Speech to Infants as Hyperspeech: Knowledge-Driven Processes in Early Word Recognition

Anne Fernald
Department of Psychology, Stanford University, Stanford, Calif., USA

Abstract
The intelligibility of a word in continuous speech depends on the clarity of the word and on linguistic and nonlinguistic contextual information available to the listener. Despite limited knowledge of language and the world, infants in the first 2 years are already beginning to make use of contextual information in processing speech. Adults interacting with infants tend to modify their speech in ways that serve to maximize predictability for the immature listener by highlighting focussed words and using frequent repetition and formulaic utterances. Infant-directed speech is viewed as a form of 'hyperspeech' which facilitates comprehension, not by modifying phonetic properties of individual words but rather by providing contextual support on perceptual levels accessible to infants even in the earliest stages of language learning.

Copyright © 2000 S. Karger AG, Basel

Introduction
Spoken language understanding by adults is influenced not only by the acoustic patterns associated with individual words, but also by how effectively the listener makes use of prosodic, syntactic, semantic, and discourse-level features of the context in which the words occur [see Altmann, 1997]. For infants first learning to recognize and understand spoken words, sophisticated knowledge of language structure is not yet available as a resource in speech processing. However, caretakers speaking to infants may intuitively organize their speech in ways that provide perceptual advantages for the inexperienced listener (e.g., Ferguson, 1977; Fernald, 1992). The idea that infant-directed (ID) speech is a ‘listener-oriented’ speech mode which could help language learners to identify potentially ambiguous words was considered by Lindblom [1990] in the context of his hyper- and hyperspeech (H and H) theory. But Davis and Lindblom (in press) later rejected this proposal as inadequate for explaining how infants solve the problem of recognizing words which are highly variable in their acoustic realization. I argue here that this rejection was too hasty: the H and H perspective provides a valu-
able framework for investigating the contributions of ID speech to the development of speech processing and language understanding.

If every phoneme in the language had a physical definition which was invariant across different talkers and in all phonetic contexts, the task of word recognition would in principle be straightforward. But the same intended speech sound varies considerably when spoken by different people [Johnson and Mullenix, 1997], and a particular phoneme is physically different in different phonetic contexts even when spoken by the same person [see Perkell and Klatt, 1986]. Lindblom [1990] has proposed that experienced listeners can recognize ambiguous words despite extensive phonetic variability because they also make use of many other kinds of information beyond the words themselves. According to H and H theory, a rapidly spoken word which is phonetically underspecified will still be understood by the knowledgeable listener if sufficient 'signal-independent' contextual information is available to enable lexical access. Moreover, speakers collaborate with listeners to provide sufficient information by monitoring the listener's state of knowledge and other aspects of the communicative situation and adjusting their speech output appropriately to maximize intelligibility. Lindblom [1990] predicts that when contextual support is minimal, speakers will tend to compensate by articulating more clearly, using a kind of 'hyperspeech'. Thus the H and H argument goes well beyond the observation that top-down information can facilitate word recognition, by focusing on the speaker's role in enhancing the listener's understanding within the dynamic context of conversation.

When the listener is a linguistically inexperienced infant, however, it is not clear how signal-complementary information could play such an important role. For the child learning language, the extensive phonetic variability in continuous speech would seem to present a problem that could not possibly be offset by access to linguistic information at other levels. In addition to the lack of acoustic invariance in the speech signal, infants also face the challenge of how to identify word and phrase boundaries which are generally not physically demarcated in continuous speech among adults [Fernald and McRoberts, 1996]. The preverbal infant could be seen as the limiting case of a listener without access to higher levels of language structure as a resource in disambiguating spoken words. H and H theory would predict, first, that infants must rely much more heavily than adults on explicit signal information in speech if they cannot yet make use of signal-complementary information. And second, that adults speaking to infants should compensate for these limitations by using a hyperspeech mode, in order to reduce phonetic variability and provide clearer exemplars for the inexperienced listener.

