 2005). Further research combining brain and behavioural tests is needed to assess the extent of learning under tasks that require different levels of attention.

Investigating statistical language learning in neonates, the results of Studies IIa and IIb suggest that neonates can learn transitional probabilities between consecutive syllables during active sleep. Furthermore, a recent study suggests that neonates can learn simple

rules embedded in speech (Gervain et al., 2008). As these were among the first studies to investigate complex learning processes in neonates, the stimulus regularities employed were relatively restricted. In the future, more complex rules and statistics should be used to test the boundaries of neonatal learning skills. Also, non-linguistic studies will provide information concerning the language-specificity of these skills. Thus far, research has concentrated on the availability of skills to learn different aspects of the sensory input, and we are still far from drawing conclusions about the practical applications of these skills. However, this will be an important direction toward which the research ultimately needs to move. Only when we know how infants employ different learning strategies, can we efficiently utilise our knowledge of the availability of these skills in an individual infant to estimate the state of that infant's development and, when necessary, intervene with remediation.

In Study III, we found that 5-month-old infants are able to integrate auditory and visual syllables into a fused percept. In the future, even younger infants should be studied. Recently, Bristow and others (Bristow et al., 2009) studied the formation of cross-modal phoneme representations in 2.5-month-old infants. However, their results did not provide information about the actual integration between auditory and visual modalities. To test for the actual integration at this age, the design proposed in Study III could be used, assuming that the auditory-visual mismatch response could be recorded this early in development. We observed the mismatch response for the misfit combination stimulus (VbAg) in 5-month-old infants. If the response was found in even younger infants, it would suggest that the permissible phoneme sequences are learned very early in development, or that the combined percept by its nature elicits a mismatch response unlike the "more natural" fused percept.

Finally, in Study IV we showed that visual speech can influence the learning of phoneme categories. To widen the scope of our results, a study using our research design, but replacing the visual articulations with non-linguistic visual stimuli should be conducted. This would shed light upon the role of visual speech in the learning, i.e., whether visual speech is given a special role, or if any clear and attractive visual cue can influence the learning in a similar fashion. A challenge for the future is to investigate whether infants actually use visual cues for learning, or whether visual cues are only used under special circumstances. Studying blind infants, completely deprived of these cues, could provide information about alternative learning strategies and the overall importance of visual speech in language development. Additionally, using natural phonetic continua present in a natural language non-native to the infant could erase the bias of previous learning experiences while still keeping the stimuli natural.

8 Conclusions

The main focus of the presented studies was on language learning in early infancy. With a novel paradigm, we found that the statistical properties of syllables embedded in a continuous stream modulate the N1m and N400m ERP components in adults. Similarly, the statistical properties modulated the neonatal ERP components, showing for the first time that neonates can extract such information, store it to long-term memory, and use it to perceive continuous speech.

We also found that even 5-month-old infants integrate auditory and visual syllables that result in a permissible percept in the ambient language. Additionally, the infants noticed a mismatch between the auditory and visual components of a stimulus that was not permissible in the ambient language and in adults results into an unnatural combined percept. Furthermore, we found evidence that visual input can influence learning of phoneme categories in 6-month-old infants.

Together, these results show that newborn infants have powerful computational capabilities for extracting statistical properties from speech. Additionally, the visual component of speech can be utilised efficiently early in development. Further research is needed to assess the role of these learning strategies in normal and atypical language development.

9 Acknowledgements

10 References

Abla, D., Katahira, K., & Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *Journal of Cognitive Neuroscience, 20*(6), 952-964.

Alho, K., Sainio, K., Sajaniemi, N., Reinikainen, K., & Näätänen, R. (1990). Event-related brain potential of human newborns to pitch change of an acoustic stimulus. *Electroencephalography and Clinical Neurophysiology, 77*(2), 151-155.

