

Language development in infants: what do humans hear in the first months of life?

Alan Langus and Marina Nespor

International School for Advanced Studies (SISSA)
Language, Cognition and Development Laboratory – Trieste

Rhythm is order in movement
Plato, *The Laws*, Book II, 93.

Keywords: Language acquisition, infants' research, rhythm, rhythmic classes, Iambic-Trochaic Law.

Abstract

In this article we will discuss experimental work on language acquisition we have carried out in the last few years. We will first describe the most common methods used in infant research. Then we will concentrate on linguistic rhythm and the aspects of language that might be learned in the first year of life on the basis of signals contained in the speech stream. These include the basic level of rhythm carried by the two most basic phonological categories of consonants and vowels, and rhythmic alternation at the phrasal level and its signal to syntax. Because linguistic rhythm is one of the first aspects of language that infants perceive and represent, the research discussed may help to diagnose hearing problems and lead to new ways of how to train individuals with speech impairments.

1. Introduction

Rhythm pervades the universe: natural as well as cultural phenomena are governed by rhythm. The waves of the sea move rhythmically. Day and night alternate rhythmically. Our heart beats rhythmically, as does that of other animals. We breathe rhythmically and we walk rhythmically. Music in all cultures is governed by rhythm. So is dance. But what is rhythm? To our knowledge, the most essential – thus more general – definition of rhythm can be found in Plato: *Rhythm is order in movement*. While the elements that establish order may vary in different natural and cultural phenomena, order is always established by alternation, or the regular recurrence of a pattern.

What is rhythm in language? What are the elements that establish order in the movement of speech? Linguistic rhythm, as rhythm in music, is hierarchical in nature. There are different elements that establish rhythmic alternation – or order – at different levels of the rhythmic hierarchy (1, 2). For example, on the most basic level, linguistic rhythm is signaled by the space occupied by vowels in the speech stream (V%) and the standard deviation of consonantal intervals (ΔC) (3). However, if we move higher on the prosodic hierarchy – to the phonological phrase level – we find that rhythm is signaled through the

systematic recurrence of prosodic cues such as pitch, intensity or duration in the prominent elements of phonological phrases.

Because of infants' sensitivity to prosody, and because the alternation of rhythm at different levels of the rhythmic hierarchy signals different linguistic properties, it offers an ideal opportunity for investigating the early stages of language acquisition. The study of rhythm perception in early infancy can therefore not only tell us whether infants can discriminate languages on the basis of rhythm, but also when infants become sensitive to the specific cues that are carried by rhythm. This is particularly relevant, since rhythmic alternation has been shown to offer cues to the size of the syllabic repertoire (3), to the mean length of words (4) and to basic syntactic properties of language (5, 6). Since the sound pattern of a language, in particular its prosody gives cues to different grammatical properties, it is feasible that infants start their first steps into the acquisition of grammar at the pre-lexical stage, i.e. before knowing the meaning of words.

To give the reader a flavor of infant research on language acquisition, we will therefore begin with a brief description of the various techniques used to test newborns and young infants. We will then concentrate on what infants know about language when they are born and how they progressively tune in to the language of their environment. This will serve as a basis for discussing the aspects of linguistic rhythm infants know, how infants process linguistic rhythm at different levels of the rhythmic hierarchy, and what they learn about their language of exposure through rhythm during the first year of life.

2. How do we know what infants know?

Unlike in experiments with human adults, preverbal infants cannot be asked what they know or told to cooperate in specific tasks. Neither can they be extensively trained like laboratory animals to perform tasks that they intuitively do not understand. It is important to bear in mind that infants' voluntary behavior is highly limited: infants begin to intentionally grasp things around 5-6-months of age, utter their first words around their first birthday and begin to utter multi-word utterances around 2-years of life. As a consequence, to investigate what infants know before they can overtly tell us, researchers have to rely on ingenious ways of interpreting and using infants' naturally occurring responses to environmental stimuli. In the following, we briefly describe the arsenal of research methods ranging from behavioral methods that rely on infants' behavioral responses such, as sucking and looking, to neural responses in the brain measured with imaging techniques such as EEG and fNIRS.

2.1 Behavioral methods

Sucking is one of the first behavioral instincts of infants. While nutritive sucking is essential for infants to acquire food, non-nutritive sucking is considered a natural reflex for infants to demonstrate their need for contact, to feel secure and relaxed. As infants tend to suck more when presented with novel environmental stimuli than with familiar ones (7), the reflex has extensively been used to determine infants' sensitivity to speech (8). In a typical non-nutritive sucking experiment a pacifier is held in the infants' mouth.

