Filozofická fakulta Univerzity Palackého

Katedra anglistiky a amerikanistiky

Research on Processes Related to Speech Segmentation in Infants: A Review

(Bakalářská práce)

2022 **Miroslava Schejbalová**

Filozofická fakulta Univerzity Palackého
Katedra anglistiky a amerikanistiky

Přehled výzkumu procesů souvisejících se segmentací řeči u kojenců

Research on Processes Related to Speech Segmentation in Infants: a Review

(bakalářská práce)

Autor: Miroslava Schejbalová
Studijní obor: Angličtina se zaměřením na komunitní tlumočení a překlad
Vedoucí práce: Mgr. Václav Jonáš Podlipský, Ph.D.
Olomouc 2022

Prohlašuji, že jsem tuto diplomovou práci vypracovala samostatně a uvedla jsem úplný seznam citované a použité literatury.

V Olomouci dne 22. 6. 2022 ……………………………………

Miroslava Schejbalová

I would like to express my gratitude to my supervisor Mgr. Václav Jonáš Podlipský, PhD for his help, time, sharing valuable advice, and for his overall support and guidance throughout the writing of this thesis – and besides the thesis supervision, also for giving me the amazing opportunity to take part in his and his colleagues' ongoing infant speech segmentation research project.

Also, many thanks to my partners and friends who reminded me to stretch and drink water.

Table of contents

Introduction

Speech segmentation, a process used by speakers of a certain language to extract concrete individual words from the continuous speech stream, is one of the important investigated topics in the fields of First Language Acquisition (FLA) and Psycholinguistics (Jusczyk et al. 1993, Aslin et al. 1998, Bavin 2009, Cutler 2012, and others).

The aim of this thesis is to serve as an introductory overview of the recent research trends and future challenges in the field of infant speech segmentation, for those interested either in the topic of speech segmentation or certain related topics of child language development.

Relevant research papers were divided into certain categories according to the most prominent and investigated segmentation cues, and theories related to them: *Prosodic Bootstrapping* (syllable prominence, lexical stress, rhythm, and intonation as segmentation cues), *Allophony* and *Statistical learning* (frequency and transitional probability as segmentation cues). The most used research methods such as *High-Amplitude Sucking (HAS) Procedure* (Eimas et al., 1971) and *Headturn Preference Procedure (HPP)* (Jusczyk et al., 1993) are also explained.

The literature review provides the reader with a brief introduction into the topic of child language acquisition, followed by not fully exhaustive but a representational compilation of research papers put into perspective, from the first breakthrough findings (such as Saffran et al. 1996a, Johnson and Jusczyk 2001, Thiessen et al. 2003) to the most recent studies (such as Frota et al. 2018, Chong 2018, Dal bel et al 2021 or Paillereau et al. 2021a) which can guide the reader into some of the future challenges of the field, possibly providing them with enough primary background information, possible reference, and overall inspiration to start working on their own research proposal regarding one of the recent and/or less investigated ideas and topics in infant speech segmentation presented in this thesis.

1 First language acquisition

1.1 Early language development

In many religions around the world, it is believed that the ability to speak has been given to humans by a divine source. This has led to some of the earliest hypotheses suggesting that if a child was left in the wild, they would not only learn how to speak by themselves, but they would also be able to speak the "original" ancient God-given language (Tomasello, 2008).

While the recent scientific research is certainly no longer examining which language was given to the humanity by Gods, some parallels surprisingly similar to this mindset are still present when it comes to some of the theories concerned with the "nature vs. nurture" problematic of the First Language Acquisition (FLA). Apart from the nativists and the behaviorists, there has also been a significant amount of research based on other theories, such as the cognitive theory which, alongside with the statistical learning approach, ties closely to the research of speech segmentation (Clark, 2016).

1.1.1 The input theory

This theory, also regarded as "behaviorism", rejects the idea of language skills being completely innate. According to Skinner (1957), every aspect of human behavior is of the same significance and can be learned by experience, by absorbing the input information from the surrounding world. This approach supports the idea of children being a "tabula rasa", slowly adapting to the stimuli of the language they are surrounded by—for example, by repeating and imitating the phonemes of a language, attaining only the ones they hear frequently enough, and forgetting the ones that are not reinforced (Skinner, 1957). Clark (2016, p. 2) also argues that "even if children are born with a learning mechanism dedicated to language, the main proposals have focused only on syntactic structure. The rest has to be learnt".

1.1.2 The nativist approach

On the other hand, the innate theory has been heavily opposed by the nativists, such as Chomsky (1965), who, with his "universal grammar" theory, argues that the learning mechanism is universal to all languages. According to Chomsky (1965, p. 46), “(the) work of the past years has provided considerable support for a conception of the language faculty (…) as a schematism that narrowly constrains the form of the grammar attained, rather than a system of generalized inductive and taxonomic procedures for acquiring knowledge. Thus, it seems (…) not unreasonable to approach the study of language as we would the study of some organ of the body.” This can be understood as if the human language, often compared to an organ by Chomsky, was something already present in all of us, rooted in our biology, rather than just a general process of brain development, which is what the cognitive theory is concerned with.

1.1.3 The cognitive theory

This theory is heavily based on cognitive psychology and explores the connection between the stages of cognitive development and the stages of language development (Piaget, 1926). Wyatt (1965) describes several levels of development based on the theory of Piaget which are closely linked to language acquisition, and which offer a framework for further psycholinguistic study of this process. These levels consist of the “psychological level” (relationships form between the speakers, and therefore there is a need to communicate), “linguistic level” (the whole process of selecting sound from speech and connecting a meaning to them while putting them to sequences), “physiological level” (the neurological development of brain capacity as well as the development the motor skills needed for speaking), and “acoustic level” when the perception of sound waves is possible (Wyatt, 1965, p. 16). According to her, all of them are present in the process of FLA and many researchers have indeed described how the different aspects of these levels affect the process of speech segmentation (Saffran, Aslin and Newport 1996a, Jusczyk 1999b, and others).

1.1.4 Stages of First Language Acquisition

Scientist of many approaches agree that learning a language is a process as natural and as crucial to children and is deeply connected to their other cognitive abilities, such as perception, attention, memory, problem solving, and learning from experience (Gottfredson, 1997). But this certainly does not happen overnight, and it is a process consisting of several developmental stages beginning in pregnancy as it has been discovered that newborn babies are able to recognize their mothers’ voice which they can hear the most during their time in the womb as well as their mother's native language which they also recognize and prefer (Gervain, 2018).

According to Clark (2016), infants absorb the language input passively until reaching the age of six to ten months (this depends on the speed of the individual infant's development) when most infants begin to babble which leads to the production of their first words and first word combinations until the children reach their second year. Since then, their language skills grow rapidly as the children take active part in contributing into conversations with others around them (Clark, 2016). This allows them to become fully sufficient users of their native language, and therefore the FLA process is complete.

According to Fromkin et al (2011, p. 351), "language acquisition involves development in various components—the lexicon, phonology, morphology, and syntax, as well as pragmatics. These different modules interact in complex ways to chart an overall course of language development." But before the lexicon of children can expand, before they can start forming syntactical structures and discover pragmatic rules – they need to be able to master one of the stepping stones of the FLA: to recognize the individual words from the continuous speech stream of their native language.

As Clark (2016) puts it, children would not be able to attach meanings the units of speech (individual words or phrases) without a successful recognition of such units. This fact makes the process of speech segmentation extremely significant for any subsequent human linguistic development.

1.2 Infant audition

Without regards to what sounds surround them since birth, whether it is the ancient goat bleating *bekos* or just the mumble of an excited Southern American aunt visiting the hospital, human infants are capable of producing sounds of crying and cooing, which many parents associate with urge of their offsprings to express their needs and emotions (Clark, 2016). But apart from expressing their biological needs, infants are also able to perceive and respond to different properties of language and to react to linguistic input. And, indeed, many experiments involving the use of specially designed pacifiers demonstrate that when there is a variation of stimuli an infant is exposed to, they will increase their sucking rate, as opposed to the infant lowering the sucking rate when exposed to the same stimuli repeatedly (Fromkin et al, 2011).

But to respond to any acoustic stimuli, they need to be able to perceive it. But how clearly are, in fact, infants able to hear and use the auditory systems they are born with? According to Kuhn and Siegler (2006), infants clearly do possess the auditory capacity that allows them to represent critical acoustic features of speech, however, their auditory apparatus is still not fully developed, and therefore what acoustic information they specifically use to make delicate phonetic and speech distinction is still a subject of investigation.

1.2.1 Development of infant hearing dispositions

As mentioned in previous chapter, infants hearing develops fast enough for them to be able to distinguish their mother's voice during pregnancy (Gervain, 2018). According to Kuhn and Siegler (2006), the human fetus usually develops the sense of hearing in the second trimester as the inner ear fully emerges and starts to function, and therefore, like other mammals, the human fetus is then already able to have neural reactions to sound. Older fetuses as well as preterm 28-weeks old infants are able to react to sound with behavioral responses. And it is the inner ear which analyzes sounds into frequency bands, a code which contains the crucial information needed for auditory perception, such as "frequency, periodicity, intensity, temporal fluctuations, location and spectral shape" (Kuhn and Siegler 2006, p. 59).

Human brain is then able to calculate the differences between the ears and the sound spectrum, while segregating the sound source form the other background sounds, known as "auditory scene analysis" (Kuhn and Siegler 2006, p. 70). Apart from that, other cognitive processes, such as attention, motivation and memory highly contribute to the infants' ability to percept the speech stream (Kuhn and Siegler, 2006).

