Biographical Sketch for Online Version

Robin Panneton was born in Los Angeles, California; studied at the University of North Carolina, Greensboro, where she earned her MA in 1982, and her PhD in 1985, both under the direction of Dr. Anthony DeCasper. After postdoctoral positions at the University of Rochester (with Dr. Richard Aslin) and the Rose F. Kennedy Center (with Dr. Diane Kurtzberg), she accepted her current faculty position in the Department of Psychology at Virginia Tech. Her scientific interests focus on the properties of both voices and faces that direct and educate infants about various aspects of their native languages. Her research projects involve infants across the first postnatal year after birth.

Megan McIlreavy was born in Charleston, West Virginia; studied at Virginia Tech, Blacksburg, Virginia, where she earned her MS in 2003, and her PhD in 2006, both under the supervision of Dr. Robin Panneton. She is currently a Franklin Postdoctoral Teaching Fellow in the Department of Psychology at the University of Georgia. Her research interests focus on the development of infants’ visual and auditory attention capacities throughout the first postnatal year.
Naureen Bhullar was born in Rajasthan, India; did her undergraduate studies in India. She earned her MS in 2004 and PhD in 2007 at Virginia Tech both under the supervision of Dr. Robin Panneton. Her major scientific interests include infants’ non-native speech perception and the influence of speaker emotion on infants’ lexical processing. She will be an assistant professor in the Department of Psychology at Widener University starting in August, 2007.
Preverbal Development and Speech Perception during Infancy

R Panneton, Virginia Tech, Blacksburg, VA, USA
M McIlreavy, University of Georgia, Athens, GA, USA
N Bhullar, Widener University, Chester, PA, USA

ã 2008 Elsevier Ltd. All rights reserved.

Introduction

During the first year after birth, most humans are continuously embedded in a social milieu within which they acquire information about people and objects, emotions and expectations, causes and consequences. Along with growing social awareness and knowledge, infants begin to understand and produce gestures in their native language system (vocal and manual) that convey their perceptions, feelings, and intentions, as well as those of their social partners. Often, when we think about language acquisition, we focus primarily on word learning. For hearing infants, it involves much more than this, including the perception and production of individual sounds, combinations of sounds, intonations, stress patterns, orderings of units (words, phrases, sentences), and accents. Language also involves movement in the face (e.g., lips and mouths), body posture, and interpersonal routines.

As complicated as this may seem, most toddlers are making requests, naming objects, identifying people, and beckoning attention by the age of 18 months. Without a doubt, infants listen to and learn from the communication that is directed to them. According to an influential model by developmentalist Anne Fernald, vocal communication between caretakers and infants serves development in the preverbal period in three ways: (1) speech captures and maintains infants’ attention to others, (2) speech conveys emotion to the infant, not only about the speaker’s feelings, but also about the speaker’s interpretations of the infant’s feelings, helping to regulate emotional experiences, and (3) speech provides lexical/linguistic information necessary for native language learning. Importantly, all of these functions continuously influence infants and children, although certain milestones within each function may be age-related (e.g., infants may be more sensitive to some aspects of vocal emotion during early as opposed to later infancy).

The aim of this article is to characterize the functions listed above via the existing literature on language-related accomplishments that take place for many infants in their first postnatal year, with an eye on the processes that nurture and shape language development. To do so, we will first briefly consider how developmental psychologists are able to ascertain infants’ understanding of language when they are not yet producing it, and when infants appear to be sensitive to language-relevant
information. Next, we will discuss language learning during infancy from two perspectives: from a social/emotional view that describes the richness of the interactive framework within which almost all early language learning takes place; from a perceptual/biological view that describes the various components of language which infants can and do perceive in the course of learning how to communicate, and how the biology of infants is becoming organized and specialized in ways that promote language processing. Both of these perspectives have generated a considerable amount of research, and contribute in important ways to the ultimate organization and structure of children’s emerging communicative competencies. We will also consider aspects of preverbal development that highlight the importance of ecology in considering just how infants learn language in their everyday environments (outside of ideal laboratory settings) but are not yet well understood, and in need of future research attention.

How Do We Study Language Perception in a Non-Verbal Infant?

It is easy to understand why the topic of language learning in infants has garnered so much attention and inquiry from developmentalists. Almost all infants learn language and are trying to actively communicate with others by the end of the first postnatal year (Figure 1).

But ironically, it is difficult to go about studying language learning when our primary subjects do not yet use it! Developmental researchers have dealt with this by devising interesting techniques that capitalize on several aspects of infants’ rapidly improving perceptuomotor skills. For example, as infants get older (usually by 4 months of age), they become proficient at controlling their own head movement, and can independently (Figure 2) turn to the left and to the right. One popular task used in language perception research takes advantage of this skill by allowing infants to hear certain words when they turn to their right (e.g., native language words) compared to other kinds of words on their left (e.g., foreign language words). In this particular task, we would conclude that infants recognize and prefer their native language if they turned their heads more often to the right (we would also test more infants with these words on their left to make sure there was no turning bias). There are quite a few different kinds of behavioral techniques that are used to ask questions about infant language processing, and capitalize on (Figure 3) different behavioral competencies. However, a full presentation of these methods is beyond the scope of this article. Thus, we summarize the most commonly used behavioral tasks for studying language perception during infancy in Table 1.