The pacifier is connected to two loudspeakers that react to the subject's sucking so that sucking with enough energy causes the loudspeakers to emit a linguistic sound, for example *ta*. In the first phase of the experiment, called habituation, each high amplitude sucking triggers the same sound. After a while the subjects suck less and when the sucking rate reaches a habituation criterion (usually 50%), half of the neonates – the experimental group – hears a sound that differs minimally from the habituation sound, for example *da*. The other half of the subjects – the control group – keeps listening to the habituation sound. If the new stimulus triggers high amplitude sucking while the old one makes the subject suck progressively less, it can be concluded that the infant can distinguish already at birth two syllables that differ only in the consonantal onset.

The second group of behavioral methods used in infant research relies on infants' looking-behavior. While it may seem counterintuitive to use infants' looking-behavior to determine what infants know about the sound of their mother tongue, these methods are among the most common ones for investigating infants' knowledge about spoken language. Looking-time experiments rely on the observation that infants tend to be more attentive to environmental stimuli that they have not seen or heard before than to familiar ones. As a consequence, they look longer to novel than to familiar stimuli (9). For example, in a typical looking-time experiment an infant sees a visual stimulus on a screen, for example a check-board, and hears simultaneously an auditory stimulus until the infant's looking decreases to a criterion (usually 50%). When the infant has been habituated, the auditory stimulus changes while the visual stimulus remains the same. If the infant discriminates the new sound from the familiar one, it is expected to regain its interest in the visual display. The difference between the looking-time to novel and familiar auditory stimuli is taken as a measure of discrimination.

While simple looking-time measures have successfully been used even in newborns the methodology requires the experimenter to code the looking-behavior manually offline. To reduce the manual work, to avoid subjective biases of the coders and to obtain a more precise measure for infants' looking-behavior, today researchers use eye-trackers. Eye-trackers emit infrared light that is reflected on the cornea of the eyeball. This reflection of the light on the cornea is recorded by a sensor that then calculates the exact point of gaze of each eye up to 120 times per second. Because modern eye-trackers no longer need to be fixed on the subjects head, the eye-tracking method allow researchers to adapt classic looking-time paradigms and behavioral methods to measure infants' behavioral reflexes to environmental stimuli highly precisely in real time. However, because during the first weeks of life the human eye goes through considerable changes, it is only possible to measure eye-movements after 3-months of age.

2.2 Neuroimaging methods

Because human infants lack fine control and coordination over their motor behavior, researchers have increasingly turned to imaging techniques that measure neural responses to environmental stimuli. As neuroimaging techniques do not require an overt behavioral response they are ideally suited for studying cognitive and linguistic processing in very young infants. With the advancement of brain imaging techniques, it has become

increasingly possible to investigate the biological foundations of language with very young infant populations and even with newborns. In recent years, using advanced imaging techniques; researchers have therefore also started to study the specific brain areas that are dedicated to language perception in newborns and young infants. This line of research mainly uses two imaging techniques.

Electroencephalography (EEG) is the recording of electrical activity along the scalp by measuring the voltage fluctuations resulting from ionic current flows within the neurons of the brain. EEGs can detect changes over milliseconds. Electroencephalography (EEG) and event-related potentials (ERPs) provide noninvasive methodological techniques that allow researchers to examine the relation between brain and behavior in infancy. Because EEG and ERPs can be utilized across the entire lifespan, they provide two of the very few methodological tools that can directly measure and compare cognitive and linguistic processing at different ages. Even though EEG and ERPs both reflect the electrical activity of the brain, and both are collected in a similar manner, they represent slightly different aspects of brain function: EEG measures the brain's ongoing electrical activity; ERPs show changes in electrical activity in response to a discrete stimulus or event. EEG provides information regarding the resting state of the brain, synchrony between regions, or spectral changes in response to a cognitive event. In contrast, deflections in the ERP reflect specific aspects of sensory and cognitive processes associated with various stimuli. Due to the high temporal resolution, ERPs are well suited for studying the mental processing of environmental stimuli as they unfold.

However, infant research is increasingly also interested in understanding brain organization during the first weeks of life. fNIRS (functional near-infrared spectroscopy) is a non-invasive imaging method ideally suited to test auditory competence in newborn infants. fNIRS measures infants' brain activity through hemodynamic responses associated with neuron behavior. It relies on the fact that when monochromatic light travels through a medium, some of it is absorbed in the medium, some of it is scattered and some of it is transmitted. For example, familiar auditory stimuli cause a decrease of oxyhemoglobin and novel auditory stimuli an increase in the oxyhemoglobin (BenavidesetalPNAS). The properties of the vascular response measured using fNIRS are therefore comparable to those described for the BOLD (blood oxygen level dependence) effect in fMRI (10). This makes the methodology highly relevant for studying brain organization and brain responses in very young infants who cannot be tested with fMRI. In fact, fNIRS has successfully been used to show the cortical organization of the newborn brain (10), memory for words (11) and cortical processing of linguistic structures (12). For a recent overview of the methodology see Gervain et al. (13).