1.2.2 The coding processes of the speech signal

From the physical point of view, when the molecules of air are disturbed, a sound is produced, and this is true to the spoken word as well – pushing the air from the lungs through the vocal folds makes them vibrate which then pushes out molecules of air from the mouth and nose in a wave-like manner (Fromkin et al, 2011). According to her and her colleagues, we can describe the produced sound by the speed in which the air variances occur (and can be visible when transcribed in a form of a sound-wave diagram)

– known as “fundamental frequency” or being perceived by the human ear, as “pitch” (Fromkin et al, 2011, p. 378).

According to Kuhn and Siegler (2006), there are two mechanisms which allow the auditory system to recognize and code frequency, each of them limited by certain physical aspects of the human ear anatomy. Firstly, the hair cells in the basilar membrane of the inner ear vibrate according to the sound frequency, but since the membrane has different levels of stiffness in different areas, those areas can respond only to frequencies. In response, each hair cell then transfers this information in a form of neural representation of sound to the auditory nerves.

1.3 The linguistic input perceived by infants

Another factor which plays a role in infant speech perception is the quality of the input itself. Kuhn and Siegler (2006) also note that until at least the age of 6 months, the sounds perceived by infants do not possess the same quality as they do when perceived by adults, but later their hearing develops enough to identify the sounds with the qualities like the adult-like sharpness and detail. This biological shortcoming can be partially compensated by adult caregivers altering their speech to be better understood by infants: adults alter their speech by changing their pitch, exaggerating intonation, slowing the pace of speech, or pronouncing more precisely – using what is also known as “Motherese” (Clark, 2016).

As Clark (2016) notes, not in every culture, but in many of them, adult speakers tend to change the way they speak to children – this phenomenon of underestimating the language skills of the other person, whether it is based on their age, education, or social status, is quite common, and especially visible in the child-directed speech (CDS).

However, according to Fromkin et al (2011), Motherese itself is not the main driving force behind language development, even though the exaggerated properties of speech present in Motherese may help the child to feel reassured and to get its attention. There is also a possibility that this act of exaggerating important details and making the speech more contrasting, and therefore easily distinguishable from background sounds, directly affects the speech perception process (Kuhn and Siegler, 2006). In their research paper, Abu-Zhaya et al. (2016) even discovered that even tactile cues, such a parent touching the baby as they speak, can be used as an aid to speech segmentation.

These utterances, usually clearly articulated and separated by prominent pauses are therefore more digestible for the infants, as these qualities make these utterances more manageable for analysis and repeating recurring words with consistency and highlighting any new information to reassure the infant, might also positively affect the process of extracting units of information from the speech stream, such as the segmentation of individual words (Clark, 2016).

2 Speech segmentation

2.1 Segmentation cues for adults

Lexical decision experiments have shown that when adults segment speech that is when they recognize words in the continuous stream of speech, one of the strategies they use is based on their lexicon (Clark, 2016). When listening to the speech signal, a search is initiated for possible candidates from an existing mental lexicon of the brain (Fromkin et al., 2011). However, according to Kuhn and Siegler (2006), infants without a lexicon are unable to segment the speech stream in a lexicon-based way. This means that they are faced with a crucial segmentation problem: How exactly should infants locate the units of speech without the prior knowledge of said units (Clark, 2016)?

Adult listeners can experience a similar segmentation problem when they encounter an utterance in an unknown foreign language. The experience adults have with a foreign language, or the lack of it, affects whether they can separate certain words from the speech stream or not (Kuhn and Siegler, 2006). This is possible since adult brains are able to apply the knowledge of previous patterns on the current observed situation, referred to as "top-down processing", as well as to perceive information via senses and later extract patterns from such input, referred to as "bottom-up processing" (Goldstein, 2011).

Infants, with less language experience and cognitive abilities not fully developed, may therefore seemingly rely solely on the "bottom-up processing", however, this may not be the case: researchers such as Mersad and Nazzi (2012) propose that infants are also be able to build on their top-down language perception experience to a certain extent [language perception as such already beginning in the prenatal stages (Gervain, 2018)] and use in combination with bottom-up perception.

2.2 Segmentation cues for infants

An empirical study conducted by Christophe et al. (1994, p. 1570) was concerned with the idea of infants being able to use other perceptional cues than pauses during language acquisition to find word boundaries – and if there are such word boundaries, infants should at least perceive them. To test this hypothesis, Christophe et al. (1994)

exposed 3-day-old French infants to a number of bisyllabic stimuli (either from within words or between words, i.e., "*mati*") in order to find out whether they can discriminate between the stimuli with or without a word boundary.

To measure the infants' reactions to the stimuli, Christophe et al. (1994) used the High-Amplitude Sucking (HAS) Procedure, and the results of the experiment suggested that because infants reacted to the change of the stimuli, they must have been able to detect a difference, possibly a perception cue.

A similar study was later conducted by Myers et al. (1996) who were concerned with the level of sensitivity of infants towards word boundaries in the speech stream, this time using stimuli extracted from English words. In their series of five experiments, they examined infants' sensitivity during the exposure to word units in continuous fluent speech, with a pause added either at a boundary between two words ("Coincident version") or between two syllables within one word ("Noncoincident version"), to test their idea that infants are sensitive to phrasal units (Myers et al., 1996, p. 1).

In the first two experiments, Myers et al. (1996) found out that 11-month-old infants were listening significantly longer to pauses inserted at a boundary between two words rather than within one word, suggesting that English-learning infants prefer markers of phrasal units, while infants of ages 4.5-months and 9-months showcased no preference for either boundary, possibly because of lower sensitivity to word units. After altering the stimuli by removing most of the sources of information (allophonic or phonotactic cues) and leaving only basic prosodic information, 11-month-old infants showed no preference to either of the two types of stimuli which, according to Myers et al. (1996) might indicate that by the age of 11 months, infants are sensitive to word boundaries in fluent speech, but instead of prosodic cues like "accent" (Christophe et al., 1994), this sensitivity might depend on different segmentation cues.

Therefore, in contrast to the initial idea that words are heard in isolation, and "a fragment of speech placed between two known words would be hypothesized as being a new word" (Christophe et al., 1994), new theories based on research of the speech segmentation problem were concerned with the existence of concrete segmentation cues relevant to the process of speech segmentation by infants (Saffran, Aslin and Newport 1996a, Jusczyk 1999b, Mattys 1999, and others).

The first solution to the segmentation problem was the idea of the existence of word-boundary cues related to prosody: lexical stress, vowel harmony, tone phenomena, or other cues that could indicate words, like allophonic variations (see section "4 Allophonic cues"), such as the fact that "aspirated unvoiced stops occur only word-initially in Tamil" (Trubetzkoy, 1939 as cited in Christophe et al., 1994, p. 1570). This idea later developed into the *Prosodic Bootstrapping Hypothesis* (Lust, 2006) (see section "3 Prosodic Bootstrapping").

A different solution than the Prosodic Bootstrapping Hypothesis is the *Statistical Learning Hypothesis* (Lust, 2006) (see section "5 Statistical learning") which states that instead of cues in the speech signal, word boundaries are perceived by the infants based on the frequency of transitional probability of certain speech sounds (Saffran et al., 1996a) This hypothesis is based on previous research, such as Hayes and Clark (1970) who proposed that "if the number of phonemes that can possibly follow a given string of phonemes is low, we probably are inside a word; if on the contrary it is high (each possible phoneme then having a low probability of occurrence), we probably are at a word boundary" (Hayes and Clark, 1970 as cited in Christophe et al., 1994, p.1571).

2.3 Methods used in infant speech segmentation research

2.3.1 Headturn Preference Procedure

The *Headturn Preference Procedure* (used by i.e., Jusczyk et al. 1993, Houston et al. 2004, Höhle et al. 2009) starts in a special soundproof booth, where the parent sits on a chair with their child on their lap, two red lamps on the left and right side and one green one at the center (Cheong and Uehara, 2021, p. 3). When the trial begins, the green light turns of and one of the red lights turn on, blinking until the child looks at it, and then then the stimuli starts to play, until the child loses attention – the looking time at different stimuli is then measured (Cheong and Uehara, 2021, p. 3).

2.3.2 Non-Nutritive High-Amplitude Sucking Procedure

The "*Non-Nutritive Sucking Paradigm*", also known as the "*High-Amplitude Sucking (HAS) Procedure*" was first used by Eimas et al. (1971) in their study where 1- and 4-month-old infants were presented with auditory stimuli and reacted to the stimuli by increasing the rate of their sucking on a pacifier every time they encountered a noticeable change in it (Kuhn and Siegler, 2006). The procedure consists of giving the infant a pacifier with the ability to measure the rate of the infants sucking on it while being exposed to an auditory stimulus sound; when the infant is listening consciously to a certain sound, his or her sucking rate has a high amplitude, after the sound is repeated, the rate of sucking lowers, and "if a significant difference is observed in sucking rate toward the two sound stimuli, the infant is presumed to detect a difference between the two sounds" (Cheong and Uehara, 2021, p. 3) which also allows researchers to measure the reactions of infants younger than 6 months who are unable to successfully turn their heads in the direction of a sound yet.