In addition to the clever behavioral methods that researchers have developed for the purpose of understanding preverbal development, this field of study has also greatly benefited from advances in a variety of psychophysiological techniques. Efforts to adapt brain-relevant recording procedures for use with infants and young children, such as scalp electroencephalography (EEG), magnetic resonance imaging (MRI), and optical topography (OT), have led to a greater understanding of
neural structure and function (Figure 4) underlying language development. A sample of such neuropsychological methods are summarized in Table 2, and we will integrate findings from this literature with those from behavioral studies as we present some of the major findings from research on preverbal development. In doing so, it is important to acknowledge that the infant brain is not simply a smaller version of that found in the adult. Studies on the organization of adults’ nervous systems clearly show that certain areas of the brain tend to be specialized for language (e.g., left temporal cortex), and that when such areas are compromised through disease or injury, language impairments are highly likely. Infants' brains are quite different in both structure and function, and that we should exercise appropriate caution in drawing parallels between findings with infants and adults. In spite of this caveat, it is important to explore evidence for similar and different neurological/neurophysiological

<table>
<thead>
<tr>
<th>Testing method</th>
<th>Examples of primary tasks</th>
<th>Dependent measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Habituation</td>
<td>Infant is presented with a sound (e.g., the phoneme /ba/) repeatedly and then is tested with the familiar sound and a novel sound (e.g., /pa/).</td>
<td>If infant attends more to novel than to familiar sound.</td>
</tr>
<tr>
<td>Preference</td>
<td>Infant is presented with a visual display accompanied by one sound (e.g., native language) then by the same display accompanied by a second sound (e.g., foreign language); sounds stay on as long as infant looks at the display</td>
<td>If infant looks longer at the visual display when one of the two sounds is available.</td>
</tr>
<tr>
<td>Conditioned head turn</td>
<td>Infant is presented with a continuously repeating background sound (/ba/) which occasionally switches to a novel sound (/pa/); when infant looks in the direction of the sound that changes, gets reinforced</td>
<td>If infant turns head to the side whenever sound changes compared to when there is no change.</td>
</tr>
<tr>
<td>Recognition/ matching</td>
<td>Infant is presented with two side x side visual displays of a speaking head and hears a centrally located sound-track that corresponds to one of the displays</td>
<td>If infant looks longer to the side that corresponds to the sound-track.</td>
</tr>
<tr>
<td>Segmentation</td>
<td>Infant is presented with a stream of acoustic input (e.g., phonemes, words, sentences) and then tested in a head-turn task with two different sound types (familiar on one side; novel on the other).</td>
<td>If the infants looks more the side that produces the familiar (or sometimes novel) stream of sounds.</td>
</tr>
<tr>
<td>Word/object pairing</td>
<td>Infant is repeatedly shown object A paired with word A, and a object B paired with word B. During test the infants sees+hears object A + word A, but also sees+hears a recombination (object A + word B).</td>
<td>If the infant attends longer on the ‘switch’ trial (object A + word B) than on the familiar trial.</td>
</tr>
</tbody>
</table>
Late in prenatal development (at the start of the third trimester), human fetuses can hear (i.e., detect and/or react to sound). Fetuses are exposed to a variety of acoustic information, much of which consists of maternal speech sounds. Prenatally occurring maternal speech can be considered vibroacoustic because it resonates through the mothers’ skeleton; maternal speech also occurs in a higher frequency band than most other in utero sounds, and it is louder at the fetal ear than typical airborne voices as it is amplified by the bone structure of the female body. Prenatal work has shown that fetuses respond to their mothers’ voices with greater heart rate changes (more heart rate decelerations, indicative of increased attention). After birth, newborns prefer their mothers’ voice compared to an unknown female’s voice, perhaps due to their exposure to the maternal voice in utero. Interestingly, newborns do not prefer the voices of their fathers, suggesting that male voices are not as readily available for fetal perception.

A more direct demonstration of fetal learning comes from a study with pregnant women in Paris, who read a nursery rhyme aloud three times a day for 4 weeks toward the end of their pregnancies. Prior to delivery, two nursery rhymes were presented via loudspeakers to the fetuses; the one the mothers had been reciting and a novel rhyme. The fetuses showed significant changes in heart rate to the familiar rhyme but no change in heart rate to the unfamiliar rhyme. Other evidence of prenatal language learning has shown that when pregnant women recite a passage regularly, 2- to 3-day-old newborns modify their sucking patterns in ways that allow them to hear the familiar passage. Newborns prefer recordings of their mothers’ voices filtered in ways that mimic how such voices sound in utero over unfiltered versions that newborns experience after birth. Also, newborns prefer recordings of their native language (even when spoken by unfamiliar females).