3. Tuning into the language of the environment

Humans start acquiring language from birth (10, 12), if not before (14, 15). The auditory system of humans becomes functional around the 25th week of gestation. This enables the fetus to begin processing linguistic information before birth. The auditory information from the environment is filtered and frequencies above 1000 Hz are attenuated (16). This

suggests that they have, at best, limited access to the acoustic information necessary for discriminating segmental information, especially consonants. The fetal hearing experience may thus affect some aspects of speech perception more than others. However, the sound that does reach fetuses provides sufficient information for the perception of suprasegmental aspects of speech. Importantly, fetuses also encode speech information into memory—especially suprasegmental information. Both newborns and fetuses demonstrate the ability to discriminate their native language from a foreign language (17-20) and their mother's voice from another woman's voice (21, 22).

Because most fine-grained segmental information is filtered out before it reaches the fetus, fetuses can encode only some suprasegmental characteristics of speech. It is therefore interesting that human infants are prepared for language perception the day they are born. For example, Peña et al. (10) played newborn human infants speech streams, the same speech streams played backward and silence while measuring infants' brain activity with fNIRS. Backward speech is the best possible control, since playing speech backward preserves the physical characteristics of the speech signal, but results in an acoustic realization that the human vocal tract cannot produce. Importantly, newborn infants' brain responses showed an increased cerebral activity in the left hemisphere, just like adults, when listening to speech as compared to when listening to backward speech or silence. These results therefore suggest that the ability to perceive speech does not require experience with spoken language after birth, and that the cortical areas responsible for the perception of spoken language are functional already at birth.

Newborn human infants also show remarkable abilities in speech perception that they cannot have developed in utero. During the very first days of life, newborns distinguish any pair of phonemes attested in one of the languages of the world (8, 23, 24). This ability to discriminate phonemes matures around 28 weeks of gestation as shown also in pre-term infants (15). Furthermore, it has also been shown that newborns can distinguish the location of word primary stress: they hear the difference between *pápa* and *papá* or *méta* and *metá* (25). The ability to distinguish both phonemes and the phonological realization of lexical stress shows that at birth human infants are capable of distinguishing all the sounds used in natural languages. At birth human infants are thus ready to learn any language to which they may be exposed.

The question is how infants tune in to the sounds of their language(s) of exposure. Adult humans are not particularly good at distinguishing sounds they are not familiar with. When perceiving sounds from foreign languages they tend to assimilate these to sounds that exist in their own language. For example, a native speaker of French – a language with systematic word final stress – loses the capability of distinguishing stress in other locations. In an experiment where monolingual native speakers of French were asked to say if a word they heard was *wásuma* or *wasúma*, were at chance (26, 27). Similarly, monolingual speakers of Japanese - a language with very restricted possibilities of consonant clusters - asked to say if they heard *ebzo* or *ebuzo* were at chance (28).

Infants slowly restrict their discrimination abilities to speech sounds that are systematically present in their environment. Foreign vowels start being ignored towards

the sixth month of life (29, 30) and foreign consonants towards the ninth month of life (31). By the first year of life infants have lost the ability to discriminate phonemes from foreign languages that are not present in their mother tongue. This focusing on the sounds present in the environment has expressively been called *Learning by forgetting* (32): infants learn to ignore sounds that are not used to distinguish meaning in the language of their surrounds, or that are not used systematically. Interestingly, they ignore not only the sounds of foreign languages but also the individual variability of speakers. For example, if certain individuals speak Italian with a uvular r [ʀ] rather than an alveolar r [r], an infant successfully learns to ignore this difference.

While it is difficult to sketch the exact time frame in which infants narrow down to specific aspects of spoken language, there is some evidence that suggests that native language influences vocal production already at birth. For example, a recent study monitored the early sound production of French and German newborns and found that the melodic contours of the cries of the two groups differed significantly (33). The cries of the French newborns showed a raising melody while those of the German group had a falling melody. These differences between newborns in the two linguistic environments suggest that the influence of the native language begins in utero. The tuning into the native language may therefore at least partially be already present in newborns, and gradually shape the perception of spoken language throughout the first year of life.

4. The basic rhythmic level signals syllabic structure

Having established that human infants are born with the ability to tell apart the different segments of the world languages, we can now begin our journey into the perception of rhythm. The basic rhythmic level accounts for the different rhythmic classes attested in the languages of the world. The presence of rhythmic classes was first proposed by Lloyd James (34) that English sounds very different from Spanish: the first resembles the noise of messages in the Morse code; the second the sound of a machine gun (34). There have been several proposals as to which phonetic element is responsible for these different rhythms. For example, the specific elements that established isochrony are interstress intervals in English, and syllables in Spanish (35) and all languages belong to one of these rhythmic classes: they are either stress-timed or syllable-timed (36). A third class was added later to account for the rhythm of Japanese (37), where the phonological constituent that establishes isochrony was proposed to be the mora. However, subsequent work by different phoneticians revealed that isochrony between different constituents is not actually present in the speech signal (38, 39).