2.3.3 Intermodal Matching Procedure

In a recently published study, Cheong and Uehara (2021) examined the ability of Japanese infants and toddlers to segment words containing simple morae from the continuous speech stream by using a new procedure: *Intermodal Matching Procedure* (a revised version of a method developed by Kobayashi et al., 2005). According to Cheong and Uehara (2021, p. 3), the methods such as HPP and HAS "could not directly demonstrate the method of speech sound segmentation by infants and toddlers" while their new procedure can directly "indicate the number of units that an infant or toddler perceives within a single word" Cheong and Uehara (2021, p. 7), and could be beneficial for more precise research in the future.

2.3.4 Other methods

When studying speech segmentation, Corpus-based studies are also used to prove related theories – for example, the existence of stressed initial syllables in the English vocabulary was tested by Cutler and Carter (1987) who used a corpus-based analysis of British English lexical words, and Rasmus et al. (1999) analyzed phonetic data from a corpus with the data from eight different languages in order to find whether they differed in rhythm. In fetuses and infants younger than 6 months, procedures such as cardiotocography, electrocardiography, electroencephalography, functional near-infrared spectroscopy, magnetoencephalography, and ultrasonography are also used (Chládková and Paillereau, 2020, p. 18).

3 Prosodic Bootstrapping

According to Fromkin et al. (2011, p. 212), *prosody* describes the qualities of certain suprasegmental features of speech sounds, such as length (duration), pitch and stress which are "over and above the segmental values such as place or manner of articulation." Prosodic features are language-specific: for instance, while in English a lengthened vowel may indicate stress and tonic accent which are properties of units larger than a single segment, in Japanese, Finnish or Czech, a change in the length of a vowel is used to contrast different vowel phonemes (single segments) and it can change the meaning of the entire word (Fromkin et al., 2011, p. 212).

The term *bootstrapping*, according to Höhle (2009, p. 359) describes "the assumption that the child is genetically equipped with a specific program to get the process of language acquisition started." Therefore, the Prosodic Bootstrapping Hypothesis is based on the idea that infants are able to extract this prosodic information from the continuous speech stream and use it to attach grammatical properties or syntactic structure to the speech signal (Lust, 2006, p. 290).

3.1 Syllable prominence as a cue

One of the prosodic features is *stress*: a stressed syllable is usually louder, pronounced with a higher pitch and for a slightly longer duration than unstressed syllables (Fromkin et al, 2011). However, when using the term "stress" in the field of speech segmentation, it is important to distinguish between the language-specific types of syllable prominence. For example, the English language has lexical stress, which means that its users must store mental information about which syllable in each lexical word is stressed, and the phonetic difference between stressed and unstressed syllables is greater (Ladefoged and Johnson, 2011). Both second language learners and infants must therefore memorize the stress placement, because it is movable; unlike in Czech where the stress placement is fixed (Ladefoged and Johnson, 2011) and therefore is no need to lexicalize stress positions and the phonetic realization of stress is less prominent. These differences in syllable prominence might play a role when segmenting speech in languages with different stress systems.

Languages also differ in the phonetic cues to stress, such as duration, intensity – an amount of acoustic energy of a syllable relative to the adjacent syllables, or pitch – how fast the vocal folds vibrate (Ladefoged and Johnson, 2011, p. 245). The relative position of the stress within a word also changes in different languages, with stress being usually on the first syllable in Czech or the penultimate (second to last) syllable in Swahili (Ladefoged and Johnson, 2011, p. 249).

As Cutler and Carter (1987, p. 133) examined in their corpus-based analysis of British English lexical words, the existence of stressed initial syllables in the English vocabulary, which the authors refer to as "Strong", is predominant (up to 85% of lexical words uttered on a day to day conversation basis). This could suggest that these Strong syllables may serve as a segmentation cue to infants, at least in English.

3.1.1 Trochaic vs iambic pattern preference in English

To test the idea of Strong initial syllables being dominant in English in relation to infants' speech segmentation, Jusczyk et al. (1993) examined the preferences of 9-month-old English-learning infants regarding the English stress patterns using a Headturn Preference Procedure (HPP). The results in the research paper of Jusczyk et al. (1993) showed that the infants indeed listened to English words with a Weak-Strong stress pattern (iambic pattern) for a shorter amount of time than to the English words with a Strong-Weak pattern (trochaic pattern), which the infants seemed to prefer, in accordance with the notion that majority of words in English begin with a Strong syllable (Cutler and Carter, 1987).

Albin and Echols (1996), on the other hand, were concerned with the word-final syllables in English. In their research paper, they tested whether word-final syllables are somehow highlighted in infant-directed speech compared to adult-directed speech, and whether even unstressed final syllables may have some degree of perception prominence. The results of the experiment supported previous ideas about infant-directed speech having a slower speech tempo and a wider pitch range, and suggested that not only word initial, but also word final syllables are salient to infants (Albin and Echols, 1996).

Indeed, according to Saffran et al., (1996a, p. 616) "final vowel lengthening serves a cue to syntactic and word boundaries in a number of languages, including English" and because infants are sensitive to rhythmic structures (Christophe at al., 1994), vowel duration could be possibly used as a segmentation cue by infants. Nevertheless, according to Saffran et al. (1996a), English-speaking infants would also have to determine the regularities in lengthening by a distributional analysis.

Jusczyk et al. (1999b, p. 159) conducted an extensive study to also explore how the English-speaking infants can segment bisyllabic words from the continuous speech stream. In one of their studies, Jusczyk et al. (1999b) found out that when 7.5-month-old infants detected Strong or Weak stress pattern of words in the context of sentences, they responded to the words as a whole, not only to the Strong syllables, resulting in errors segmentation errors, such as interpreting a Weak/Strong word followed by a Weak syllable ("guitar is") as a Strong/Weak word ("taris") which could indicate that the preference for a trochaic stress pattern is high in infants (this tendency was later also discovered by Houston et al. [2004] in the process of infants extracting trisyllabic words).

However, 10.5 month old infants seemed to be able to segment Weak/Strong words correctly, suggesting that later in their development, infants learn to rely on multiple sources of boundary information (Jusczyk et al., 1999b), this "multiple-cue approach" to speech segmentation of infants was also supported by Mattys et al. (1999) who explored the way how prosodic and phonotactic cues interact with each other, suggesting that 9-month-old infants are able to rely on both, yet at that age, prosody may be more dominant. Consequently, when Johnson and Seidl (2009b, p. 139) exposed 11-month-old infants to stimuli containing only statistical cues (see section "5 Statistical learning"), the infants preferred words according to statistics, but with conflicting stress cues were added, infants preference for prosody outranked their reliance on statistical cues.

3.1.2 The role of language specific patterns

However, Curtin et al. (2005) were interested in finding out to which degree is the stress pattern of a language able to shape the language development of infants. In their experiments, Curtin et al. (2005) exposed 7- and 9-month-old infants, to an artificial language consisting of a continuous stream of CV (consonant-vowel) syllables with the only cue being stress for 2 minutes. The syllables were cut into continuous stream which consisted of several tri-syllabic sequences with the middle syllable being stressed, so that when the infants were later presented with tri-syllabic stimuli in isolation, either with initial/or final stress, with medial stress or an unfamiliar control item, infants' reaction to them could be measured (Curtin et al., 2005). The result of the experiment by Curtin et al.

(2005) exhibited that English-speaking infants preferred initial stress sequences, and, therefore, were sensitive to lexical stress and biased towards trochaic stress pattern even when lexical stress was the only available cue. But what if infants can alter their segmentation strategies?

Thiessen and Saffran (2007, p. 79) hypothesized whether an exposure to a certain input with a clear relation between acoustic information and word boundaries was enough for infants to learn to use the cue for later word segmentation. To test this idea, Thiessen and Saffran (2007) enhanced a particular phonological pattern (such as iambic pattern, uncommon to English) by pauses and exposed 9-month-old infants to it. As a result, infants were able to successfully segment iambic words after the exposure, showing that they are sensitive to the way stress is distributed across words (Thiessen and Saffran (2007). This could indicate that infants can adapt to stress patterns of different languages very rapidly, even by a short exposure to them, and experiments with infants acquiring other languages than English (or French) may be beneficial for future research in the field of speech segmentation.

Another study, this time with Dutch learning infants, was conducted by Kooijman et al. (2009). In their paper, Kooijman et al. (2009, p. 591) focused their research on the fact that, like in English, Dutch speakers prefer the trochaic stress pattern (even though languages with a predominantly trochaic stress pattern also contain words with a iambic pattern). However, 10-month-old Dutch infants still exhibited a great reliance on strong syllables, even when exposed to stimuli with a Weak-Strong pattern. This result may be because, even though Dutch and English stress patterns are similar, they are not identical, leading to slightly different developmental trajectories of Dutch and English infants (Kooijman et al., 2009).

A similar study, based on the findings of Jusczyk et al. (1999b), was conducted by Höhle et al. (2009) who explored the trochaic bias in infant learners of German and French using the HPP. While German infants showed preference for trochaic stress pattern at 6 months, French 6-month-old infants did not show such significant preference while segmenting speech, even though French infants were able to determine between the two stress patterns: a finding which demonstrated that “the trochaic bias is acquired by 6 months of age, is language specific and can be predicted by the rhythmic properties of the language in acquisition” (Höhle et al., 2009, p. 262). Their results underline the significance of cross-linguistic research of speech segmentation by infant learners of languages with different stress patterns, together with taking language specific child directed speech into account as well, as for example Segal et al. (2009) did with their study of prosodic characteristics of Hebrew child directed speech.