To explore the hypothesis that phonetic variability is lower in ID than in adult-directed (AD) speech, Davis and Lindblom [in press] analyzed vowel formant frequencies in speech to a 6-month-old infant. Contrary to the prediction that mothers articulate more clearly in ID speech, they found substantial phonetic variability in speech to the child. Davis and Lindblom [in press] then concluded that H and H theory is not useful in explaining the variation in ID speech, based on their assumption that preverbal infants could not possibly have access to signal-independent information to help them identify words in the face of extensive phonetic ambiguity. Since it goes without saying that such knowledge is largely still to be developed by infants, phonetic variation in BT [ID speech directed to a 6-month-old child cannot be successfully accounted for by invoking H and H theory' [Davis and Lindblom, in press, p. 15].
However, recent research on infant speech processing suggests that some of the central insights motivating H and H theory are indeed relevant to understanding how ID speech facilitates speech perception and word recognition by infants. In light of new experimental evidence on infants’ attention to structure in speech [e.g. Jusczyk and Aslin, 1995; Saffran et al., 1996], on the one hand, and on the role of ID speech in facilitating word recognition [Fernald et al., in press], on the other, Davis and Lindblom’s [in press] assumption that infants cannot exploit signal-complementary information seems (happily) too pessimistic. It no longer goes without saying that preverbal infants rely exclusively on explicit signal information in recognizing spoken words. In making this case, I take liberties with the original formulation of H and H theory by focusing on only one half of the argument, exploring the notion of listener-oriented hyperspeech in interactions with infants without discussing Lindblom’s [1990] complementary notion of production constraints leading to the use of hypospeech. Moreover, while H and H theory focuses on speaker’s accommodations to listeners at the segmental level, the idea of hyperspeech proposed here is extended to include other kinds of modifications which may enhance intelligibility. Although the new findings from infancy research are still a long way from resolving the issue of how inexperienced listeners cope with variability at the phonetic level, they suggest that even very young infants are adept at discerning structure on many levels as they listen to spoken language, and that ID speech can indeed be seen as an effective form of hyperspeech which serves functions related to those proposed by Lindblom [1990].

**Signal-Complementary Information in Speech Signal Processing by Adults**

The fact that adult listeners do not rely only on acoustic information in the signal to identify words in continuous speech has been documented extensively in experiments exploring the role of contextual factors in spoken language understanding. Early research by Pickett and Pollack [1963a, b] showed that when words are excised from their context, they are often phonetically ambiguous and difficult to understand in isolation. Warren’s [1970] parallel demonstration of the ‘phoneeme restoration effect’ showed that when sufficient context is available, listeners can understand a word even when a segment within the word is replaced by noise, often failing to notice that the segment is missing. What Lindblom [1990] refers to as ‘signal-complementary’ or ‘signal-independent’ information includes such classic examples of contextual cues used to compensate for incomplete phonetic specification. But these terms can also be extended to a broader range of sources of linguistic and nonlinguistic knowledge which the experienced listener can exploit in making judgments and predictions about word identity. The many forms of organization in spoken language which are potentially informative sources of signal-complementary knowledge include the following:

**Coarticulation Effects.** Since the vocal tract changes shape continuously and anticipates what is coming next, a speech sound reflects not only the intended sound but also previous and subsequent sounds in the sequence. Thus the vowel in the word *job* is acoustically different from the vowel in either *sob* or *jog*. By splicing the initial consonant and vowel from *job* to the final consonant from *jog*, Marslen-Wilson and Warren [1994] created a word which was perceived as *jog* but in which there was a mismatch between the place of articulation indicated by the vowel transition and the
place of articulation indicated by the postvocalic consonant release. They found that subjects in a lexical decision task responded more slowly and less accurately to words in which the articulation cues were misleading, indicating that listeners make use of the smallest detail possible to distinguish among words in the mental lexicon.

Phonotactics. Languages differ not only in the sounds they use but also in restrictions on how these sounds can be ordered to form syllables and words. For example, in English the consonant cluster [fr] can occur at the end of a word, as in raft and thief, but not at the beginning of a word, while the opposite pattern holds for the consonant cluster [fr]. Knowledge of such language-specific phonotactic rules can guide the listener in the correct segmentation of words in continuous speech.

Prosodic Regularities. Languages also have characteristic rhythmic properties such as the strong/weak stress pattern characteristic of many words in English [Cutler and Carter, 1987]. Cross-linguistic research shows that listeners make use of such prosodic regularities in identifying word boundaries in continuous speech [see Cutler et al., 1997], and that listeners' segmentation strategies reflect the rhythmic structure of their native language [e.g., Otake et al., 1993].