Although it is unclear which aspects of maternal language are perceived prenatally, it appears that some information specific to the mother (her vocal signature) and her native tongue (language rhythm) are evident in the prenatal environment, and that this experience biases the newborn toward certain language sounds after birth. In contrast, one recent study with newborns suggests that not all early biasing comes from prenatal experience with mothers’ voices. In this study, newborns showed a preference (Figure 5) for nonsense words (e.g., ‘neem’) over nonspeech sounds that shared every acoustic dimension with the nonsense words, except the spectral properties that allowed the words to sound ‘speech-like’. Similar findings have come from our laboratory with 1-month-olds in which they showed a preference for normal speech over speech that was filtered to preserve its pitch patterning, but obscured the words (this is in

### Table 2  Current psychophysiological methodologies for studying infants’ language perception.

<table>
<thead>
<tr>
<th>Methodology</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>OT (optical topography)</td>
<td>The recording of changes of vascular blood flow from the scalp by measuring the absorption of infrared light in the brain (2–3 cm from the surface); cerebral blood flow increases in areas that are more active depending on the task. Optical fibers are placed on the scalp to both emit and detect changes in blood flow. Can be administered during sleep.</td>
</tr>
<tr>
<td>EEG (electroencephalogram)</td>
<td>The recording of electrical activity from the scalp, which reflects patterns of cortical activation/deactivation caused by synchronous firing of neuron assemblies. Although the temporal resolution of EEG activity is poor, it is a useful measure of the relationship between patterns of brain activity in relation to perceptual/ cognitive function. Must be administered while awake.</td>
</tr>
<tr>
<td>ERP (event-related potential)</td>
<td>The recording of electrical activity from the scalp, similar to EEG except that the change in cortical activation/deactivation is time-locked to the onset of a discrete event. Of primary interest is the latency to significant changes in the amplitude of activation patterns (i.e., waveforms) that reflect brain activity during event processing. Can be administered during sleep.</td>
</tr>
<tr>
<td>MMN (mismatch negativity)</td>
<td>An ERP waveform component that is generated by an unexpected change in some repetitive string of events (e.g., a sudden change in the pitch of successive tones). MMN is evoked in response to changes in the physical features of a sound stimulus such as frequency, location, intensity and duration and also changes in patterns of sound. Can be administered during sleep.</td>
</tr>
<tr>
<td>fMRI (functional MRI)</td>
<td>The recording of changes in cerebral metabolic activity (blood volume shifts and blood oxygenation) in specific areas of cortex as a function of task parameters. The spatial resolution of fMRI (localization of cortical sources) is superior to EEG, but temporal resolution is similar to ERP. Can be administered during sleep.</td>
</tr>
<tr>
<td>HR (heart rate change)</td>
<td>The recording of average length of interbeat intervals (heart period) during event processing. Increases in heart period reflect heart rate deceleration (associated with attention) whereas decreases in heart period reflect heart rate acceleration (associated with alerting or attention termination). Administered while awake.</td>
</tr>
</tbody>
</table>
Language learning after birth takes place in a rich, interactive context in which caretakers, siblings, and others communicate with infants in order to regulate their attention, modulate their emotions and arousal levels, and teach them about themselves and the world around them. Moreover, infants are also surrounded by ongoing communication events between other partners and have numerous opportunities to observe dynamic properties of these exchanges. Thus, the consideration of early language learning is best served by a full description of this social milieu, especially with regard to the emergent communication patterns between infants and others.

Across many of the worlds’ cultures, adults interact vocally with infants via speech and song. Infant-directed speech (IDS) often sounds considerably different from the speech that adults use when interacting with other adults, that is, adult-directed speech (ADS). When comparing the two, IDS is characterized by higher pitch, greater pitch variability, fewer words per utterance, longer pauses between utterances, slower rate, more repetition, and greater hyperarticulation (speech that is clearly enunciated). Generally, these are referred to as ‘prosodic modifications’ in that they primarily influence the way that speech ‘sounds’ to the listener. Adults often modify speech prosody to convey emotion, increase emphasis, and highlight intention, particularly in their speech to infants and children. When addressing infants, such modifications occur in the speech of mothers, fathers, other adults, and even children, and occur within a wide variety of language cultures (French, Italian, German, Japanese, Mandarin, Spanish, Australian, British English, American English, Japanese Sign Language, American Sign Language).

Several studies have shown that infants (especially when young) prefer IDS over ADS, although this depends on certain aspects of the testing context. For example, 1-month-olds show no preference for IDS over ADS when both are spoken by their own mother. However, they do prefer IDS when both are spoken by an unknown female. This further suggests the potency of the maternal voice for infant attention, possibly due to prenatal experience with the mother's voice, as young infants' preference for the maternal voice can override their preference for IDS. In contrast, 4-month-olds prefer IDS over ADS even when both are spoken by their mothers, showing their emerging interest in this enhanced speaking style and perhaps the diminished potency of the maternal voice as infants age. When paired with a smiling female face, female IDS effectively increases infants’ attention, but female ADS does not. As infants get older, IDS has been found to increase attention, aid in emotion regulation, highlight linguistic information, and enable infants to better discriminate speaker gender, speech content, and temporal synchrony of facial information (coordination of lips, voice, and face movements).

Recently, it has been argued that infants’ attention to IDS is primarily due to their perception of its emotional valence, especially for positive vocal emotion or ‘happy talk’. Typically, IDS is emotionally rich, which may help to highlight the communicative intent (Figure 6) of the speaker. In fact, adults more accurately classify speaker intent when listening to IDS compared to ADS utterances. Interestingly, when IDS and ADS are equated
for emotion valence (both speech types are rated by adults as 'happy'), infants show no preference for IDS over ADS. Likewise, when pitch characteristics are held constant, but one group of IDS utterances are judged as 'happy' and another judged as 'sad', infants prefer the happy speech. Therefore, it may be more accurate to say that infants are particularly drawn to speech that conveys a high degree of positive emotion, and that IDS is more likely to do so (at least in some cultures) than ADS. This being said, it is not the case that happy speech is the only speech to which infants respond, as IDS is often employed and effective when calming distressed infants or prohibiting activity in older infants. The vocal components of soothing and prohibitive speech are not the same as those in happy speech, yet infants readily respond to both.