Which property of the signal is then responsible for the machine-gun vs. Morse code effect, i.e. for the undeniable rhythmic difference between English, Dutch or Russian on the one hand, and Spanish, Greek or Italian, on the other hand? These languages differ in their syllable structures: stress-timed languages have a greater variety of syllable types than syllable-timed languages and as a result, they have heavier syllables. Furthermore, in stress-timed languages, unstressed syllables usually have a reduced vocalic system (sometimes reduced to just one vowel, schwa), and unstressed vowels are consistently shorter than stressed vowels, or even absent (40). These features combine with one

another to give the impression that some syllables are far more salient than others in stress-timed languages, and that all syllables tend to be equally salient in syllable-timed languages (41). It was therefore proposed that the measures responsible for this basic rhythmic structure are the space occupied by vowels in the speech stream ($V\%$) and the standard deviation of consonantal intervals (ΔC) (3).

This proposal of basic linguistic rhythm where languages fall into different rhythmic classes on the basis of the percentage of time occupied by vowels and on the regularity with which vowels recur – that in turn predicts the richness of the syllabic repertoire – has received considerable support from infant studies. For example, French newborns discriminate two languages such as English and Japanese that belong to different rhythmic classes, but not languages such as English and Dutch that belong to the same rhythmic class (20). It can thus be said that newborns identify the rhythmic class of their language of exposure. In fact, infants fail to discriminate between languages belonging to the same rhythmic class throughout the first months of life (42). Only 4-month-old bilingual infants exposed to two languages from the same rhythmic class (Spanish and Catalan) can discriminate the two languages they are actually exposed to (43). There is thus strong evidence for the proposal that during the first months of life infants rely primarily on the basic rhythmic level to discriminate between languages. Telling apart languages from the same rhythmic class appears to require experience with the specific languages.

What could infants learn through the identification of the rhythmic class of their native language? The ability to discriminate languages from different rhythmic classes shows that infants – in addition to being able to discriminate between different phonemes – can use quantitative information about consonants and vowels to tell apart languages they may encounter in their surrounding. These results furthermore suggest that newborn infants may have an idea of the richness of the syllabic repertoire of their language of exposure. A high $\%V$ implies that there are not so many consonants, and the syllables must be quite simple. A low $\%V$ implies instead many consonants, thus a complex syllabic structure (3). A simple syllabic structure is in turn predictive of word length since the number of possible monosyllables that could be used as alone standing words is restricted. This is the case, for example, of Japanese, where most words are polysyllabic. Languages with complex syllabic structure – such as English or Dutch – are instead rich in monosyllables. The identification of the rhythmic class could thus also offer a bias as to the mean length of common words and thus provide a valuable aid to speech segmentation (4).

The sensitivity to this basic level of linguistic rhythm in early stages of cognitive development is also important because there is a partial division of labor between vowels and consonants. While the main role of consonants concerns the lexicon, the main role of vowels is that of allowing the identification of the rhythmic class as well as of the development of a bias as to the mean length of common words (44). For example, in an artificial speech experiment adult participants prefer to use consonants to vowels in word identification (45) and they tend to use relations among vowels to extract generalizations while disregarding the relations among consonants for the same purpose (46). These

asymmetries are not by-products of low-level acoustic differences between consonants and vowels (47, 48), but result from the categorical representations of consonants and vowels themselves. This bias for using consonants for word learning and vowels for rule learning is found in infants by the latest of 12 months of age (49). The ability to perceive linguistic rhythm – to discriminate, represent and use phonemes categorically – from the earliest stages of language acquisition may therefore be necessary for the acquisition of words and rules of the mother tongue.

5. Phrasal rhythm signals word order

At a hierarchical level higher than that of the alternation between vowels and consonants, rhythm is determined by the alternation of stresses. It is important to note that the different levels of the prosodic hierarchy are organized so that lower levels are exhaustively contained into higher ones (50). This is best exemplified by considering the prosodic constituents most relevant for the present paper: the Phonological Phrases. The Phonological Phrase extends from the left edge of a phrase to the right edge of its head in head-complement languages; and from the left edge of a head to the left edge of its phrase in complement-head languages (1). While the number of Phonological Phrases contained in an Intonational Phrase that immediately governs them may vary, Phonological Phrases never straddle Intonational Phrase boundaries: Phonological Phrases are exhaustively contained in Intonational Phrases.

The most well known use for Phonological Phrases is for speech segmentation. Because syntactic structure is automatically mapped onto prosodic structure during speech production (1), many prosodic cues signal syntactic and word boundaries. For example, among the most common cues for Phonological Phrase boundaries is final lengthening (51-54). Adult listeners use phonological phrase boundaries to identify where words begin or end (55). Similarly, 10-month-old infants can use phonological phrase boundaries to find words from continuous speech (56). Furthermore, because prosodic boundaries are necessary for infants to detect rule-like regularities from continuous speech (57), phonological phrase boundaries may also help infants to extract rules from different levels of the prosodic hierarchy (58).