Subsequently, Butler and Frota (2018) investigated 4- to 10-month-old infants learning European Portuguese (EP) and their preference to positional prosodic cues, and their results suggested that, probably due to the mixed prosodic features of EP, such infants find cues at the edge of an utterance the most salient.

In another research paper, Frota et al. (2020, p. 1) decided to test 5-6 month old EP-learning infants' sensitivity to the contrast between trochaic and iambic stress, because "previous studies have examined this ability in languages that are either clearly stress-based (favoring the development of a preference for trochaic stress, like English and German) or syllable-based (favoring the development of no stress preferences, like French, Spanish, and Catalan)", and the EP language contains conflicting rhythmic and stress-related cues, and could challenge some of the notions about infant perception of stress. Indeed, 6-month-old infants exhibited asymmetrical preference stress patterns, while favoring iambic stress the most which is a result suggesting that early acquisition of lexical stress is driven by language-specific phonetic tendencies that affect the frequency of trochaic and iambic stress (Frota et al., 2020, p. 12).

Infant speech processing based on prosody was also investigated by Ragó et al. (2021, p. 1), this time focusing on the lexical status of words in a fixed stress language – Hungarian. In their research paper, Ragó et al. (2021, p. 9) tested 6- to 10-month-old infants by exposure to a frequent word and non-word stimuli with either legal or illegal stress, finding out that the infants exhibited expectation for the typical native Hungarian stress pattern even in non-words, probably because in Hungarian, stress does not modify meaning and is consistent, unlike in English.

3.1.3 Cross-linguistic approach to speech segmentation and recent discoveries

Tyler and Cutler (2009) compared how English, French, and Dutch adult listeners used certain prosodic cues to segment the continuous speech stream, using two experiments with Artificial Language Learning (ALL, a technique of exposing listeners to a continuous speech stream of non-existent but phonologically legal words for a certain period, and later testing their recognition abilities) to explore the role of vowel lengthening and pitch movement. While all three groups of listeners were more successful in the recognition task when the last syllable was lengthened, so, therefore, vowel lengthening may be a language-universal cue Tyler and Cutler (2009). When pitch movement was introduced, English listeners preferred the first syllable, French listeners the last syllable and Dutch listeners, indicating that the place of pitch movement is language specific (Tyler and Cutler, 2009, p. 367). The ALL technique could be beneficial to similar cross-linguistic research of infants which seems to be one of the recent directions of the research concerned with prosodic bootstrapping as a segmentation cue.

To further explore the possibility of lexical stress constraining statistical learning, Hay and Saffran (2012) used ALL to familiarize English-learning infants with stimuli (either an artificial language or a tone sequence) altered in intensity or duration of either the first or second syllable, while also containing statistical cues. In the recognition process, infants discriminated that had higher intensity as initial, and syllables with longer duration as final, and, therefore, their statistical learning proved to be constrained by prosody (Hay and Saffran, 2012, p. 610)

Mugitani et al. (2009) was also interested in the duration of syllables, specifically in the role of vowel duration as a phonetic cue in languages where the length of a vowel is phonemic (it can change a meaning of a word, Japanese) or where the vowel length is not phonemic (English). They found out that the ability to discriminate the change in vowel lengths was the same for 18-month-old English infants and 10-month-old Japanese infants, but at 18 months, the perception of Japanese infants was altered by emerging native phonological rules which are sensitive only to long to short vowel change and not vice versa (Mugitani et al., 2009, p. 236).

Similarly, a recent research study of Paillereau et al. (2021b, p. 1) explored the perceptual sensitivity to vowel quality length in the first year of Czech-learning infants. Regarding phonemic vowel lengthening, in Czech (similar to Japanese and Finnish), the acoustic duration of a vowel (differentiated as "short" and "long" vowels) is phonemically contrasting (Paillereau et al., 2021b, p. 1) and its "vocalic system (consists of) five contrasting vowel qualities specified in terms of height and backness (which) combine with two degrees of length" (Paillereau et al., 2021b, p. 2). In their research, they found out that infants between 4 to 10 months failed to detect changes in vowel quality, while being able to detect changes in vowel duration (Paillereau et al., 2021b, p. 4), and suggested that future research could compare their findings with "a language with similar vowel qualities as Czech but no vowel length contrasts (such as Greek or Spanish)" (Paillereau et al., 2021b, p. 5).

3.2 Rhythm as a cue

Apart from syllable prominence, another important prosodic factor of language is the *rhythm* of the speech stream: the perceived tempo of the alteration between stressed and unstressed syllables (Ladefoged and Johnson, 2011), and the perception of rhythm was hypothesized to drive early discrimination between languages (Gervain, 2018).

Indeed, rhythm is language specific: English, for example, is a "stress-timed" language, and every English utterance is divided into rhythmic feet (a stressed syllable surrounded by a number of unstressed syllables) which tend to have the same duration, even though the stressed syllables in feet are more prominent and the unstressed ones are reduced to keep the rhythm consistent, while in French, a "syllable-timed" language, each syllable tends to have the same duration, regardless of being stressed or unstressed (Fromkin et al. 2011, p. 212). Apart from syllables and rhythmic feet, an existence of a different rhythmic unit – the mora – was supported by the Otake et al. (1993) study of the Japanese language, a "mora-timed language" where "syllables can be composed of either one mora (vowel or consonant-vowel), or two morae (VV, CVV, VC or CVC)" (Bertoncini et al., 1995, p. 312).

In their research paper, Bertoncini et al. (1995) investigated the sensitivity of newborn infants to different rhythmical units of speech used in various languages (syllables vs morae) and their ability to use those units to discriminate multisyllabic words from the linguistic input. Using the HAS procedure Bertoncini et al. (1995) tested 3-day-old French infants by exposing them to a list of Japanese words – as a result, infants were able to discriminate between bisyllabic and trisyllabic words, however, they were unable to distinguish between two words with the same number of syllables while being composed of different number of morae. This suggested that "syllables are particularly salient units during the initial stage of speech processing, irrespective of which language and rhythmical structure is heard" (Bertoncini et al., 1995, p. 311).

Echols et al. (1997, p. 202) later studied the role of rhythmic in the speech segmentation of words by both adult and infant speakers of English and found out that while 7-month-old infants showed no preference for a certain rhythmic unit, both adults and 9-month-old infants distinguished previously a heard trochaic sequence, although the 9-month-olds failed to recognize previously heard trochaic units.

In order to explain how detecting the type of rhythm in the input by infants might correlate with the chosen units of speech (syllables, feet or morae), Rasmus et al. (1999) analyzed phonetic data from a corpus of eight languages of several rhythm classes, and by measuring the "interval(s) located between the onset and the offset of a vowel (or consonant)", referred to by the authors as "vocalic" and "consonantal" intervals (Rasmus et al., 1999, p. 271), they computed that Polish, English and Dutch were indeed possibly stress-times, Japanese mora-timed, and French, Spanish, Italian and Catalan syllable-timed (Rasmus et al., 1999, p. 272-273). However, many irregularities in their research suggested that such classification is not definitive, as many languages may not entirely fit into either category, as for example Portuguese, pointed out by Butler et al. (2018).

Mersad et al. (2010) published a review of studies regarding early word segmentation of American English-learning infants based on rhythmical cues. Based on this review, Mersad et al. (2010, p. 37) suggested that at the very beginning of segmenting speech, "infants rely on the underlying rhythmic unit of their native language (…) independently on the lexical level", until they "start specifying other language-specific word boundary cues (allophonic, phonotactic, …)" – a theory referred to as *Early Rhythmic Segmentation Hypothesis* (Mersad et al. 2010, p. 48). Another useful review related to the studies of rhythmic-based segmentation was published by Bjelica (2012) who compared several approaches related to speech rhythm in English and Serbian.

In a recent research study concerned with the development of infants' sensitivity to native versus non-native rhythm, Paillereau et al. (2021a, p. 1) suggests that rhythmic classes (such as "stress-timed" and "syllable timed") are problematic because they do not include enough data on how infants discriminate languages which do not fall into either rhythmic category**.** The results of their experiment show that while 6-month-old infants prefer nonnative rhythm, 4-month-old infants prefer the native one – suggesting that infants acquire sensitivity to rhythmic cues as early as at 4-months (Paillereau et al. 2021a, p. 17).

This study further pinpoints the importance of future cross-linguistic research including rhythmically unclassified languages (Paillereau et al. 2021a, p. 19), as for example the Czech language.

3.3 Intonation as a cue

Patterns of varying pitch within an utterance represent a prosodic feature known as *intonation* (Ladefoged and Johnson, 2011, p. 25). Pitch can convey a lexical difference in certain languages, called tone languages where words can contrast with each other only in tone to have a different meaning, for example in Mandarin Chinese, Thai, Vietnamese, or several African languages (Fromkin et al., 2011, p. 213). The intonation of a sentence can display information about the speaker's attitude, gender, or age, and it can also coincide with syntactic units (Ladefoged and Johnson, 2011, p. 25). According to Clark (2016), when talking to infants, a wider range, together with more careful pronunciation and slower tempo, is also used in infant-directed speech which leads to catching the attention of the infant as well as making the utterances more catered to the infants' hearing and processing development.