Alku, P., Tiitinen, H., & Näätänen, R. (1999). A method for generating natural-sounding speech stimuli for cognitive brain research. *Clinical Neurophysiology, 110*(8), 1329-1333.

Amiel-Tison, C. (1995). Clinical assessment of the infant nervous system. In M. I. Levene, & R. J. Lilford (Eds.), *Fetal and neonatal neurology and neurosurgery* (Second ed., pp. 83-104). London: Churchill Livingstone.

Amiel-Tison, C. (2002). Update of the amiel-tison neurologic assessment for the term neonate or at 40 weeks corrected age. *Pediatric Neurology, 27*(3), 196-212.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*(4), 321-324.

Best, C. T., & Jones, C. (1998). Stimulus-alternation preference procedure to test infant speech discrimination. *Infant Behavior and Development, 21*, 295-295.

Bosch, L., & Sebastián Gallés, N. (1997). Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition, 65*(1), 33-69.

Boysson-Bardies, B., & Vihman, M. M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language, 67*(2), 297-319.

Brazelton, T. B., & Nugent, J. K. (1995). *The neonatal behavioral assessment scale*. Cambridge, MA: Mac Keith Press.

Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., Baillet, S., & Mangin, J. (2009). Hearing faces: How the infant brain matches the face it sees with the speech it hears. *Journal of Cognitive Neuroscience, 21*(5), 905-921.

Burnham, D., & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology, 45*(4), 204-220.

Carral, V., Huotilainen, M., Ruusuvirta, T., Fellman, V., Näätänen, R., & Escera, C. (2005). A kind of auditory 'primitive intelligence' already present at birth. *The European Journal of Neuroscience, 21*(11), 3201-3204.

Ceponiene, R., Cheour, M., & Näätänen, R. (1998). Interstimulus interval and auditory event-related potentials in children: Evidence for multiple generators. *Electroencephalography and Clinical Neurophysiology, 108*(4), 345-354.

Chambers, K. E., Onishi, K. H., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition, 87*(2), B69-B77.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience, 1*(5), 351-353.

Cohen, D., Cuffin, B. N., Yunokuchi, K., Maniewski, R., Purcell, C., Cosgrove, G. R., Ives, J., Kennedy, J. G., & Schomer, D. L. (1990). MEG versus EEG localization test using implanted sources in the human brain. *Annals of Neurology, 28*(6), 811-817.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development, 61*(5), 1584-1595.

Csibra, G., Davis, G., Spratling, M. W., & Johnson, M. H. (2000). Gamma oscillations and object processing in the infant brain. *Science, 290*(5496), 1582-1585.

Cunillera, T., Toro, J. M., Sebastián Gallés, N., & Rodriguez-Fornells, A. (2006). The effects of stress and statistical cues on continuous speech segmentation: An event-related brain potential study. *Brain Research, 1123*(1), 168-178.

DeBoer, T., Scott, L. S., & Nelson, C. A. (2005). ERPs in developmental populations. In T. C. Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 263-297). Cambridge, MA: MIT Press.

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

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., & Pena, M. (2001). Electrophysiological evidence for automatic phonetic processing in neonates. *Neuroreport, 12*(14), 3155-3158.

Desjardins, R. N., & Werker, J. F. (2004). Is the integration of heard and seen speech mandatory for infants? *Developmental Psychobiology, 45*(4), 187-203.

Draganova, R., Eswaran, H., Murphy, P., Huotilainen, M., Lowery, C., & Preissl, H. (2005). Sound frequency change detection in fetuses and newborns, a magnetoencephalographic study. *NeuroImage, 28*(2), 354-361.

Draganova, R., Eswaran, H., Murphy, P., Lowery, C., & Preissl, H. (2007). Serial

magnetoencephalographic study of fetal and newborn auditory discriminative evoked

responses. *Early Human Development, 83*(3), 199-207.

Dror, R. O., Willsky, A. S., & Adelson, E. H. (2004). Statistical characterization of real-

world illumination. *Journal of Vision, 4*(9), 821-837.

Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in

infants. *Science, 171*(3968), 303-306.

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior

and Development, 8*(2), 181.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by

infants. *Proceedings of the National Academy of Sciences of the United States of

America, 99*(24), 15822-15826.

Gervain, J., Macagno, F., Cogoi, S., Pena, M., & Mehler, J. (2008). The neonate brain

detects speech structure. *Proceedings of the National Academy of Sciences of the

United States of America, 105*(37), 14222-14227.

Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological

Science, 13*(5), 431-436.

Green, K. P., & Kuhl, P. K. (1989). The role of visual information in the processing of

place and manner features in speech perception. *Perception and Psychophysics, 45*(1),

34-42.

Grigg-Damberger, M., Gozal, D., Marcus, C. L., Quan, S. F., Rosen, C. L., Chervin, R. D.,

Wise, M., Picchietti, D. L., Sheldon, S. H., & Iber, C. (2007). The visual scoring of

sleep and arousal in infants and children. *Journal of Clinical Sleep Medicine, 3*(2), 201-240.

Grossmann, T., Johnson, M. H., Farroni, T., & Csibra, G. (2007). Social perception in the infant brain: Gamma oscillatory activity in response to eye gaze. *Social Cognitive and Affective Neuroscience, 2*(4), 284-291.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*(5598), 1569-1579.

He, C., Hotson, L., & Trainor, L. J. (2007). Mismatch responses to pitch changes in early infancy. *Journal of Cognitive Neuroscience, 19*(5), 878-892.

He, C., Hotson, L., & Trainor, L. J. (2009). Development of infant mismatch responses to auditory pattern changes between 2 and 4 months old. *European Journal of Neuroscience, 29*(4), 861-867.

Huotilainen, M., Kujala, A., Hotakainen, M., Parkkonen, L., Taulu, S., Simola, J., Nenonen, J., Karjalainen, M., & Näätänen, R. (2005). Short-term memory functions of the human fetus recorded with magnetoencephalography. *Neuroreport, 16*(1), 81-84.

Huotilainen, M., Kujala, A., Hotakainen, M., Shestakova, A., Kushnerenko, E., Parkkonen, L., Fellman, V., & Näätänen, R. (2003). Auditory magnetic responses of healthy newborns. *Neuroreport, 14*(14), 1871-1875.

Huotilainen, M., Winkler, I., Alho, K., Escera, C., Virtanen, J., Ilmoniemi, R. J., Jääskeläinen, I. P., Pekkonen, E., & Näätänen, R. (1998). Combined mapping of human auditory EEG and MEG responses. *Electroencephalography and Clinical Neurophysiology, 108*(4), 370-379.

47

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

Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences, 3*(9), 323-328.

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

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in english-learning infants. *Cognitive Psychology, 39*(3-4), 159-207.

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior and Development, 18*(1), 111-116.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83*(2), B35-42.

Ko, D. Y., Kufta, C., Scaffidi, D., & Sato, S. (1998). Source localization determined by magnetoencephalography and electroencephalography in temporal lobe epilepsy: Comparison with electrocorticography: Technical case report. *Neurosurgery, 42*(2), 414-421.

Korzyukov, O., Alho, K., Kujala, A., Gumenyuk, V., Ilmoniemi, R. J., Virtanen, J., Kropotov, J., & Näätänen, R. (1999). Electromagnetic responses of the human

auditory cortex generated by sensory-memory based processing of tone-frequency changes. *Neuroscience Letters, 276*(3), 169-172.

Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *The Journal of the Acoustical Society of America, 66*(6), 1668-1679.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5*(11), 831-843.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science, 218*(4577), 1138-1141.

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.

Kurtzberg, D. (1984). Differential maturation of cortical auditory evoked potentials to speech sounds in normal fullterm and very low-birthweight infants. *Developmental Medicine and Child Neurology, 26*(4), 466.