However, the systematic recurrence of Phonological Phrases is also important because the location as well as the physical manifestation of stress in Phonological Phrases give a cue to the word order of a language. Of the six logically possible orders of Subject (S), Object (O) and Verb (V), the basic word order of the great majority of the languages of the world (76%) is either SOV – as in Turkish, Japanese or Basque - or SVO – like English, Hausa or Greek (59). Nespor et al. (6) proposed that the stress in Phonological Phrases differs systematically between SVO and SOV languages. Languages with the SVO order have stress in final position (e.g. on the Noun in a Verb-Noun sequence) manifested mainly through duration, i.e. prominent syllables are longer than less prominent ones. Languages with the SOV order have stress in initial position (e.g. on the Noun in a Noun-Verb sequence) manifested mainly through pitch and intensity, i.e. prominent syllables are more intense and have a higher pitch than less prominent ones.

Given the uniformity of word order across constituents of different types within one language, could the specific word order of the language of exposure be inferred from the manifestation of phrasal rhythm in the prelexical period? To begin answering this question, French 6- and 12-week-old infants were exposed to French (SVO) and Turkish (SOV) sentences that differed solely on the location of relative prominence within phonological phrases (5). Turkish and French were chosen in this study because their syllabic structure is similar in complexity and because they both have word final stress. To eliminate the possible effect of segments – e.g. French but not Turkish has a uvular *r* [ʁ]; Turkish but French has a high-unrounded posterior vowel – all sentences were resynthesized as to have the same segments for each category. The findings of the study show that infants at this young age could in fact discriminate Turkish and French sentences exclusively on the basis of phrasal rhythm, highlighting the fact that phonological phrase level rhythm is perceived already during the first months of life.

If infants can discriminate two languages solely on the basis of their phrasal prominence, are they also able to hear if in a sequence of alternating weak and strong beats, the strong beat is initial or final in its group? This puzzle could be solved if a perceptual mechanism could indicate if a strong element is initial or final. A perceptual mechanism first proposed to account for grouping in music (60-62) – the Iambic-Trochaic Law (ITL) – was proposed to be responsible for the different acoustic manifestation of phrasal stress depending on whether it is initial or final in its phrase (6). The ITL states that a sequence of sounds that alternates in duration is grouped iambically (weak-strong), while a sequence of sounds that alternates in intensity – and pitch for speech at the phonological phrase level – is grouped trochaically (strong-weak). Whether a language has the Verb-Object or the Object-Verb order is thus signaled by phrasal rhythm.

Adults and 7-month olds were tested for their memory of sequences of syllables that alternate in either pitch or duration (63). Exposed to flat syllables in the test phase, adults were better at remembering pairs of syllables that during familiarization had short syllables preceding long syllables, or high-pitched syllables preceding low-pitched syllables. Instead, infants familiarized with syllables alternating in pitch, showed a preference to listen to pairs of syllables that had high pitch in the first syllable. However, no preference was found when the familiarization stream alternated in duration. One possible explanation for this asymmetry is related to the syntactic difference between SVO and SOV languages. It is widely accepted that SVO is syntactically unmarked and that all languages are at an underlying syntactic level SVO. The surface SOV order would then be result of movement (64, 65). External confirmation for this theory comes from historical change: change in word order is unidirectional from SOV to SVO (66, 67). It is therefore possible that infants are sensitive early on only to the prosodic profile that indicates that the order of their language of exposure is *not* the unmarked one.

Taken together, there is strong evidence from language acquisition that the perception of linguistic rhythm is one of the first aspects of language young infants perceive and process. Because rhythm is, at different levels of the rhythmic hierarchy, carried by the core elements of the speech signal, it is helpful for a variety of tasks that the young

language learners face: from speech segmentation to learning one of the basic properties of the syntactic structure of the language to be acquired.

6. Sensitivity to speech sounds by cochlear implanted infants

While studies on language acquisition zoom in on the factors that influence speech perception and processing, it is often difficult to disentangle which abilities require early exposure to linguistic input. As newborns are able to discriminate all the sounds found in the world languages, listening to speech in early infancy would appear to primarily help to narrow their cognitive abilities and tune into their mother tongue. However, because native language acquisition is constrained by critical periods – or the window of opportunity – during which native competence of a language can be acquired (68), it is also likely that this narrowing effect of speech input is sensitive to the specific stage of language acquisition. One source of valuable information about how listening to speech modulates our cognitive abilities are individuals who are born deaf and gain hearing through cochlear implants.