The manner in which adults convey their emotions in speech to infants changes developmentally across the first postnatal year. It has been found that mother's adapt their patterns of interaction (including their voices but also facial and body gestures) according to the age of their infants. More specifically, mothers modify their voices primarily to calm and soothe a newborn and gently introduce them into social interactions. There is also an important social dynamic that is evident in these early communicative patterns not only for the infant but for adults as well. Studies that have compared caregivers' speaking and singing styles both in the presence and absence of their infants have found that mothers exaggerate their prosody more when interacting with their infants. That is, they sing and speak in a higher pitched voice, with a slower tempo, place more pauses between phrases, and have a more loving tone but only when their infant is actually present. Comparatively, the extent of these prosodic modulations declines when mothers sing as if their infant was present.

Although infants and caregivers contribute uniquely to the social situation, most often the caregiver is the more responsible agent for the coordination of the social interaction. Over time, these interactions lay the foundation for infants' understanding of their own emotions as well as the emotional messages being expressed to them. From 3 to 9 months, mothers typically alter their different intonation patterns that help optimize interactions (elicit attention and encourage participation), particularly as this is a time in which infants are able to initiate and contribute as an active partner in the social interchange. Infants appear to differentiate these interactive patterns, even by their vocal characteristics alone. When familiarized with different female voices, all conveying the same intention (e.g., eliciting activity), infants do not increase their attention to a new voice with the same intention, but do increase their attention to a new voice from a different intention category (e.g., calming). Additionally, 5-month-olds have shown differences in facial affect when presented with Approval and Prohibitive IDS (more smiling to Approvals, more negative emotion to Prohibitives). Interestingly, infants did not show any affective differences in response to Approval and Prohibitive ADS. These results suggest that infants can discriminate affective vocal expressions occurring in IDS and that it is more effective than ADS in eliciting infant emotion.

By 9 months, mothers actually attenuate some of the vocal exaggerations associated with positive emotion, replacing them with an increase in their use of directive utterances ("Don't put that in your mouth, sweetie."). As we will see in the next section, it is during this time that infants are becoming more aware of and tuned into the linguistic structure of their language, and thus may not rely as much on exaggerated vocal emotion to maintain their attention to speech. In fact, recent evidence from our laboratory suggests that the preference for IDS over ADS seen in younger infants is somewhat weaker in 8-month-olds. One interesting question that arises from studies investigating infants' speech preferences is whether the vocal emotion in IDS differentially affects infant learning at various ages. In her recent dissertation project, N. Bhullar found that 11- to 13-month-old infants discriminated a change in a target word carried in a set of IDS sentences (that remained constant) when they were presented by a dynamic female speaker who was...
portraying a happy expression. In contrast, another group of same-aged infants did not discriminate the word change under similar conditions except that the female speaker was portraying a sad expression.

In one of the few psychophysiological studies to focus on the perception of emotional prosody in speech, 7-month-old infants’ event-related potentials (ERPs) were compared to happy, neutral, and angry speech. The amplitude of the first negative ERP wave was greatest for angry speech over frontal areas of the infants’ brains (indicating more emotional attention to this event) and the amplitude of the first slow, positive ERP was greatest for happy and angry speech over temporal areas, but not so for neutral speech. This ERP pattern suggests enhanced sensory processing of emotionally loaded linguistic events (happy and angry speech). These psychophysiological data support the hypothesis that infants are primarily responsive to the emotional valence of speech, but they do not address the issue of whether infants prefer one emotional valence to another (e.g., happy over angry speech).

To this point, it is clear that the language ecology of the infant is replete with speech that appears tailored to their interests and needs, even if the primary motivation of adults who talk to infants is to convey their feelings of warmth and nurturance. From this input, infants construct their initial understanding of their native language, so we now turn to those processes that allow infants to decipher the information they will need to become competent communicators with those around them.

Infants’ Perception of the Speech Stream

As is clear from the section above, the input to the infant is a critical aspect of language development. Given the complexities of language, however, it is not immediately obvious how any given language learner, but especially an infant, manages to listen to, comprehend, and then produce meaningful speech. As it turns out, infants can (and do) attend to many different aspects of language, and this perceptual multiplicity is highly adaptive because perception at one level almost always informs the perceiver about language at other levels (in the developmental literature, this is called ‘bootstrapping’).

By way of illustrating this point, we will discuss infants’ perception of prosody (the intonation and rhythm patterns of language) and phonology (the individual speech sounds of language) as each pertains to the acquisition of language-relevant information. Although prosody and phonology continuously interact in natural speech (see Figure 7), it has been fruitful to consider them as separate sources of information for infants’ language learning. Recall that prosody can reside above the level of individual speech sounds (e.g., the emotional tone of utterances), but it can also be linguistically relevant. Prosodic features provide important cues for phrase, clause, and word segmentation, and contribute to languages’ stereotypical sounds.