In order to find out whether infant-directed speech helps infants with word segmentation, Thiessen and Saffran (2005) conducted a research study where one group of infants between the ages of 6.5 and 7.5 months was exposed to nonsensical syllable combinations with intonation characteristic for adult-directed speech, while the other group of infants was exposed to the same stimuli but with intonation typical for infant-directed speech: intonation and statistical cues (see section "5 Statistical learning") being the only available cues in the stimuli. In accordance with their proposal, the second group listening to infant-directed-style speech was more successful in the segmentation task, suggesting that "infant-directed speech facilitates word segmentation" (Thiessen and Saffran, 2005, p. 53).

According to Männel and Friederici (2009), another factor that intonation plays a role in is the process of identifying intonational phrase boundaries (prosodic boundaries between basic units of intonation [Bavin, 2009, P.62]) which may be language specific and serve the infants as another salient prosodic cue in speech segmentation, as Männel and Friederici (2009) suggested in their research concerned with German and English infants and their processing of intonational phrase boundaries based in the intonational systems of their respective languages.

Another research concerned with German infants by Zahner et al. (2016, p. 3) tested how much may the 9-month-old infants rely on pitch and metric stress when segmenting speech using HPP, with the results suggesting that only when the pitch of the stressed syllables was high, were they perceived as stressed Zahner et al. (2016), showing that pitch and metric stress are related.

Instead of measuring how intonation helps in speech segmenting, Chong et al. (2018) focused merely on the role of intonation in discrimination between languages by infants. In their paper, The Chong et al. (2018, p. 795) exposed 7-month-old English-learning infants to stimuli containing either American English or German, and, when the natural pitch variation of the stimuli was switched out for a monotone, the infants were unable to distinguish between the two languages, supporting the idea of Männel and Friederici (2009).

4 Allophonic cues

Apart from changing the prominence or the tone of syllables, in natural utterance, the signal of speech sound is also distorted and influenced by the surrounding speech sounds on the acoustic level, by a process known as assimilation (Fromkin et al., 2011). An example of changing certain features of a speech sound would be that, according to phonological rules, the phoneme /p/ in pit [pʰɪt] is aspirated, while the /p/ in *spit* [spɪt] is not: in this example, the [pʰ] and [p] are *allophones* of the same phoneme /p/ (Fromkin et al., 2011, p. 232).

In English, aspiration typically occurs at the beginning of a stressed syllable (Fromkin et al., 2011), and, therefore, *allophony* could be used as a segmentation clue by adults who are aware of the phonotactic rules of their language. But what about infants who are yet in the process of learning phonotactic rules – are they able to detect such rules and use them to segment speech?

To find out whether infants are in fact sensitive to allophonic differences or not, Hohne and Jusczyk (1994) tested the abilities of 2-month-old infants to discriminate allophones of the same phoneme in pairs of similar words (*"nitrates"* vs *"night rates"*). After Hohne and Jusczyk (1994) habituated the infants to the sounds of a certain pair, half of the infants was presented with a member of the same pair, and the other with a member of a different one – and using the HAS procedure, the latter group noticed the difference, even after the stimuli was cross-spliced (to minimalize prosodic differences between the pairs, the stimuli altered so that the only difference was in the allophones). Therefore, the research of Hohne and Jusczyk (1994) suggested that this sensitivity to allophonic cues might help infants to later use allophonic information to segment fluent speech.

Jusczyk et al. (1999a) later conducted a research study to find out whether English-learning infants are able to use this sensitivity to allophonic cues to determine word boundaries and thus segment words from a continuous stream of speech.

In their research paper, Jusczyk et al. (1999a) exposed 9-month-old infants, and later 10.5-month-old infants, to pairs of two-syllable items acoustically similar, different only by certain allophones (*nitrates/night rates* used in the previous research), or to two-syllable items not acoustically related at all (*doctor/hamlet*) in the stream of fluent speech. After being familiarized by Jusczyk et al. (1999a) with the first half of the pair of items and then being presented with the second one, the 9-month-old infants were able to perceive a significant difference in familiarity between *doctor/hamlet* but not between *nitrates/night rates* (even though 2-month-old infants were able to distinguish the differences between *"nitrates"* and *"night rates"* when the stimuli was in isolation [Hohne and Jusczyk, 1994]).

However, by the age of 10.5 months, relaying only on allophonic cues was enough for the infants to identify the difference in familiarity between *nitrates/night rates*. According to Jusczyk et al. (1999a), this could indicate that the sensitivity of 9-month-old infants to allophonic cues may not be strong enough, because being able to identify the cues in the context of fluent speech might be too demanding to the infants' processing abilities.

Therefore, the answer to the previous question would be: yes, infants are sensitive to allophonic cues and, together with prosodic bootstrapping, they might use the cues to segment words from the continuous speech stream. But only when they are approximately older than 9 months; when the infants are younger, the allophonic cues might merely help them while they rely on statistical cues, such as frequency or transitional probability (Jusczyk et al., 1999a).

5 Statistical learning

Statistical learning approaches suggest that, when learning a language, the learner studies the input and uses predictability of certain features based on their tendency to occur together in certain combinations (Bavin, 2009, p. 35). Humans are sensitive to identifying patterns, and, according to Saffran et al. (1996b), they can do so not only in different types of input, such as vision, hearing of music or other senses, but also in language. The central idea of these approaches is that the learning of certain patterns in the process of language acquisition leads to the discovery of reoccurring units of information such as syllables, later developing into segmentation of words (Aslin et al., 1998).

If this process is to be the main tool for infants to acquire their mother tongue, then researchers might ask the following questions: (i) What statistical cues are infants sensitive to and when? (ii) How do they identify them and use them in the learning process? and (iii) What are the limitations of statistical learning? (Bavin, 2009, p. 36)

To describe and categorize the types of statistical cues used by learners of a language, we could use the terms "conditional statistics" and "distributional statistics" (Bavin, 2009, p. 37). According to Bavin (2009), conditional statistics can predict how likely *y* is to occur, given whether *x* occurs as well or not. Conditional statistics are also concerned not only with the frequency of the common occurrence of *x* and *y*, but also with the strength of their relationship.

Bavin (2009, p. 39) then describes distributional statistics as "an additional group of statistics" which reflects "the relative frequency of an event". Similar to his example, if two events *x* and *y* occur together, for example a hundred times, the percentage of each event (*x* eighty times, *y* twenty times) informs us about the variability and tendencies of this pair of events and could be used to describe the distributional probability of *x* or *y* in percent's (x = 80%, y = 20%). And even though both conditional and distributional statistics offer a different point of statistical view, they both may derive from the same mechanisms concerned with frequency (Bavin, 2009, p. 37).

5.1.1 Transitional probability (TP)

The phonotactic transitional probability theories are concerned with children learning to segment speech not necessarily into words (with the idea that frequent chunks of speech sounds might possibly be words), but rather into syllables – which children may be able to analyze more easily and, based on the frequency of possible combinations, to determine the valid combinations within words or across word boundaries (Saffran et al., 1996a). This might seem like a complicated deduction process, however, according to Christophe et al. (1994, p. 1571), "for all we know, infants might very well function like tape recorders".

Speech segmentation may be easier for adult speakers of a language with a lexicon (whether it is a fully developed mental lexicon or a lexicon consisting of sequences of speech sounds without a meaning connected to them which Saffran et al. [1996a] refer to as a "proto lexicon"), because they can refer to it and can estimate the probability of a particular sequence of speech sounds according to the syllable position (Dal Ben et al., 2021). This ability to use phonotactic probability would allow speakers of a certain language to be able to decide which sequences are legal or permitted, and when there is an unlikely or impossible sequence within a syllable (Fromkin et al., 2011, p. 261).

Saffran et al. (1996b) were concerned with this approach and asked the following question: With the words rarely divided by pauses, could determining the distributional cues to word boundaries solve the segmentation problem? In their paper, Saffran et al. (1996b) commented on a previous approach, a theory that children learn new words in isolation without any segmentation process involved, based on words being defined as "minimum free forms" (Bloomfield, 1933 as cited in Saffran et al. 1996b), and argued that words in isolation are often ungrammatical, and even in child-directed speech, the words do rarely occur in isolation which makes the approach based on Bloomfield's definition problematic.

In their paper, Saffran et al. (1996b) used a pair of two syllables forming a word "*baby*" [beɪbi] as an example, and stated that to compute the transitional probability of *y* = [bi], one could divide the frequency of the pair *xy* = [beɪbi] by the frequency of *x* = [beɪ]. As a result, a high transitional probability (TP) would then indicate that the existence of *x* is able to predict *y* strongly and a lower TP would suggest a weaker prediction (Saffran et al., 1996b, p. 610). The syllables in "*baby*", as well as "*bacon*" or "*basic*" do not span a word boundary, and, therefore, Saffran et al. (1996b, p. 610) called them "word-internal" pairs. The same computation would also be possible to use with "word-external" pairs, such as "*bay*"+"*too*", but, according to Saffran et al. (1996b, p. 610), the transitional probabilities of x = [beɪ] and y = [tuː] tend to be lower, yet enough for the listener to notice the difference.

To test this idea, Saffran et al. (1996b) exposed adults for 21 minutes to an artificial language with no pauses between words, and after exposure to the stimuli, one half was tested to choose between the words from the artificial language and "nonwords" (syllables form the language but in different order) and the other half had to choose between the words from the artificial language and "part-words" (the last syllable was different). To ensure that the TPs inside the words would be variable, some syllables occurred more time than others (*bu* four times, *ta* only ones), and the overall results suggested that adult learners were able to segment the artificial words based on their transitional probability (Saffran et al., 1996b, p. 613)

But how could the use of such probability be measured in infants, when they are unable to communicate whether they do or do not recognize a word as clearly as adults who were able to press keys on the keyboard for "word" and "nonwords" in Saffran et al. (1996b)?