Speech intelligibility in cochlear-implanted individuals is directly related to the quality of the prosody they hear (69). There is some evidence that suggests that cochlear-implanted children perform significantly worse than normal controls in the perception of amplitude, pitch and temporal structure of spoken language (70). Studies in languages such as Chinese and Thai, where the lexical meaning of words is modulated by tones, show that the main problem for cochlear-implanted children lies in the perception of pitch (71). This is also evident from studies that show that cochlear-implanted children have also problems in interpreting intonation and its relation with the illocutionary force of a sentence – e.g. distinguishing a statement from a question. As a consequence they also have a significantly poorer performance than normally hearing individuals in the perception and production of emotional prosody (72).

It remains an open question whether prosodic processing can only emerge if the speech input is present during the window of opportunity – the critical period during which language can be acquired, or at least can be acquired natively. However, because rhythm interprets the grammatical structure of sentences at different levels of the prosodic hierarchy, the inability to process prosody may cause serious difficulties in acquiring language in individuals who receive a cochlear-implant after a certain age. On the one hand, the research described above may be useful to diagnose hearing impairments in infants and young children. On the other hand, it may help to find ways to train the perception and production of prosody in cochlear-implanted individuals – both children and adults. One possibility, albeit an unexplored one, may be to complement auditory training with training in the perception and production of rhythm in a non-acoustic modality, e.g. in the visual modality. For example, it has been shown that the iambic-trochaic law determines grouping of sequences also in vision, where adult participants can group sequences of shapes according to visual frequency, intensity and duration just like in the auditory domain (73).

8. Conclusions

From the very beginning of life, humans are sensitive to all consonantal and vocalic distinctions present in the different languages of the world. Their sensitivity to prosody – rhythm and intonation – appears to emerge already in the last weeks of gestation. Knowledge of vowels and consonants, that carry the basic level of linguistic rhythm, is necessary to learn the lexicon. There is evidence that consonants are privileged in learning words because of their relative stability. In contrast, vowels, because of their variability, provide more information about the grammatical structure of languages. Specifically, the percentage of time vowels occupy in the speech stream can give a cue to the size of the syllabic repertoire of the language of exposure and thus to the average length of common words. And, because vowels are the main carrier of prosody they also give a cue to one of the basic syntactic properties of language: word order. Cochlear-implanted children have been shown to be especially impaired in prosody in general and in the perception – and thus the production – of tonal distinctions, in particular. This may suggest that the prosodic processing of speech, especially of intonation, is a precarious ability that requires experience with speech from the earliest stages of development.

Acknowledgements

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Program (*FP7/2007-2013*) / ERC grant agreement n° 269502 (*PASCAL*).

References:

1. Nespors M, Vogel I. *Prosodic Phonology*. Dordrecht: Foris; 1986.
2. Nespors M, Vogel I. On the rhythm parameter in phonology. In: Roca I, editor. *Logical issues in language acquisition*. Dordrecht: Foris; 1990. p. 1-19.
3. Ramus F, Nespors M, Mehler J. Correlates of linguistic rhythm in the speech signal. *Cognition* 1999;73:265-292.
4. Nespors M, Shukla M, Mehler J. Stress-timed vs. syllable timed languages. In: van Oostendorp M, Ewen CJ, Hume E, Rice K, editors. *The Blackwell Companion to Phonology*. Wiley-Blackwell; 2011. p. 1147-1159.
5. Nespors M, Guasti MT, Christophe A. Selecting word order: the Rhythmic Activation Principle. In: Kleinhenz U, editor. *Interfaces in Phonology*. Berlin: Akademie Verlag; 1996. p. 1-26.
6. Nespors M, Shukla M, van de Vijver R, Avesani C, Schraudolph H, Donati C. Different phrasal prominence realization in VO and OV languages. *Lingue e Linguaggio* 2008;7:1-28.