Phonology often refers to individual speech sounds, such as consonants, vowels, clicks, and tone changes (e.g., in Thai) that are linguistically significant. The world’s languages vary greatly in phonotactic structure, with some languages allowing certain combinations of speech sounds while excluding others (e.g., th-but not dp-at the start of an English word). Phonology also involves the combinations of phonemes into higher-order elements that are lexically meaningful (morphemes, syllables, words). Clearly, infants learning a language must perceive and produce its phonotactic elements, so a considerable amount of research has concentrated on this developmental process.

Perception of Prosody

Because the prosody of IDS is highly exaggerated and infants’ attend more to IDS than to other speaking styles, IDS is likely to be advantageous for its language-promoting properties. One such prosodic source is linguistic rhythm, which varies across the world’s languages. Some languages are stress-timed (their rhythm is determined by differential vocal emphasis on certain syllables) whereas others are syllable-timed (their rhythm is fairly constant because all syllables receive similar emphasis, but some syllables may receive stress primarily through consonant/vowel lengthening) or mora-timed (their rhythm is influenced more by voicing durations, in...
that some syllables contain vowels that are longer than others). Within stress-timed languages (such as English and Dutch), some are dominated by a strong-weak trochaic pattern (as in English words such as ‘feline’ and ‘person’) compared to a weak-strong iambic pattern (such as the English words ‘sustain’ and ‘decide’). Other kinds of prosodic features include intonation contours as they relate to communicative intent (e.g., rising pitch movement toward the end of a sentence to convey a question), lengthening of vowels in final syllables to indicate that a sentence is coming to an end, and periods of silence that indicate separation between successive utterances/words.

As early as 4 days old, newborns discriminate their native from a non-native language, even if the speech is acoustically filtered but leaving the prosodic contours intact. However, no discrimination is seen if newborns hear these filtered contours played in reverse, suggesting that the dynamics of forward pitch movement in speech is one early source of familiarity for infants. Newborns can distinguish between two languages with different stress/rhythm patterns (e.g., Dutch and Japanese) but have considerable difficulty discriminating two languages within the same stress category (e.g., Dutch and English). With more experience and better perceptual skills, 4-month-olds (and older infants) begin to show discrimination between languages within the same rhythm class, such as Spanish and Catalan, and even between two accented versions of the same language (American English and British English). In general, it appears that language stress or rhythm is one of the first ways in which infants begin to categorize languages, with broad rhythmic classes acting as perceptual ‘anchors’ from which they refine their perception of differences across and within their own and other languages.

Several neuropsychological studies have found support for infants’ perception of linguistic prosody, one of the major cues for speech segmentation (e.g., stress patterns, phrases, clauses, words). Five-month-old, but not 4-month-old, German infants showed significant mismatched-negativity (MMN) effects in brain activity while hearing strings of bisyllabic words with their native stress pattern (e.g., trochaic) punctuated by an occasional word with a different stress pattern (e.g., iambic). Although these results show infants’ attunement to the prevalent stress pattern in their native language, the results are slightly different from those found with behavioral studies. As mentioned above, infants show preferences for their native stress patterns in individual words at around 7 months of age, but not earlier. Thus, the discrepancy is in the finding that 5-month-olds discriminated a change in syllable-level stress whereas no preference for dominant native stress pattern has been seen until about 7.5 months of age. This most likely reflects younger infants’ sensitivities to cues that do differentiate words before they are able to perceive that one such pattern predominates in their own native system. This is similar to the finding that younger infants discriminate both native and non-native phoneme contrasts, before they become attuned to those that are present in their own language. However, it would be interesting to couple the use of ERP and behavioral protocols in the same sample of infants to verify that discrimination using one technique does not necessarily predict discrimination in the other, contingent on infant age.

With age and experience, infants’ perception of prosodic stress/rhythm continues to improve, helping infants with a particularly difficult task: segmenting ongoing speech into syllables and words, particularly in the second half of the first postnatal year. Studies have shown that by 6 months, American infants prefer native over non-native words (e.g., English vs. Norwegian), and by 9 months, infants prefer bisyllabic words with the predominant stress patterns of their native language (e.g., trochaic or iambic), even if the words are filtered to reduce phonetic information. Within this same age range, infants are also able to use stress patterns to isolate words from sentences, but more so if the words reflect their language’s dominant stress pattern. For example, American 7-month-olds recognize words like doctor and candle (trochaic) after hearing these in sentences, but they do not recognize words like guitar (iambic), suggesting that infants use the strong syllable (the one with the primary stress) as the start of a word. Of course, as infants get older, they become more adept at extracting words from sentences with any given stress pattern, including weak syllables, but this is a harder perceptual feat and no doubt requires more language experience.

There is evidence that 2-3 day-old-infants can use prosodic cues to help mark the onset of phrase and clause boundaries in continuous speech. These cues are helpful to newborns because they highlight certain units of language and allow for further processing (they can focus on phrase-level vs. sentence-level strings). Likewise, 6-month-olds listen longer to IDS sentences in which natural clauses are bounded by short pauses as opposed to sentences in which pauses have been artificially inserted, suggesting that even silence can act as a perceptual cue to phonological boundaries. Interestingly, this same pattern is seen in older infants with words in that they prefer pauses between words rather than within them, but only when the words are not filtered. This is important because the finding that this preference does not emerge when the word-level information is not available demonstrates that as infants grow older and gain more experience with language, they progressively integrate both prosodic (e.g., pauses) and phonological (e.g., specific consonants and vowels) cues for speech segmentation.