Lexical Patterns. Syllables that co-occur in a fixed order are likely to constitute words, and listeners can use the first part of a syllable sequence to make accurate predictions about what follows. Research on the time course of lexical access shows that adult listeners process speech incrementally, generating and rejecting hypotheses about the identity of a word based on what they have heard so far [e.g., Marslen-Wilson and Zwitserlood, 1989]. For example, a syllable sequence such as elef activates several English words including elegant, elevator, and elephant, all consistent with the word-initial phonetic information. However, when the listener hears the next segment is /l/, the word elephant can be uniquely identified even before the final syllable is completed. Such distributional regularities in patterns of syllables provide another kind of signal-complementary information which listeners can use in word recognition.

Semantic Relations. Semantic priming can fundamentally shape the perception of a physically ambiguous speech signal. To use Lindblom's [1990] example, the phrase less'n five is interpreted quite differently when preceded by the question What was your homework assignment? than when preceded by the question How many people came today? Although the pronunciation is identical in both cases, less'n five would be perceived unambiguously as lesson five in the first case, and as less than five in the second. This phenomenon is just one of many in the substantial literature in cognitive psychology on concept-driven or top-down processing in both the auditory and visual domains [e.g., Stanovich and West, 1983].

Syntactic Cues. While the English word to could in principle be followed by a word in several classes, in the context of the phrase beginning to, it can only be followed by a verb. Listeners take such grammatical cues into account when interpreting an ambiguous sequence such as /ilstak/, which is perceived differently in different syntactic frames – as a verb in wants to attack or as a determiner plus noun in sit on a tack. Thus knowledge of syntactic rules can also constrain alternatives in word recognition, although context effects due to syntactic congruity are weaker than effects due to semantic congruity [e.g., Tyler and Wessels, 1983].

Discourse Level Cues. The content and coherence of ongoing discourse provides another source of contextual information which can facilitate recognition of words without exclusive reliance on sensory information. Numerous studies have shown how real-world knowledge is brought to bear in language comprehension, as when well-
known scripts are activated and increase the speed of recognition of related words [e.g. Sharkey and Mitchell, 1985].

Nonlinguistic Information. Listeners also integrate information from the visual context with linguistic information in the process of identifying spoken words [e.g. Tanenhaus et al., 1995]. For example, when listeners follow explicit instructions to manipulate real objects, they closely monitor the scene before them. Using an eye-tracking camera, Tanenhaus et al. [1995] showed that listeners’ eye movements to the objects named are time-locked to the referring words, and that visual information is continuously integrated in the course of arriving at a correct interpretation of the utterance.

These few examples make the point that in natural discourse, adult listeners have numerous and diverse sources of signal-complementary information which they can rely on simultaneously in transforming an ambiguous acoustic signal into a meaningful representation.

Do Infants Have Access to Signal-Complementary Information?

Until recently, it was difficult to imagine how preverbal infants who show no signs of understanding the words they hear could possibly be making use of top-down information in speech processing. The pioneering early studies on the development of speech perception tested infants’ ability to discriminate isolated syllables [see Aslin et al., 1998]. There was initially little communication between the researchers trained in psychoacoustics who studied speech processing skills evident in the first months of life and the language acquisition specialists, typically trained in linguistics, who studied the development of comprehension abilities that start to emerge in spontaneous behavior at the end of the 1st year. In more recent research, however, these two traditions have begun to converge on common questions about the early development of receptive language skills. The focus on infants’ perception of isolated syllables has shifted toward explorations of infants’ ability to discern patterns in continuous speech which are relevant to several different aspects of linguistic structure.

The forms of signal-complementary information potentially useful to adults in word recognition draw on knowledge of the phonological, semantic, syntactic, and discourse organization of the language as well as on nonlinguistic knowledge. Because infants are still in the process of learning about the phonology of the ambient language, and have just begun to build a lexicon and to figure out syntactic relations, they have limited resources for disambiguating spoken words using signal-independent knowledge. However, even in the early months, infants’ attention to regularities in the speech they hear provides a foundation for emerging linguistic capabilities which will become more evident in the 2nd year of life. Long before they understand word meanings, infants perceive speech sounds in categories shaped by exposure to the ambient language [e.g. Kuhl et al., 1992; Werker and Tees, 1983]. By 7 months of age, infants are also sensitive to phonotactic patterns typical of the language [e.g. Jusczyk et al., 1994]. That is, they show evidence of recognizing not only which speech sounds are common in the language they are hearing but also which sequences of speech sounds are legitimate. Around this age, infants also attend differentially to language-specific stress patterns. Jusczyk et al. [1993] found that English-learning infants show a listening preference for the strong-weak stress pattern typical of the majority of English words. Other
studies show that infants recognize recurrent patterns in strings of speech sounds that more closely resemble natural speech. Jusczyk and Aslin [1995] found that when 8-month-olds are presented repeatedly with a bisyllabic word embedded in continuous speech, they can later recognize the word as familiar when it is presented in isolation. Even when familiarized only briefly with nonsense syllable strings that do not represent actual language samples, 9-month-olds appear to extract wordlike units by noticing which syllables co-occur [Saffran et al., 1996]. These studies indicate that over the 1st year, infants become increasingly skilled in making detailed distributional analyses of acoustic-phonetic features of spoken language.