7. Siqueland ER, DeLucia CA. Visual reinforcement of nonnutritive sucking in human infants. *Science* 1969;165:1144-1146.
8. Eimas PD, Siqueland ER, Jusczyk P, Vigorito J. Speech perception in infants. *Science* 1971;171:303-306.
9. Gottlieb G. Krasnegor NA. *Measurement of Audition and Vision in the First Year of Postnatal Life: A Methodological Overview*. Connecticut: Ablex; 1985.
10. Peña M, Maki A, Kovačić D, Dehaene-Lambertz G, Koizumi H, Bouquet F, Mehler J. Sounds and Silence: An optical topography study of language recognition at birth. *Proceedings of the National Academy of Science* 2003;100:11702-11705.
11. Benavides-Varela S, Hochmann JR, Macagno F, Nespor M, Mehler J. Newborns' brain activity signals the origin of word memories. *Proceedings of the National Academy of Sciences of the USA* 2012;109:17908-17913.
12. Gervain J, Macagno F, Cogoi S, Peña M, Mehler J. The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences of the USA* 2008;105:14222-14227.
13. Gervain J, Mehler J, Werker JF, Nelson CA, Csibra G, Lloyd-Fox S, Shukla M, Aslin RN. Near-infrared spectroscopy: A report from the McDonnell infant methodology consortium. *Developmental Cognitive Neuroscience* 2011;1:22-46.
14. Peña M, Pittaluga E, Mehler J. Language acquisition in premature and full-term infants. *Proceedings of the National Academy of Sciences of the USA* 2010;107:3823-3828.
15. Mahmoudzadeh M, Dehaene-Lambertz G, Fournier M, Kongolo G, Goudjil S, Dubois J, Grebe R, Wallois F. Syllabic discrimination in premature human infants prior to complete formation of cortical layers. *Proceedings of the National Academy of Sciences of the USA* 2013;110:4431-4432.
16. Lecanuet JP, Granier-Deferre C, Jacquet AY, Capponi I, Ledru L. Prenatal discrimination of a male and female voice uttering the same sentence. *Early Development and Parenting* 1993;2:217-228.
17. Kisilevsky B, Hains S, Brown C, Lee C, Cowperthwaite B, Stutzman S, Swansburg Lee K, Xie X, Huang H, Ye HH, Zhang K, Wang A. Fetal sensitivity to properties of maternal speech and language. *Infant Behavior and Development* 2009;32:59-71.
18. Mehler J, Jusczyk P, Lambertz G, Halsted N, Bertoncini J, Amiel-Tison C. A Precursor of Language Acquisition in Young Infants. *Cognition* 1988;29:143-178.

19. Moon C, Panneton-Cooper R, Fifer WP. Two-day-olds prefer their native language. *Infant Behavior and Development* 1993;16:495-500.
20. Nazzi T, Bertoncini J, Mehler J. Language Discrimination by Newborns: Toward an Understanding of the Role of Rhythm. *Journal of Experimental Psychology: Human Perception and Performance* 1998;24:1-11.
21. DeCasper AJ, Fifer WP. Of human bonding: Newborns prefer their mothers' voices. *Science* 1980;208:1174-1176.
22. Kisilevsky B, Hains S, Lee K, Xie X, Huang H, Zhang K, Wang Z. Effects of experience on fetal voice recognition. *Psychobiological Science* 2003;14:220-224.
23. Eimas PD, Miller JL. Contextual effects in infant speech perception. *Science* 1980;209:1140-1141.
24. Dehaene-Lambertz G, Peña M. Electrophysiological evidence for automatic phonetic processing in neonates. *NeuroReport* 2001;12:3155-3158.
25. Sansavini A, Bertoncini J, Giovanelli G. Newborns discriminate the rhythm of multisyllabic stressed word. *Developmental Psychology* 1997;33:3-11.
26. Cutler A, Mehler J, Norris D, Segui J. The Syllable's Differing Role in the Segmentation of French and English. *Journal of Memory and Language* 1986;25:385-400.
27. Cutler A, Mehler J, Norris D, Segui J. The Monolingual Nature of Speech Segmentation by Bilinguals. *Cognitive Psychology* 1992;24:381-410.
28. Dupoux E, Kakehi K, Hirose Y, Pallier C, Mehler J. Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 1999;25:1568-1578.
29. Kuhl PK. Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics* 1991;50:93-107.
30. Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 1992;255:606-608.
31. Werker JF, Tees RC. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 1984;7:49-63.
32. Mehler J, Dupoux E. *Naître Humain*. Paris: Odile Jacob; 1990.

33. Mampe B, Friederici A, Christophe A, Wermke K. Newborns' cry melody is shaped by their native language. *Current Biology* 2009;19:1994-1997.
34. Lloyd James A. *Speech Signals in Telephony*. London; 1940.
35. Pike KL. *The Intonation of American English*. Ann Arbor: University of Michigan Press; 1945.
36. Abercrombie D. *Elements of General Phonetics*. Chicago: Aldine; 1967.
37. Ladefoged P. *A Course in Phonetics*. New York: Harcourt Brace Jovanovich; 1975.
38. Borzone de Manrique AM, Signorini A. Segmental durations and the rhythm in Spanish. *Journal of Phonetics* 1983;11:117-128.
39. Wenk B, Wiolland F. Is French really syllable-timed? *Journal of Phonetics* 1982;10:193-216.
40. Bertinetto PM. *Strutture prosodiche dell'italiano. Accento, quantità, sillaba, giuntura, fondamenti metrici*. Firenze: Accademia della Crusca; 1981.
41. Dauer R. *Phonetic and phonological components of language rhythm*. Paper presented at the XIth International Congress of Phonetic Sciences, Tallinn; 1987.
42. Christophe A, Morton J. Is Dutch native English? Linguistic analysis by 2-month-olds. *Developmental Science* 1998;1:215-219.
43. Bosch L, Sebastian-Galles N. Evidence of Early Language Discrimination Abilities in Infants From Bilingual Environments. *Infancy* 2001;2:29-49.
44. Nespors M, Peña M, Mehler J. On the different Roles of Vowels and Consonants in Speech Processing and Language Acquisition. *Lingue & Linguaggio* 2003;2:203-231.
45. Bonatti L, Peña M, Nespors M, Mehler J. Linguistic Constraints on Statistical Computations: The Role of Consonants and Vowels in Continuous Speech Processing. *Psychological Science* 2005;16:451 - 459.
46. Toro JM, Bonatti LL, Nespors M, Mehler J. Finding words and rules in a speech stream: functional differences between vowels and consonants. *Psychological Science* 2008;19:137-144.
47. Toro JM, Shukla M, Nespors M, Endress A. The quest for generalizations over consonants: Asymmetries between consonants and vowels are not the by-product of acoustic differences. *Perception and Psychophysics* 2008;70:1515-1525.