To tackle these questions, Saffran et al. (1996a) conducted another study, this time on statistical learning by 8-month-old infants with a following idea: since infants are sensitive to new fresh input (such as an unfamiliar faces or sounds), the newer the input, the longer they pay attention to it – therefore, after having extracted information (based on transitional probabilities only) during the exposure to the auditory material, the infants should have different looking times for the familiar and the novel test stimuli.

They used an artificial language (consisting of 4 made-up words, 3 syllables each), and exposed the 8-month-old infants to it. During the habituation trial of listening to monotone female synthetic speech for 2 minutes, the infants were supposed to use the auditory material (Condition 1: *tupiro, golabu, bidaku, padoti*, Condition 2: *dapiku, tilado, burobi, pagotu*) as a potential learning experience. The transitional probability (TP) of syllables in the nonexistent language within words was 1.0 (*tu* always followed *pi*), and the TP across word boundaries of syllables was 0.33 (*ro* could be followed by the beginning of one of the other three words – either *go*, *bi* or *pa*).

Saffran et al. (1996a) believed that such difference was enough for sensitive listeners to would put boundaries between the words right there, and the results of this research suggested that 8-month-old infants are sensitive to transitional probabilities only after 2 minutes of exposure (Saffran et al., 1996a, p. 621).

However, Aslin et al. (1998, p. 321) found out that in the study by Saffran et al. (1996a), the transitional probability between the syllables and frequency of the syllable appearance were unintentionally confounded. To achieve clearer results, they once again conducted the Saffran et al. (1996a) study, this time controlling that the frequency of items in the stimuli was equal and constant (Aslin et al., 1998, p. 321), with the result suggesting that infants do in fact rely on the different TPs, and not solely on frequency (Aslin et al., 1998, p. 323).

5.1.2 The scope and limitations of statistical learning

Saffran et al. (1999, p. 27) later conducted a study where they aimed to determine, whether are adults and infants able to use their statistical learning abilities to segment other acoustic input than just language, or if the "statistical learning ability is uniquely tied to linguistic materials". Therefore, they replicated the Saffran et al. (1996a) experiment, this time using tone sequences organized into "tone words", with the result suggesting that the same mechanisms of statistical learning could be used to analyze non-linguistic input, such as tones (Saffran et al., 1999, p. 27), which may indicate that statistical learning is not unique to language learning but to other cognitive processes.

While statistical learning mechanisms proved to be useful in segmenting both linguistic and non-linguistic input (Saffran et al., 1999), the previous the research papers discovered that infants can also choose from a variety of cues present in the continuous stream of speech, such as stress (Jusczyk et al., 1993), rhythm (Bertoncini et al., 1995), or allophony (Hohne and Jusczyk (1994). In their research paper, Johnson and Jusczyk (2001, p. 321) investigated the way infants integrate these different cues into their segmentation by replicating the Saffran et al. (1996a) study using the HPP and natural syllables instead of artificially generated stimuli. However, after adding conflicting segmentation cues (transitional probability pitted against stress cues and later cues produced by coarticulation), the 8-month-old infants reversed their preferences and relied more heavily on speech cues than statistics (Johnson and Jusczyk (2001, p. 563)

According to Thiessen and Saffran (2003, p. 707), "the hypothesis that stress is the earliest cue used for word segmentation presents something of a chicken-and-egg problem", and to explore the attention infant pay to conflicting cues through the different ages, they replicated the Johnson and Jusczyk (2001) study with 7- to 9-month-old infants (Thiessen and Saffran, 2003, p. 708). Their results suggest that "statistical segmentation helps infants to acquire the vocabulary necessary to discover the regularity of word stress in English", (Thiessen and Saffran, 2003, p. 715) which might be possible even if the sensitivity to such cues develops simultaneously, because infants can use each of them according to their current needs throughput their development.

5.1.3 Statistical cues in artificial vs natural speech

According to Pelucchi et al. (2009, p. 674), the studies of infant speech acquisition which supported statistical learning mechanisms have relied on artificial languages: stimuli consisting of carefully chosen syllables, simplified and far from the qualities of a real speech (even though, according to Thiessen and Saffran [2003, p. 706], "(t)his stripping has been a necessary step in the effort to establish that infants can segment speech on the basis of statistical cues alone.").

Therefore, according to Pelucchi et al. (2009), the applicability of statistical learning on the course of natural language learning needs to be explored. Consequently, the results of their experiments on 8-month-old English learning infants tracking transitional probabilities between Italian words (four-syllable sequences extracted from natural infant directed speech serving as a stimuli) suggest that infants are sensitive to statistical cues in natural speech and even in a different language (Pelucchi et al. (2009, p.674).

However, Johnson and Tyler (2010, p. 339) conducted a study with different results. They also aimed to explore whether are infants able to use the same transitional probabilities when segmenting natural speech compared to the highly simplified artificial language stimuli. They tested 5.5- and 8-month-old Dutch infants in a procedure where the infants had to segment either "an artificial language containing four words of uniform length (all CVCV) or four words of varying length (two CVCV, two CVCVCV)", with equal transitional probabilities (Johnson and Tyler, 2010, p. 339), and while both groups segmented the uniform words successfully, they failed to segment the words with different length.

This was in contrast with the Pelucchi et al. (2009) experiment where the novel stimuli were not artificial, yet all segments were of the same length, and it is up to debate whether what role this difference played in the segmentation process. Anyhow, Johnson and Tyler (2010, p. 343) therefore argue that “infants’ ability to track transitional probabilities between syllables is much more fragile than earlier studies have suggested.” By such statement, they might not deny the ability of infants’ sensibility to transitional probabilities but are merely pointing out the need for further research using natural speech stimuli. For additional comprehensive compilation on statistical learning in language acquisition, see Romberg and Saffran (2010, p. 906) who published a review where they emphasize the importance of “studying statistical language learning in context: within language, within the infant learner, and within the environment as a whole.”

One of the aspects of the natural learning environment could be the voice of the speaker, as Estes and Lew-Williams (2015) suggest. In their research paper, Estes and Lew-Williams (2015, p. 1) tested the ability of 8- and 10-month-old infants to segment speech using statistical cues even when the continuous linguistic input was produced by eight different female speakers with acoustic variations. When listening to eight alternating voices with high variation, infants were able to segment words, even when a new male voice was added, however, when the stimuli consisted of only two voices, similar to a situation with two parents being around the infant, the infants were not able to segment the words, possibly because the low variation in input may encourage them to rather focus on difference rather than average similarities (Estes and Lew-Williams, 2015, p. 12).

5.1.4 The future of statistics: multiple segmentation cues and bilingualism

Saksida et al. (2017, p. 1) based their research paper on the idea of statistical tendencies in spoken language being high enough for infants without prior language knowledge to recognize them and to segment speech accordingly. They conducted a corpora-based analysis of child directed speech data form nine languages, and found out, that when modeling possible “statistics-based speech segmentations”, different languages required different ideal statistical segmentation strategies, and that infants may rely mostly on other non-statistical cues to segment words successfully (Saksida et al., 2017, p. 1).

To find out which cues are infants able to use right after birth, Fló et al. (2019, p. 1) measured the abilities of newborn infants to segment speech after 3 minutes of continuous stimuli with spectroscopy (“a non-invasive brain imaging technique” [Fló et al., 2019, p. 4]), testing both the use of statistical and prosodic segmentation by violating either type of a boundary in the stimuli. The data collected by Fló et al. (2019, p. 12) during their experiment suggest that, regardless of which cue will become more salient in the future of their development, the neonates were able to use both types of cues at such an early age (2-7 days).

A rather interesting development of findings can be observed in the studies of bilingual infants and their segmentation abilities, as for example a recent study by Antovich and Graf Estes (2020, p. 1), in which they test the notion that infants growing up in a bilingual language environment need to somehow manage to detect the irregularities in segmentation cues, sometimes even contrasting, in the two different languages they are learning at the same time. In their research paper, Antovich and Graf Estes (2020, p. 1), observed the abilities of 16-month-old infants to segment from the dual speech stream, and in accordance with their proposal, the monolingual infants were unsuccessful in this task, while the bilingual infants were able to predict statistical cues in both languages, probably based on their real-life experience with attuning to two languages at once.

Conclusion

This paper aimed to put into perspective various research papers in the field of infant speech segmentation. The structure of this thesis was divided into an introduction to the topic of first language acquisition (concerned with defining the terms and discussing the role of speech segmentation in the FLA process), and an overview of the recent research trends and future challenges in the field of infant speech segmentation, as well as scientific methods used in the studies of infant speech segmentation.

After the introductory part in Section 1, the Section 2 was focused on the difference between how adults and infants perceive segmentation cues. It also discussed the methods used to measure infants' segmentation abilities. While the High-Amplitude Sucking (HAS) Procedure (Eimas et al., 1971) has proved to be useful to measure infants who are younger than 6 months and thus unable to focus on the stimuli by looking as in the Headturn Preference Procedure (HPP) (Jusczyk et al., 1993), since both methods are unable to directly indicate the processes of infants' consciousness, other more precise brain and physiology related methods, such as cardiotocography, electrocardiography, electroencephalography, functional near-infrared spectroscopy, magnetoencephalography, and ultrasonography (Chládková and Paillereau, 2020, p. 18) are now used. A new method to replace HAS and HPP could be the recently introduced Intermodal Matching Procedure (IMP), which can directly "indicate the number of units that an infant or toddler perceives within a single word" (Cheong and Uehara, 2021, p. 7).