Given that the abilities to recognize familiar sequences of speech sounds and to parse the speech stream into wordlike units are emerging over the 1st year, can these be viewed as potential sources of signal-complementary knowledge in the sense intended by Lindblom [1990]? For example, does early awareness of phonotactic regularities or typical stress patterns in English actually help the infant to identify ambiguous words? Research directly addressing this question is just beginning, but Jusczyk et al. [1993] have found that English-learning infants more readily segment embedded words which have the strong-weak stress pattern prevalent in English than words with the less common weak-strong stress pattern [Jusczyk, 1997]. This finding suggests that infants' segmentation strategies become more efficient through early experience, as they are adapted to exploit recurrent patterns which provide cues to word boundaries in the ambient language. It is important to note, however, that the 'word segmentation' skills demonstrated at this age are more appropriately viewed as an increasingly fine-tuned ability to recognize familiar phonetic patterns. Although competence in identifying sequences of sounds as coherent acoustic patterns is obviously prerequisite for comprehension, this can occur without any association between a particular sound sequence and a word meaning. For example, Hallé and de Boysson-Bardies [1994] found that 10-month-old French-learning infants listened longer to words likely to be familiar to them than to less common words, showing that infants can have some kind of acoustic-phonetic representation for frequently heard words before there is any evidence for comprehension. While the 10-month-old infant's ability to identify a spoken word as an exemplar of a familiar sound sequence constitutes word recognition in only a limited sense, it is nevertheless an essential step in the process. Infants' rudimentary awareness of phonological regularities and recurrent lexical patterns functions as signal-complementary knowledge in facilitating segmentation, even at this early stage of language development.

Toward the end of the 1st year infants begin to learn meanings for words, at first very gradually. Around the age of 18 months, many infants pick up speed in word learning, a shift in rate of acquisition which is often referred to as the 'vocabulary burst'. During this period they also pick up speed in another important domain, recognizing familiar spoken words more quickly and reliably. Research in our laboratory has shown that infants make dramatic gains in speech processing efficiency between the ages of 15 and 24 months [Fernald et al., 1998]. We tested infants in a procedure where they looked at pictures of familiar objects while listening to speech naming one of the objects. By closely monitoring their eye movements, we could assess the speed and accuracy of word recognition. When listening to a sentence such as Where is the baby?, 15-month-old infants shifted their gaze to the matching picture only after the target word has been completely spoken. However, 24-month-old infants typically responded about 300 ms faster, shifting their gaze to the correct picture before the end of the tar-
get word. Thus around the time of the vocabulary burst, infants not only increase the rate of learning new words, but also increase the speed and efficiency with which they recognize and understand familiar words in continuous speech.

In two further studies, we asked whether infants are able to process speech incrementally, an ability central to the speed and efficiency of spoken language understanding by adults. Swingley et al. [1999] showed 24-month-old infants pairs of pictures of objects whose names have either substantial phonetic overlap at word onset (doggie/dolly) or no overlap (e.g. doggie/tree). While doggie is distinguishable from tree at the beginning of the words, doggie and doll could not be discriminated by adults until the occurrence of the second consonant. The question of interest was whether infants hearing the word doggie would be slower to shift to the correct picture on dog/doll trials than on dog/tree trials. As with adults, 2-year-olds rapidly distinguished doggie from tree, but were slower to distinguish doggie from doll, suggesting that they were monitoring the speech signal continuously and taking advantage of word-initial information to identify these words. In the next study we presented 18- to 21-month-old infants with truncated target words in which only the first part of the word was available. If infants can identify familiar words when only the word-initial information is presented, this would provide even more convincing evidence for the early development of incremental processing. Infants heard sentences containing either a whole target word (e.g. baby, doggie, kitty), or a partial target word constructed by deleting the final segments of a whole word (e.g. /ba/, /da/, /ki/). We found that 18- and 21-month-old infants were able to recognize spoken words rapidly and reliably on the basis of partial segmental information.