Infants may also integrate across prosodic and phonological levels in their speech preferences given the results of several recent experiments in our laboratory. We have found that American 6-month-olds prefer American...
female IDS over Australian female IDS (even when the same sentences are being spoken), but not if the speech is low-pass filtered (the words are no longer available). In other experiments, we also found that 10-month-olds preferred native utterances (i.e., American) over non-native utterances (e.g., Mandarin), but only within IDS; when using ADS, no preferences were found. It is possible that when listening to ADS, infants did not attend as much to the phonological information in order to recognize utterances as native speech. In sum, the prosodic aspects of language are readily perceived by infants during the entire first postnatal year, and appear to provide important cues to native language structure and to breaking into the speech stream in order to learn about its lexical elements. Next we consider infants’ perception at this more elemental level.

Perception of Phonology

Phonemes are the basic sound units in any given language that have become incorporated into formal language systems. For many of the world’s languages, phonemes consist of various combinations of consonants (C) and vowels (V). For other languages, a phoneme can also be a CV t-tone combination. For example, in Thai, ma (rising pitch) is a different phoneme from ma (falling pitch). Phonemes can be differentiated at many levels, such as (1) their place and/or manner of articulation (e.g., whether the lips are closed or open during production), (2) their voicing properties (e.g., whether activity in the larynx begins prior to full production), and (3) degree of aspiration (or airflow) during production.

From the newborn period onward, infants from all language cultures appear capable of discriminating phonemes (notice a change from one to another), with two features of early phoneme perception being especially noteworthy. First, infants (like adults) perceive phonemes categorically. That is, they discriminate the phonemes /ba/ and /pa/, because they come from two distinct categories (according to an acoustic feature called voice onset time). However, they do not discriminate two versions of ‘pa’ [pa₁ vs pa₂] or two versions of ba [ba₁ vs ba₂] even though acoustically these pairs are just as distinct as the ba/pa contrast, yet they do not cross category boundaries.

The second interesting aspect of phoneme perception is that younger infants respond categorically to speech contrasts that are present in their native language, and also to those that are not present in their native language (i.e., non-native phonemes they have not previously heard). This is true for both consonants and vowels, suggesting that early phonetic perception derives from more general auditory competencies. However, with age and experience, infants continue to discriminate native phonemes, but have more difficulty discriminating non-native speech sounds. This has generally been referred to as perceptual attunement, resulting from infants’ increasing attention to and encoding of native language information.

Interestingly, this pattern of initial perceptual openness followed by progressive narrowing across infancy is seen in other domains. For example, younger infants discriminate between pairs of human faces as well as primate faces, but older infants only maintain discrimination of human faces, even if the primate faces are accompanied by distinct vocalizations. Most recently, infants have even shown discrimination of video presentations of both native and non-native phonemes (with no sound track) at younger ages, but not at older ages. In the older group, only discrimination of native visual phonemes was evident.

Neurophysiological studies support these general behavioral patterns (categorical perception and perceptual attunement). Newborns and slightly older infants show distinct ERPs to categorical changes in consonants, especially those involving voice-onset-time differences. Such category-specific ERPs have been observed over several cortical areas, some involving the right or left hemisphere, and some involving both. Interestingly, distinct ERPs occur in infants when listening to phonemes with place-of-articulation differences but these effects are observed primarily over the left temporal areas (a more adult-like pattern). Discrimination of changes in phoneme categories (both consonants and vowels) has also been observed using MMN measures. For example, newborns show a distinct MMN pattern when presented with two Finnish vowels. MMN has also been observed in English 8-month-olds to the consonant/vowel (CV) pairs /da/ and /ta/. In a similar study, MMN was recorded from Finnish infants at 6 and 12 months of age in a longitudinal design, and from Estonian 12-month-old infants. Both groups of infants were tested for their discrimination of changes in Finnish and in Estonian vowels. The results showed a significant MMN response in the 6-month-olds to both native (Finnish) and non-native (Estonian) vowels, and also in the 12-month-old Estonian infants to their native vowels. However, all infants at 12 months showed diminished MMN to non-native vowels.

Likewise, in a longitudinal ERP study, American infants at 7 and 11 months of age were presented with native and non-native speech contrasts. The results showed no difference in ERP latency or magnitude in speech-related components to either native or non-native contrasts at 7 months of age, but only the native contrasts elicited these same ERP patterns at 11 months of age. This is consistent with the behavioral data reported with non-native speech discrimination. We might expect, then, that the underlying biological substrates that subserve language processing are the same across infants and adults. Infants between 13- and 17-months of age also show larger amplitude ERP responses to known than to unknown words, with
this difference evident in both hemispheres in the frontal, parietal and temporal lobes. By 20 months of age, however, this ERP enhancement is restricted to the left hemisphere over the temporal and parietal lobes only, indicating a gradual specialization of the neural systems for processing words (and much more akin to the pattern seen in adults). These results were further corroborated in a study with 14- and 20-month-olds in which they heard known words, unknown words that were phonetically similar to the known words, and unknown words that were phonetically dissimilar from the known words. Both age groups showed higher amplitude ERP responses to known than to unknown words. However, the 14-month-olds' ERP responses were similar in amplitude for known words and phonetically similar unknown words which imply that these words were confusing. In contrast, ERP responses in the 20-month-olds to the phonetically similar unknown words were the same as those to the unknown words. These findings show that the older infants improved processing of phonetic detail with experience compared to the younger infants.