Kushnerenko, E. (2001). Event-related potential correlates of sound duration: Similar pattern from birth to adulthood. *Neuroreport, 12*(17), 3777.

Kushnerenko, E., Ceponiene, R., Balan, P., Fellman, V., Huotilainen, M., & Näätänen, R. (2002a). Maturation of the auditory change detection response in infants: A longitudinal ERP study. *Neuroreport, 13*(15), 1843-1848.

Kushnerenko, E., Ceponiene, R., Balan, P., Fellman, V., Huotilainen, M., & Näätänen, R. (2002b). Maturation of the auditory event-related potentials during the first year of life. *Neuroreport, 13*(1), 47-51.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*(4427), 203-205.

Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory and Cognition, 11*(5), 539-550.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*(5947), 161-163.

Kutas, M., van Petten, C., & Besson, M. (1988). Event-related potential asymmetries during the reading of sentences. *Electroencephalography and Clinical Neurophysiology, 69*(3), 218-233.

Kutas, M., van Petten, C. K., & Kluender, R. (2006). Psycholinguistics electrified II. In M. J. Traxler, & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (Second ed., pp. 659-724). Netherlands: Academic Press.

Leppänen, P. H. T., Guttorm, T. K., Pihko, E., Takkinen, S., Eklund, K. M., & Lyytinen, H. (2004). Maturational effects on newborn ERPs measured in the mismatch negativity paradigm. *Experimental Neurology, 190*(Supplement 1), 91-101.

Liu, A. K., Dale, A. M., & Belliveau, J. W. (2002). Monte carlo simulation studies of EEG and MEG localization accuracy. *Human Brain Mapping, 16*(1), 47-62.

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

Massaro, D. W. (1984). Children's perception of visual and auditory speech. *Child Development, 55*(5), 1777-1788.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology, 38*(4), 465-494.

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science, 11*(1), 122-134.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*(3), B101-B111.

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

Molfese, D. L. (1985). Electrophysiological indices of auditory discrimination in newborn infants: The bases for predicting later language development? *Infant Behavior and Development, 8*(2), 197.

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

Myers, J., Jusczyk, P. W., Nelson, D. G. K., 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.

Näätänen, R. (1992). *Attention and brain function*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Näätänen, R., Gaillard, A. W., & Mantysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica, 42*(4), 313-329.

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic

response to sound: A review and an analysis of the component structure.

Psychophysiology, 24(4), 375-425.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001).

"Primitive intelligence" in the auditory cortex. *Trends in Neurosciences, 24*(5), 283-

288.

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.

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by english-

learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and*

Language, 43(1), 1-19.

Newman, R. S. (2003). Prosodic differences in mothers' speech to toddlers in quiet and

noisy environments. *Applied Psycholinguistics, 24*(4), 539-560.

Otsubo, H., & Snead, O. C. (2001). Topical review: Magnetoencephalography and

magnetic source imaging in children. *Journal of Child Neurology, 16*(4), 227-235.

Otten, L. J., & Rugg, M. D. (2005). Interpreting event-related brain potentials. In T. C.

Handy (Ed.), *Event-related potentials: A methods handbook* (pp. 3-16). Cambridge,

MA: MIT Press.

Paetau, R., Ahonen, A., Salonen, O., & Sams, M. (1995). Auditory evoked magnetic fields

to tones and pseudowords in healthy children and adults. *Journal of Clinical*

Neurophysiology, 12(2), 177-185.

Patterson, M. L., & Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior and Development, 22*(2), 237-247.

Patterson, M. L., & Werker, J. F. (2002). Infants' ability to match dynamic phonetic and gender information in the face and voice. *Journal of Experimental Child Psychology, 81*(1), 93-115.

Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science, 6*(2), 191-196.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development, 80*(3), 674-685.

Pena, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298*(5593), 604-607.