These findings show that by the age of 18 months, when English-learning infants on average can speak fewer than 100 words, they are already beginning to make use of signal-complementary knowledge in word recognition. In the experiments described above, infants chose the matching picture using only word-initial phonetic information, making their move before the signal was completely specified. This required integrating information from the visual context with stored knowledge of word meanings and of the sounds associated with those meanings. As with most experimental procedures, the context here was highly constrained as compared to more natural situations in which language is spoken and understood, since the object named was visually present and the infant saw only two objects at a time. However, by imposing such experimental constraints we revealed aspects of infants’ developing competence in language understanding which could not be seen in spontaneous behavior at this age. The finding that infants increase the speed of recognizing familiar spoken words by 300 ms between 15 and 24 months is very likely related to parallel developments in the ability to remember and produce longer strings of words, an ability which is essential for learning syntax. Similarly, the finding that infants, like adults, can recognize some words without complete signal specification shows that young language learners are already integrating information from the acoustic signal with linguistic and nonlinguistic contextual information as they listen to continuous speech. Although the sources of signal-complementary knowledge available are minimal compared to those used by adults, infants take advantage of whatever sources they have, even in the very early stages of developing competence in spoken language understanding.
Adult Speech to Infants as 'Hyperspeech'

Although I am ignoring central issues in the H and H argument by focusing on infant speech processing, it is important to see how early language input fits into the larger framework of Lindblom’s [1990] theory. The goal of H and H theory is to provide a solution for the so-called invariance problem, which refers to the absence of consistent associations between phonemes and particular patterns of acoustic energy in the speech signal [see Perkell and Kluft, 1986]. The invariance problem turns out to be more problematic for theorists than it is for listeners, since word recognition is in fact generally effortless and unproblematic for adults. Lindblom [1990] argues that phonetic variability exists for good reason, and that the quest for acoustic invariance is misguided; rather than searching for a prototypical pattern of physical correlates for each speech sound, the quest should be redirected toward discovering how speech signals vary as a function of the fluctuating demands of the conversational situation. According to H and H theory, speakers continually adjust their productions to accommodate the listener. In situations where the listener has difficulty understanding, the speaker provides a clear signal by using more precisely articulated hyperspeech. In the converse situation, where the listener has access to sufficient signal-complementary information to follow the speaker’s meaning, the speaker reverts to use less clearly articulated hyperspeech. Thus phonetic variability reflects the lawful covariation between modulations in the acoustic signal along the hyper-hyperspeech continuum and the communicative pressures that induce these modulations. Speech to infants is relevant in the context of H and H theory for two related reasons: First, Lindblom [1990] is interested in the nature and early development of the vowel and consonant categories formed by infants. And second, ID speech may function as a form of hyperspeech which enhances intelligibility for the naive listener. I will focus here only on the second issue, which Davis and Lindblom [in press] rejected based on their finding of extensive phonetic variability in mothers’ speech.

ID speech has been characterized as listener-oriented because adults modify their speech in various ways which might be particularly effective when interacting with immature listeners. Caretakers speaking to infants in many cultures tend to use shorter utterances, longer pauses, higher pitch, and wider pitch excursions than when talking to an adult [e.g. Fernald and Simon, 1984; Fernald et al., 1989]. The intonation of ID speech is effective in engaging the infant’s attention and interest [e.g. Fernald, 1985; Werker and McLeod, 1990] and in eliciting emotional responses [e.g. Fernald, 1992, 1993]. In addition to these attentional and affective functions, ID speech may also enhance the intelligibility of speech for the child [Ferguson, 1977], a prediction consistent with the idea that ID speech is a hyperspeech mode. Several researchers have analyzed corpora of speech addressed to infants and adults to see if ID and AD speech differ in phonetic and prosodic features which might increase intelligibility [e.g. Bernstein-Ratner, 1986]. In a cross-language comparison of ID and AD speech in English, Swedish and Russian, Kuhl et al. [1997] found that vowels were somewhat more peripheral in the vowel space in ID speech. Although Kuhl et al. [1997] interpreted these findings as evidence for the enhanced distinctiveness of vowels in speech to infants, it is important to note that there was substantial acoustic variability among ID vowels in the cross-language sample, comparable to the variability Davis and Lindblom [in press] found in their analysis of mothers’ speech in English. Thus these results do not support the idea that ID speech helps infants by providing them with speech that is less variable than AD speech at the segmental level.
Davis and Lindblom [in press] concluded that mothers do not use clearly articulated hyperspeech with infants. Their assumption that infants have no access to the kinds of signal-complementary knowledge available to adults then led to their conclusion that the H and H account is not relevant to ID speech. However, it is not the case that infants rely only on the quality of the signal in word recognition, as we saw in the research reviewed above. Even in the 1st year, infants make use of their rudimentary knowledge of regularities in the speech they are hearing to identify potential linguistic units. And by the middle of the 2nd year, they are using lexical knowledge in combination with visual context and incomplete acoustic information to identify spoken words with impressive speed and accuracy. Because Davis and Lindblom [in press] underestimated the ability of young infants to make use of signal context, the hyperspeech model should not be discredited on these grounds.