The literature compilation itself was then divided into chapters reviewing different approaches to the solution of the speech segmentation problem, based on the cues the research papers have focused on: namely Prosodic Bootstrapping (Section 3, syllable prominence, lexical stress, rhythm, and intonation as segmentation cues), Allophony (Section 4, allophones as segmentation cues) and Statistical learning (Section 5, frequency, and transitional probability as segmentation cues).

In Section 3, after an introduction into the hypotheses concerned with prosodic bootstrapping (Lust, 2006) and the existence of trochaic vs iambic bias (Jusczyk et al. 1993, Jusczyk et al. 1999b), the studies have shown that while this Strong/Weak tendency may be true for English (Cutler and Carter, 1987) and other "stress-timed" languages such as Dutch (Kooijman et al., 2009) or German (Höhle et al., 2009), it may not be the same for “syllable timed” languages, such as French (Tyler and Cutler, 2009), or “more-timed” languages such as Japanese (Mugitani et al., 2009). However, not every language fits into those categories, such as European Portuguese which has mixed prosodic features (Butler and Frota, 2018) or Czech which has (together with Finnish and Japanese) a unique phonemically contrasting vowel duration (Paillereau et al., 2021b). Moreover, as Paillereau et al. (2021b, p. 5) suggested, a further future cross-linguistic research is needed, for example by comparing Czech with “a language with similar vowel qualities (…) but no vowel length contrasts (such as Greek or Spanish)”.

In Section 4 concerned with allophony, the study of Hohne and Jusczyk (1994) suggested that only infants older than 9-months are sensitive to allophonic cues enough to used them as segmentation cues, and even then, the allophonic cues might merely help the infants while they rely on other cues, which they are able to detect earlier (Jusczyk et al., 1999a).

In Section 5, after an introduction into the hypotheses concerned with statistical learning (Saffran et al. 1996a, Aslin et al. 1998), the ideas of how statistical learning is limited (Saffran et al., 1999, Johnson and Jusczyk 2001) were explored. As the studies indicated, “statistical segmentation helps infants to acquire the vocabulary necessary to discover the regularity of word stress in English” (Thiessen and Saffran, 2003, p. 715), but might not be cue infants rely on the most, at least in certain ages. Moreover, as Pelucchi et al. (2009) suggested, the statistical segmentation needs to be tested by using natural speech and not only by the artificial language method. Some of the recent studies of this topic seemed to be mostly focused on the idea of studying statistical cues as a part of a multiple cue framework. Together with the studies of prosodic bootstrapping, there is a tendency to fill the gap in research by exploring the segmentation abilities of bilingual infants.

This paper brought together recent and relevant studies in the field of infant speech segmentation and attempts to be a helpful tool in future research, possibly providing the reader with both theoretical and experiment-based background information and overall inspiration to find ideas for future studies and writing about the topic of infant speech segmentation.

References

Abu-Zhaya, R., A. Seidl, and A. Cristia. 2016. "Multimodal infant-directed communication: how caregivers combine tactile and linguistic cues*". *Journal of Child Language*, 44:1088–1116.

Albin, D. D., and C. H. Echols. 1996. "Stressed and word-final syllables in infant-directed speech." *Infant Behavior & Development*, 19 (4):401–418.

Antovich, D. M., and K. Graf Estes. 2020. "One language or two? Navigating cross-language conflict in statistical word segmentation." *Developmental science*, 23(6)1–13.

Aslin, R., J. Saffran, and E. Newport. 1998. "Computation of Conditional Probability Statistics by 8-Month-Old Infants." *Psychological Science*, 9 (4): 321–324.

Bavin, Edith L. 2009. *The Cambridge Handbook of Child Language*. Cambridge Handbooks in Language and Linguistics. Cambridge: Cambridge University Press.

Bertoncini, J., C. Floccia, T. Nazzi, and J. Mehler. 1995. "Morae and Syllables: Rhythmical Basis of Speech Representations in Neonates." *Language and Speech*, 38:311–329.

Bjelica, M. 2012. *Speech rhythm in English and Serbian: A Critical Study of Traditional and Modern Approaches*. Filozofski Fakultet: Serbia.

Butler, J., and S. Frota. 2018. "Emerging word segmentation abilities in European Portuguese-learning infants: new evidence for the rhythmic unit and the edge factor." *Journal of Child Language*, 45(6):1294–1308.

Cheong, Y. and I. Uehara. 2021. "Segmentation of Rhythmic Units in Word Speech by Japanese Infants and Toddlers." *Frontiers in Psychology*, 12. doi: 10.3389/fpsyg.2021.626662

Chládková K., and N. Paillereau. 2020. "The what and when of universal perception: A review of early speech sound acquisition," *Language Learning* 70(4):1136–1182.

Chomsky, N. 1976. *On the Nature of Language*. Cambridge, MA: MIT Press.

Christophe, A., E. Dupoux, J. Bertoncini, and J. Mehler. 1994. "Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition." *The Journal of the Acoustical Society of America*, 95 (3): 1570–1680.

Clark, E. 2016. *First Language Acquisition*. Cambridge: Cambridge University Press.

Curtin, S., T. H. Mintz, and M. H. Christiansen. 2005. "Stress changes the representational landscape: evidence from word segmentation." *Cognition*, 96 (2005):233–262.

Cutler, A. 2012. *Native listening*. Cambridge, MA: MIT Press.

Cutler, A., and D. M. Carter. 1987. "The predominance of strong initial syllables in the English vocabulary." *Computer Speech & Language*, 2(3-4):133–142.

Dal Ben, R., D. D. H. Souza, and J. F. Hay. 2021. "When statistics collide: The use of transitional and phonotactic probability cues to word boundaries." *Memory & Cognition*, 49(7), 1300–1310.

Echols, C. H. , M. J. Crowhurst, and J. Childers. 1997. "The perception of rhythmic units in speech by infants and adults." *Journal of Memory and Language*, 36, 202–225.

Eimas, P. D., E. R. Siqueland, P. Jusczyk, and J. Vigorito. 1971. "Speech perception in infants." *Science*, 171(3968):303–306.

Estes, K. G., and C. Lew-Williams. 2015. "Listening through voices: Infant statistical word segmentation across multiple speakers." *Developmental psychology*, 51(11):1517–1528.

Fló, A., P. Brusini, F. Macagno, M. Nespor, J. Mehler, and A. Ferry. 2019. "Newborns are sensitive to multiple cues for word segmentation in continuous speech." *Developmental science*, 22(4):1–16.

Fromkin, V., R. Rodman, and N. Hyams, 2011. *An Introduction to Language.* Boston: Wadsworth Pub.

Frota, S., J. Butler, E. Uysal,, C. Severino, and M. Vigário. 2020. "European Portuguese-Learning Infants Look Longer at Iambic Stress: New Data on Language Specificity in Early Stress Perception." *Frontiers in Psychology*, 1890 (11):1–14.

Gervain, J. 2018. "The role of prenatal experience in language development." *Current Opinion in Behavioral Sciences* (21):62–67

Goldstein, E. B. 2011. *Cognitive psychology: Connecting mind, research, and everyday experience.* Australia: Wadsworth Cengage Learning.

Gottfredson, Linda S. 1997. "Why g Matters: The Complexity of Everyday Life." *Intelligenc*e a(1): 79–132.

Hay, J.F., and J. R. Saffran. 2012. "Rhythmic grouping biases constrain infant statistical learning." *Infancy: the official journal of the International Society on Infant Studies*, 17(6):610–641.

Höhle, B. 2009. "Bootstrapping mechanisms in first language acquisition." *Linguistics*, 47(2):359–382.

Höhle, B., R. Bijeljac-Babic, B. Herold, J. Weissenborn, and T. Nazzi. 2009. "Language specific prosodic preferences during the first half year of life: evidence from German and French infants." *Infant behavior & development*, 32(3):262–274.

Hohne, E.A., P. W. Jusczyk. 1994. "Two-month-old infants' sensitivity to allophonic differences." *Perception & Psychophysics,* (56):613–623.

Houston, D. M., L. M. Santelmann, and P. W. Jusczyk. 2004. "English-learning infants' segmentation of trisyllabic words from fluent speech." *Language and Cognitive Processes*, 19 (1):97–136.

Johnson E. K., and A. H. Seidl. 2009. "At eleven months, prosody still outranks statistics." *Developmental Science*, 12 (1):131–141.

Johnson, E. K. and P. W. Jusczyk. 2001. "Word Segmentation by 8-Month-Olds: When Speech Cues Count More than Statistics." *Journal of Memory and Language* 44 (4): 548–567.

Johnson, E. K., and M. D. Tyler. 2010. "Testing the limits of statistical learning for word segmentation." *Developmental science*, 13(2):339–345.

Jusczyk, P. W., A. Cutler, and N. Redanz. 1993. "Preference for the predominants stress patterns of English words." *Child Development*, 64:675–687.

Jusczyk, P. W., E. A. Hohne, and A. Bauman. 1999a. "Infants' Sensitivity to Allophonic Cues for Word Segmentation." *Perception and Psychophysics,* 61 (8): 1465–1476.