Petitto, L., & Marentette, P. (1991). Babbling in the manual mode: Evidence for the ontogeny of language. *Science, 251*(5000), 1493-1496.

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., Miller, G. A., Ritter, W., Ruchkin, D. S., Rugg, M. D., & Taylor, M. J. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology, 37*(2), 127.

Pinker, S. (1994). *The language instinct*. New York, NY: William Morrow.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20*(2), 421-435.

Pons, F., Lewkowicz, D. J., Soto-Faraco, S., & Sebastián Gallés, N. (2009). Narrowing of intersensory speech perception in infancy. *Proceedings of the National Academy of Sciences of the United States of America, 106*(26), 10598-10602.

Pulvermuller, F., & Shtyrov, Y. (2006). Language outside the focus of attention: The mismatch negativity as a tool for studying higher cognitive processes. *Progress in Neurobiology, 79*(1), 49-71.

Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P. K. (2005). Brain potentials to native and non-native speech contrasts in 7- and 11-month-old american infants. *Developmental Science, 8*(2), 162-172.

Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Perception & Psychophysics, 59*(3), 347-357.

Ruusuvirta, T., Huotilainen, M., Fellman, V., & Näätänen, R. (2003). The newborn human brain binds sound features together. *Neuroreport, 14*(16), 2117-2119.

Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition, 81*(2), 149-169.

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

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*(1), 27-52.

Saffran, J. R., & Wilson, D. P. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy, 4*(2), 273.

Sanders, L. D., & Neville, H. J. (2003). An ERP study of continuous speech processing. I. segmentation, semantics, and syntax in native speakers. *Brain Research Cognitive Brain Research, 15*(3), 228-240.

Sanders, L. D., Newport, E. L., & Neville, H. J. (2002). Segmenting nonsense: An event-related potential index of perceived onsets in continuous speech. *Nature Neuroscience, 5*(7), 700-703.

Service, E., Helenius, P., Maury, S., & Salmelin, R. (2007). Localization of syntactic and semantic brain responses using magnetoencephalography. *Journal of Cognitive Neuroscience, 19*(7), 1193-1205.

Siqueland, E. R., & DeLucia, C. A. (1969). Visual reinforcement of nonnutritive sucking in human infants. *Science, 165*(898), 1144-1146.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America, 26*(2), 212-215.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology, 50*(1), 86-132.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Toro, J. M., Nespor, M., Mehler, J., & Bonatti, L. L. (2008). Finding words and rules in a speech stream. *Psychological Science, 19*(2), 137-144.

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition, 97*(2), B25-34.

Trainor, L. J. (2008). ERP measures in auditory development research. In L. A. Schmidt, & S. J. Segalowitz (Eds.), *Developmental psychophysiology* (pp. 69-102). New York, NY: Cambridge University Press.

Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin & Review, 9*(2), 335-340.

Trainor, L. J., Samuel, S. S., Desjardins, R. N., & Sonnadara, R. R. (2001). Measuring temporal resolution in infants using mismatch negativity. *Neuroreport, 12*(11), 2443-2448.

van Petten, C. (1995). Words and sentences: Event-related brain potential measures. *Psychophysiology, 32*(6), 511-525.

van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory and Cognition, 18*(4), 380-393.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America, 102*(4), 1181-1186.

Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: The privileged status of speech for young infants. *Developmental Science, 7*(3), 270-276.

Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián Gallés, N., & Werker, J. F. (2007). Visual language discrimination in infancy. *Science, 316*(5828), 1159.

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

Williamson, S. J., & Kaufman, L. (1981). Biomagnetism. *Journal of Magnetism and Magnetic Materials, 22*(2), 129-201.

Zimmerman, J. E., Thiene, P., & Harding, J. T. (1970). Design and operation of stable rf-biased superconducting point-contact quantum devices, and a note on the properties of perfectly clean metal contacts. *Journal of Applied Physics, 41*(4), 1572-1580.