The evidence does show that ID speech is not a hyperspeech form as originally defined in H and H theory, i.e. speech in which vowels and consonants are more clearly articulated and more homogeneous than in AD speech. However, the idea that adults intuitively and dynamically accommodate speech to infants in order to make their meanings more accessible to inexperienced listeners is still plausible. My proposal is that the hyperspeech notion should not be confined to articulatory factors at the segmental level, but should be extended to a wider range of factors in speech that facilitate comprehension by the infant. Lindblom [1990] describes hyperspeech as dynamic because speakers increase signal quality to compensate for the lack of signal-complementary information available to the listener. This account implies that speakers have control over the signal but have no control over the context in which the signal occurs. An even more dynamic account of hyperspeech would explore both how speakers articulate speech sounds more or less clearly in different contexts, and also how speakers continually adjust the kinds and extent of contextual information they provide to the listener. If the hyperspeech notion is expanded to include not only ways of enhancing signal clarity but also ways of increasing the availability of signal-complementary information, all in the service of successful communication between speaker and listener, then ID speech provides some good examples of hyperspeech in action.

How ID Speech Facilitates Word Recognition

In addition to the suprasegmental modifications in ID speech reviewed above, caretakers also arrange the words they speak to infants in ways which may facilitate comprehension. For example, focused words are frequently spoken in isolation [Aslin et al., 1996; Fernald and Morikawa, 1993], a rare occurrence in AD speech. When focused words are embedded in multiword utterances, English-speaking mothers often place them on pitch peaks at the end [Fernald and Mazzie, 1991]. English-speaking mothers also increase the duration of content words when talking to children [Albin and Echols, 1996; Swanson et al., 1992]. Another distinctive characteristic of ID speech is the high rate of repetition, as mothers either repeat themselves exactly or use a theme-and-variation style in successive utterances [e.g. Fernald and Morikawa, 1993].

Whether or not these ID speech modifications actually facilitate word recognition is a question that can only be addressed by testing infants' understanding of spoken
words in controlled experiments. Using the procedure described earlier, we have begun to explore the effects of various characteristic features of ID speech on word recognition by infants in the 2nd year. In one study, we tested the hypothesis that familiar words occurring at the end of the utterance would be recognized more reliably than the same words when embedded in the middle of the utterance [see Fernald et al., in press]. Subjects were 15- and 19-month-old English-learning infants, who heard sentences containing a familiar word either in utterance-final position (e.g. Over there there's a BALL) or in utterance-medial position (e.g. There's a BALL over there). The younger infants were able to identify the target word when spoken in final position, but not in medial position; the older infants performed above chance on both final and medial words, but were still significantly better at recognizing words in final position. Thus even when a familiar word was easily recognized when it came at the end of a sentence, the same word embedded in the middle of the sentence presented a more difficult processing task. These findings suggest that the common ID speech strategy of positioning focussed words at the end of the utterance has perceptual advantages for the language-learning infant.

In another study we explored the effects of repetition of the carrier phrase on word recognition by 18-month-old infants [Fernald and McRoberts, 1999]. Infants heard familiar words positioned at the end of nonsense carrier phrases made up of permutations of the same four syllables (e.g. Ba li gu do a BALL, Gu do ba li a BALL, etc.). In the highly predictable fixed frame condition, infants heard target words in the same carrier phrase on every trial; in the less predictable variable frame condition, infants heard the target words in four different carrier phrases in random order across trials. Infants in the fixed frame condition performed better overall, and also showed a learning effect over trials. That is, within a few trials in which the same carrier phrase was used repeatedly, they were able to figure out where to listen for the target word in the meaningless string of syllables. However, when the carrier phrase was less predictable, infants were less successful in listening for the familiar word. These results provide another example of how characteristic patterns of redundancy in ID speech can serve as a kind of support system for the novice listener.