Jusczyk, P., D. Houston, and M. Newsome. 1999b. "The Beginnings of Word Segmentation in English-Learning Infants." *Cognitive Psychology,* 39 (3-4):159–207.

Jusczyk, P.W. 1997. *The Discovery of Spoken Language*. Cambridge, MA: MIT Press.

Kobayashi, T., K. Hiraki, and T. Hasegawa. 2005. "Auditory-visual intermodal matching of small numerosities in 6-month-old infants." *Developmental Science*, 8:409–419.

Kooijman, V., P. Hagoort, and A. Cutler. 2009. "Prosodic Structure in Early Word Segmentation: ERP Evidence From Dutch Ten-Month-Olds." *Infancy: the official journal of the International Society on Infant Studies*, 14 (6):591–612.

Kuhn, D. and R. Siegler. 2006. *Handbook of child psychology, Vol. 2: Cognition, Perception and Language*. Hoboken, N.J.: John Wiley & Sons.

Ladefoged, P., and K. Johnson. 2011. *A Course in Phonetics (6th ed.).* Boston, MA: Wadsworth.

Lust, Barbara. 2006. *Child Language: Acquisition and Growth.* Cambridge, United Kingdom: Cambridge University Press.

Männel, C., and A. D. Friederici. 2009. "Pauses and intonational phrasing: ERP studies in 5-month-old German infants and adults." *Journal of cognitive neuroscience*, 21(10):1988–2006.

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

Mersad, K., and T. Nazzi. 2012. "When Mommy Comes to the Rescue of Statistics: Infants Combine Top-Down and Bottom-Up Cues to Segment Speech." *Language Learning and Development*, 8:303–315.

Mersad, K., L. Goyet, and T. Nazzi. 2011. "Cross-linguistic differences in early word form segmentation: a rhythmic-based account." *Journal of Portuguese Linguistics*, 9-2 (2010)/10-1 (2011):37–65.

Mugitani, R., F. Pons, L. Fais, C. Dietrich, J. F. Werker, and S. Amano. 2009. "Perception of vowel length by Japanese- and English-learning infants," *Developmental Psychology,* 45(1):236–247.

Myers, J., P. W. Jusczyk, D. G. Kemler Nelson, J. Charles-Luce, A. L. Woodward, and K. Hirsh-Pasek. 1996. "Infants' Sensitivity to Word Boundaries in Fluent Speech." *Journal of Child Language,* 23 (1): 1–30.

Otake, T., G. Hatano, A. Cutler, and J. Mehler. 1993. "Mora or syllable? Speech segmentation in Japanese." *Journal of Memory and Language*, 32:258–278.

Paillereau, N., V. J. Podlipský, F. Smolík, Š. Šimáčková, and K. Chládková. 2021a. "The development of infants' sensitivity to native versus non-native rhythm." *Infancy*, 26(1):1–26.

Paillereau, N., V. J. Podlipský, Š. Šimáčková, F. Smolík, Z. Oceláková, and K. Chládková. 2021b. "Perceptual sensitivity to vowel quality and vowel length in the first year of life." *Journal of the Acoustical Society of America* 1(2):1–7.

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

Piaget, J. 1926. *The thought and language of the child.* New York: Harcourt, Brace, and Company.

Pinker, S. 1994. *The Language Instinct: The New Science of Language and Mind.* London: Allen Lane, the Penguin Press.

Ragó, A., Z. Varga, L. Garami, F. Honbolygó, and V. Csépe. 2021. "The effect of lexical status on prosodic processing in infants learning a fixed stress language." *Psychophysiology*, 58(12):1–13.

Ramus, F., M. Nespor, and J. Mehler. 1999. "Correlates of linguistic rhythm in the speech signal," *Cognition* 73(3):265–292.

Romberg, A. R., and J. R. Saffran. 2010. "Statistical learning and language acquisition. Wiley interdisciplinary reviews." *Cognitive science*, 1(6):906–914.

Saffran, J., E. Johnson, R. Aslin, and E. Newport. 1999. "Statistical learning of tone sequences by human infants and adults." *Cognition* 70 (1): 27–52.

Saffran, J., R. Aslin, and E. Newport. 1996a. "Statistical Learning by 8-Month-Old Infants." *Science* 274 (5294): 1926–1928.

Saffran, J., R. Aslin, and E. Newport. 1996b. "Word Segmentation: The Role of Distributional Cues." *Journal of Memory and Language,* 35 (4): 606-621.

Saffran, J.R. 2002. "Constraints on Statistical Language Learning." *Journal of Memory and Language*, 47:172–196.

Saksida, A., A. Langus, and M. Nespor. 2017. "Co-occurrence statistics as a language-dependent cue for speech segmentation." *Developmental science*, 20(3):1–11.

Segal, O. B. Nir-Sagiv, L. Kishon-Rabin, and D. Ravid. 2009. "Prosodic patterns in Hebrew child-directed speech." *Journal of Child Language*, 36(3):629–656.

Skinner, B. F. 1957. *Century psychology series. Verbal behavior*. East Norwalk, CT, US: Appleton-CenturyCrofts.

Thiessen, E. D., and J. R. Saffran. 2003. “When cues collide: use of stress and statistical cues to word boundaries by 7- to 9-month-old infants.” *Developmental psychology* 39 (4):706–716.

Thiessen, E. D., and J. R. Saffran. 2007. “Learning to Learn: Infants' Acquisition of Stress-Based Strategies for Word Segmentation.” *Language Learning and Development*, 3(1):73–100.

Thiessen, E.D., E. Hill, and J. R. Saffran. 2005. “Infant-Directed Speech Facilitates Word Segmentation”. *Infancy: the official journal of the International Society on Infant Studies*, 7(1): 53–71.

Tomasello, M. 2008. *Origins of human communication*. Cambridge, MA: MIT Press.

Tyler, M. D., and A. Cutler. 2009. “Cross-language differences in cue use for speech segmentation.” The Journal of the Acoustical Society of America, 126 (1):367–376.

Wyatt, G. L. 1969. *Language Learning and Communication Disorders in Children*. New York: The Free Press.

Yule, G. 2010. *The Study of Language*. Cambridge: Cambridge University Press.

Zahner, K., M. Schönhuber, and B. Braun. 2016. “The limits of metrical segmentation: intonation modulates infants' extraction of embedded trochees.” *Journal of Child Language*, 43(6):1338–1364.

Annotation

Author: Miroslava Schejbalová

Field of study: English for Translators and Interpreters

Title: Research on Processes Related to Speech Segmentation in Infants: a Review

Type: Bachelor thesis

Faculty and Department: Faculty of Arts, Department of English and American Studies

Supervisor: Mgr. Václav Jonáš Podlipský, Ph.D.

Number of pages: 50

Number of characters: 88 765

Description: The bachelor thesis presents with a brief introduction into the topic of child language acquisition, followed by not fully exhaustive but a representational compilation of research papers put into perspective, from the first breakthrough findings to the most recent studies, which were divided into categories according to the investigated segmentation cues and theories related to them. The most used research methods such as High-Amplitude Sucking (HAS) Procedure and Headturn Preference Procedure (HPP) are also explained. The purpose of the thesis is to serve as an introductory overview of the recent research trends and future challenges in the field of infant speech segmentation, for those interested either in the topic of speech segmentation.

Keywords: First language acquisition (FLA), behaviorism, nativism, cognitive theory, speech sounds, speech segmentation, phonetics, phonology, intonation, prosody, phonotactic cues, child-directed speech (CDS), Motherese, Headturn Preference Procedure (HPP), Artificial Language Learning (ALL), High-Amplitude Sucking (HAS) Procedure, vowel length, stress, rhythm, intonation

Anotace

Autor: Miroslava Schejbalová

Studijní obor: Angličtina se zaměřením na komunitní tlumočení a překlad

Název: Přehled výzkumu procesů souvisejících se segmentací řeči u kojenců

Typ práce: bakalářská práce

Fakulta a katedra: Filozofická fakulta, Katedra anglistiky a amerikanistiky

Vedoucí práce: Mgr. Václav Jonáš Podlipský, Ph.D.

Počet stran: 50

Počet znaků: 88 765

Charakteristika: Tato bakalářská práce představuje úvod do problematiky osvojování dětské řeči, po němž následuje ne zcela kompletní, avšak reprezentativní přehled výzkumných prací od prvních průlomových zjištění až po nejnovější studie, které byly rozděleny do kategorií podle zkoumaných segmentačních podnětů a teorií s nimi spojených. Vysvětleny jsou rovněž nejpoužívanější výzkumné metody, jako jsou dudlíková metoda (HAS) a metoda podmíněného zrakového vyhledávání (HPP). Cílem práce je sloužit jako přehled současných výzkumných trendů a budoucích výzev v oblasti segmentace dětské řeči, a to pro zájemce jak o problematiku segmentace řeči u dětí.

Klíčová slova: První osvojování jazyka (FLA), behaviorismus, nativismus, kognitivní teorie, zvuky řeči, jazykový vstup, segmentace řeči, fonetika, fonologie, intonace, prozodie, fonotaktické signály, řeč zaměřená na dítě (CDS), mateřština, metoda podmíněného zrakového vyhledávání (HPP), učení pomocí umělého jazyka (ALL), dudlíková metoda (HAS), délka samohlásek, přízvuk, rytmus, intonace