Thus, young infants show consistent brain-related responses to different speech sounds (supporting the behavioral evidence) but their brain-localization patterns in response to different phonemes appears to depend (at least to some extent) on the nature of the information in the speech sounds that make them distinct. Vocal-timing differences appear to be represented more diffusely in the infant brain whereas place/manner of articulation takes on a more adult-like representation (left-temporal localization). This could be due to less cortical specificity for timing in general (because timing is a process involved in many domains of perceptual functioning) and/or because there are multiple pathways available for speech processing in the developing nervous system, given that the infant is less experienced with speech in general. This latter possibility may help to explain one study which examined infants’ processing of their native language compared to a non-native language and to backwards speech. The results showed that areas of cortex in infants’ brains that are activated by the native language are not completely confined to the primary auditory areas but include those similar to adults in their localization (temporal region) and lateralization (left hemisphere). This early lack of specificity has also been found using ERP methods with 6-month-olds, in which same-component ERPs to words are equally large over temporal and occipital (typically referred to as ‘visual cortex’) brain regions. Interestingly, between 6 and 36 months of age, there is a gradual decrease in the ERP amplitude to vocal words (i.e., decreased processing) over occipital areas but the amplitude remains unchanged over temporal areas.

Such findings point to a common process of perceptual attunement to culturally relevant information throughout the first postnatal year. In terms of language development, infants begin building linguistic representations of phonemes so that their subsequent perception is guided by the fit between an incoming speech sound and these phonemic representations. We have also seen that prosodic information appears to assist infants in this focusing on important elements of speech (prosody bootstraps the discovery of phonemic detail). But what happens if the speech stream that infants’ hear is prosodically attenuated, as is more likely the case when the caretaker seeks to soothe and calm a distressed infant. Such soothing speech is more likely to be lower in pitch, pitch variance, amplitude, and slow. Therefore, are infants not as perceptually attuned to language in these instances? Would infants not learn language if only an ADS style of speaking was available to them?

Perception of Conditional Probability

These questions can be addressed by studies examining another process that appears to facilitate infants’ segmentation of words, called statistical learning. Statistical learning occurs when perceivers are sensitive to the conditional probabilities of the occurrence of adjacent events, over time. When applied to speech, statistical learning refers to the conditional probability that one phoneme will follow another. For example, if the probability of ‘mas’ being followed by ‘cot’ is high (let’s say close to 90%), then a listener might anticipate the sequence ‘mas-cot’. Such sensitivity to conditional probabilities leads to the perceptual grouping of high probability strings (or phoneme clusters), allowing listeners to parse units (e.g., words) from the speech stream relatively rapidly and efficiently. Recently, 8-month-old infants were played continuous strings of speech sounds with no prosodic cues present (e.g., pagotikutapagotikutapagotiku…). In which the conditional probabilities of some phonemes (pagoti) were higher than others. In a subsequent test, infants showed different amounts of attention to ‘words’ such as ‘pagoti’ compared to non-words such as ‘dotapa’ (the conditional probability of this string was very low). So even in the absence of prosodic cues, infants can use other information to help find words in speech. Nonetheless, other recent studies have found that when infants have a choice between prosodic and conditional probability cues for segmentation, they tend to rely on prosody first.

In sum, the speech that is characteristically available during interactions between adults and infants appears to be organized in a way that facilitates language processing. In fact, even adults who are learning words from a non-native language benefit from this same kind of speaking style. These benefits derive from the structure of the input itself, especially that the prosodic and phonotactic characteristics are exaggerated in ways that make speech
information more perceptually available. Overall, the evidence from neurophysiological studies supports those using only behavioral measures of infants' speech processing abilities (e.g., phoneme discrimination; discrimination of linguistic stress). However, the correspondence between adults and infants regarding cortical patterns of localization for speech is less clear. Some level of cortical organization for speech is apparent, but the precise patterning of area-specific increases in activation is dependent on method and/or speech information. Clearly, this is an area of concern for future infant speech research, and will no doubt benefit from continued improvements in technologies as well as infant-specific cortical models.

Relating Research from the Laboratory to Infants' Language-Learning Ecology

To this point, we have considered the nature of speech to infants and how it appears to be ideally suited to maximize attention and highlight information that needs to be culled from the speech stream in order to learn language. We have also seen that to some extent, early language perception is subserved by patterns of cortical organization that promote language processing (e.g., specialization of the left hemisphere for speech). In this last section, we turn to a slightly different issue in considering the context in which language learning takes place, and whether the current literature on infants' perception of speech can address issues of ecological validity. Ideally, language learning takes place in a quiet environment, one in which the speech signal can easily be attended to and identified, processed and stored for later use. In contrast, most infants face more challenging learning situations in their daily lives. Infants acquire language in homes filled with sources of multisensory information (sounds, sights, smells, textures), that can include multiple caretakers and siblings, and an endless list of potential distractors and competitors for attention. In the laboratory, most of this complexity is substantially reduced to make the perceptual task more accessible to the infant, and to increase the likelihood that learning will occur, but this is done at the risk of potentially misrepresenting the natural context for language learning, and so may reduce external validity. We will review the results of studies that have brought some of this natural complexity into the laboratory in interesting ways, and then we will also make suggestions for how future studies can be designed to more accurately model the language learning ecology during infancy.