48. New B, Nazzi T. The time course of consonant and vowel processing during word recognition. *Language and Cognitive Processes* 2012;iFirst:1-15.
49. Hochmann JR, Benavides S, Nespors M, Mehler J. Consonants and Vowels: different roles in early language acquisition. *Developmental Science* 2011;14:1445-1458.
50. Selkirk E. *Phonology and syntax: The relation between sound and structure*. Cambridge: The MIT Press; 1984.
51. Beckman M, Pierrehumbert J. Intonational structure in Japanese and English. *Phonology Yearbook* 1986;3:15–70.
52. Cooper WE, Paccia-Cooper J. *Syntax and Speech*. Cambridge: Harvard University Press; 1980.
53. Klatt DH. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America* 1976;59:1208–1221.
54. Wightman CW, Shattuck-Hufnagel S, Ostendorf M, Price PJ. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 1992;91:1707–1717.
55. Christophe A, Peperkamp S, Pallier C, Block E, Mehler J. Phonological phrase boundaries constrain lexical access: I. Adult data. *Journal of Memory and Language* 2004;51:523-547.
56. Gout A, Christophe A, Morgan J. Phonological phrase boundaries constrain lexical access: II. Infant data. *Journal of Memory and Language* 2004;51:547-567.
57. Peña M, Bonatti L, Nespors M, Mehler J. Signal-Driven Computations in Speech Processing. *Science* 2002;298:604-607.
58. Langus A, Marchetto E, Bion RAH, Nespors M. Can prosody be used to discover hierarchical structure in continuous speech? *Journal of Memory and Language* 2012;66:285-306.
59. Dryer MS. The order of subject, object and verb. In: Haspelmath M, Dryer MS, Gil D, Comrie B, editors. *The World Atlas of Language Structures*. Oxford: Oxford University Press. 2005; p.330–333.
60. Bolton T. Rhythm. *American Journal of Psychology* 1894;6:145–238.
61. Cooper G, Meyer L. *The Rhythmic Structure of Music*. Chicago: University of Chicago Press; 1960.

62. Woodrow H. Time perception. In: Stevens S. editor. *Handbook of Experimental Psychology*, New York: Wiley; 1951; p. 1224–1236.
63. Bion RAH, Benavides-Varela S, Nespor M. Acoustic markers of prominence influence infants' and adults' segmentation of speech sequences. *Language and Speech* 2011;54:123-140.
64. Kayne RS. *The Antisymmetry of Syntax*. Cambridge: MIT Press; 1994.
65. Chomsky N. *Minimalist Program*. Cambridge: MIT Press; 1995.
66. Newmeyer FJ. On the reconstruction of 'Proto-world' word order. In: Knight C, Hurford JR, Studdert-Kennedy M, editors. *The Evolutionary Emergence of Language: Social function and the origins of linguistic form*. Cambridge: Cambridge University Press; 2000.
67. Gell-Mann M, Ruhlen M. The origin and evolution of word order. *Proceedings of the National Academy of Sciences of the USA* 2011;108:17290–17295.
68. Lenneberg EH. *Biological Foundations of Language*. Wiley; 1967.
69. Chin SB, Bergeson TR, Phan J. Speech intelligibility and prosody production in children with cochlear implants. *Journal of Communication Disorders* 2012;45:355-366.
70. Meister H, Tepeli D, Wagner P, Hess W, et al. Experiments on prosody perception with cochlear implants. *HNO* 2007;55:264-70.
71. Ciocca V, Francis AL, Aisha R, Wong L. The perception of Cantonese lexical tones by early-deafened cochlear implantees, *Journal of the Acoustical Society of America* 2002;111:2250–2256.
72. Nakata T, Trehub SE, Kanda Y. Effect of cochlear implants on children's perception and production of speech prosody. *Journal of the Acoustical Society of America* 2012;131:1307-3014.
73. Peña M, Bion RAH, Nespor M. How modality specific is the iambic–trochaic law? Evidence from vision. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 2011;37:1199-1208.