Given that most natural language learning situations contain more than one sound source at any point in time, infants must be able to separate speech from background distraction, a process known as streaming. Compared to adults, infants appear to be at a disadvantage for streaming because they have: (1) higher auditory detection thresholds (sounds, including speech, need to be louder before they can be detected), (2) higher auditory discrimination thresholds (sounds, including speech, in both quiet and noise, need to be more discrepant before a change from one to another can be detected), and (3) more difficulty localizing sound in space. Also, unlike adults, infants do not listen more selectively to the frequency band within which most speech sounds occur, which means they are attending to other areas of auditory space that are not necessary for perceiving language.

Nonetheless, infants do perceive speech and learn much about their native language in spite of noisy environments and limitations in their perceptual skills. Once again, it may be IDS that helps to facilitate speech perception under demanding conditions. A series of experiments have found that infants can successfully match a voice track with its complementary video of a female speaker when a distracting male voice has been superimposed on her speech, if she is using IDS; if she is using ADS, no such matching occurs. Similarly, infants are able to learn words being repeated by a female IDS speaker even when a male distractor voice is superimposed on hers. However, they do not learn words if these voices are reversed (i.e., the words are delivered by a male ADS speaker with a female IDS distractor voice superimposed on his). Infants show this same kind of enhanced attention when the speaker's voice is someone they know (e.g., their own mother) compared to an unknown female. Thus, the combination of IDS and familiarity appears to help infants focus on both speakers and what is being said.

The ability of familiarity to aide infants' speech perception was shown in a similar study in which infants listened longer to a female speaker when she uttered the infant's own name compared to when she uttered the names of other infants, even when several background female voices were superimposed on the target voice. Importantly, infants' recognition of familiar words in the speech stream (e.g., their own names) also seems to increase the saliency of adjacent words, allowing infants to better process them. It is thought that frequency of sound patterns that are heard repetitively early in the first year are the first steps in building a lexicon. Around 6 months of age, infants respond more to utterances that contain their names than to those containing other infants' names (matched for syllable structure). Capitalizing on this finding, researchers familiarized 6-month-olds with utterances containing the infants own names or others' names, along with target words directly before and after the embedded name. They found that the infants were more likely to recognize isolated target words in a subsequent test when they had occurred adjacent to the infant's own name.

There are many cognitive processes associated with an infant's ability to comprehend speech in noise that are relevant to language learning. Speech stream segregation involves infant discrimination of speech from other sources...
of information (particularly other acoustics), perceptual identification of speech and cues occurring from a specific speaker, and selective attention to that speaker. If there is a deficiency in any of these cognitive processes, an infant's ability to hear speech in noise will be impaired. Further, it has been suggested that if an infant does have difficulty discriminating speech in a noisy environment it may have underpinnings for later language development, as poorer ability to segregate streams of speech could potentially lead to slower language acquisition. This variability in infant performance has yet to be fully explored.

Finally, most existing studies on infants' language learning have employed methods that rely primarily on the presentation of language (whether phonemes, words, or utterances) either in the absence of a face and/or in the presence of an arbitrary visual event (e.g., a checkerboard). However, language learning often takes place in face-to-face interactions between infants and caretakers, and there is often concurrent information about how the sound looks as it is articulated, the timing of the sound with its visual movement in the face, whether or not it is exaggerated in its production (both vocally and facially), and the ways in which speech is complemented by gestures. An important demonstration of the power of face+voice information on infants' perception of consonants is provided by studies involving the McGurk Effect. For example, 4.5-month-old infants were familiarized with a female speaker who was mouthing the phoneme /ga/ but the video was accompanied by her voice saying /ba/. When adults view this video, they perceive a blended phoneme of either /da/ or /tha/. When the infants were tested after familiarization, they only paid attention to the phoneme /ba/, suggesting that they perceived this phoneme as novel even though it is actually the one they heard during familiarization. So, the infants most likely perceived the blended phoneme, like adults, rather than the actual phoneme. This demonstrates the power of the face to influence infants' perception of speech, but this aspect of language learning has largely been ignored in previous research.

Taken together, it is clear that the field of infant speech perception has made great strides toward our understanding of one of the most interesting and important feats in development. With our focus on new and exciting questions, coupled with our continued creativity in designing laboratory tasks that capture the essence of the natural context in which language is learned, this field of research will continue to contribute to our ability to fully characterize language acquisition, and all of its inherent complexities.

See also: Auditory Development and Hearing Disorders (00014); Bilingualism (00021); Grammar (00073); Habituation and Novelty (00074); Language Development (00088); Language Development Theories (00089); Literacy (00094); Perceptual Development (00120); Pragmatic Development (00125); Semantic Development (00141); Speech Perception (00154).

Suggested Readings