When an adult speaking to a young child intuitively places the focussed word in a predictable and perceptually accessible position in the sentence, and uses short repetitive phrases to lead up to the word of interest, the word is easier for the child to recognize and understand. These common strategies in ID speech can be seen as ways of providing signal-complementary information which function more on a perceptual than a linguistic level in enabling the child to identify units in the stream of speech. When adults rely on contextual information in understanding speech, the expectations that inform their understanding are based on their knowledge of the language and on a sophisticated appreciation of the discourse situation. Infants can make only limited use of expectations based on linguistic knowledge, but they can make much greater use of expectations which are induced, so to speak, by the exaggerated redundancy of ID speech itself. That is, caretakers provide a form of signal-complementary information which does not depend on extensive knowledge of language and pragmatics in order to be accessible to the infant, but rather depends on more general auditory pattern recognition abilities. By maximizing predictability through frequent repetition and the use of formulaic speech patterns which highlight focussed words, ID speech provides contextual support on perceptual levels which are accessible even to pre-verbal infants.
Summary and Conclusions

Lindblom [1990] points out that H and H theory provides a ‘presupposition’ account of intraspeaker phonetic variation. In linguistics, this term is used with reference to how speakers vary the lexical, grammatical, and discourse-level information in their productions, based on assumptions about what the listener knows or needs to know in order to understand the message. Many experiments show how the intelligibility of a word in continuous speech depends partly on the clarity of the word itself, but also on the availability of contextual information which engages the listener’s stored linguistic knowledge and knowledge about the world. Such signal-complementary knowledge facilitates word recognition, and may be especially important when the signal is ambiguous. Viewing the phonetics of speech production as analogous to the generation of lexical and grammatical forms, Lindblom [1990] coined the term ‘hyperspeech’ to refer to the speaker’s tendency to adjust the quality of the signal to accommodate the listener’s needs. Although ID speech has been characterized as a listener-oriented speech register, Davis and Lindblom [in press] found extensive phonetic variability in speech to infants and thus concluded that ID speech is not a form of hyperspeech. They also reasoned that since infants have not developed the signal-complementary linguistic knowledge which would enable them to compensate for the inherent ambiguity of speech sounds, the variability in ID speech could not be understood in terms of H and H theory.

Davis and Lindblom’s [in press] conclusion makes sense if the definition of hyperspeech is limited to speech modulations at the phonetic level. However, if hyperspeech is also understood to include speech modulations of other kinds which serve to increase predictability and to generate expectations about where focused words will occur in the sequence of syllables, then ID speech can indeed be viewed as a hyperspeech form with beneficial consequences for the language-learning infant. Although phonetic variability is extensive in ID speech, infants have two sources of signal-complementary knowledge to exploit in word recognition. The first is their early sensitivity to regularities in the speech they hear which are correlated with structural aspects of the ambient language. The second is their attention to the local redundancies provided by caretakers who use a special form of hyperspeech when speaking to the infant, i.e. who repeat themselves frequently, emphasize focused words by placing them in perceptually prominent positions, and lighten the load of monitoring continuous speech by presenting new and focused information in predictable formats. Hyperspeech in this broader view is not used just to clarify the signal when contextual knowledge is lacking. ID speech also functions to increase the amount of contextual information available, by organizing the speech stream in such a way that focused words are likely to be salient even to a young listener with minimal linguistic experience. This account differs substantially from Lindblom’s [1990] formulation of hyperspeech in H and H theory and skirts the difficult question of how infants cope with phonetic variability. However, the functional view of ID speech proposed here suggests that the notion of hyperspeech can deepen our understanding of how signal-complementary information on many different linguistic and non-linguistic levels influences word recognition even in the earliest stages of language learning.
References


Aston, R.N.; Woodward, J.Z.; LaMendola, N.P.; Bejer, J.G.: Models of word segmentation in fluent maternal speech to infants; in Morgan, Demuth, Signal to syntax: bootstrapping from speech to grammar in early acquisition (Erlbaum, Hillsdale 1996).


Fernald, A.; McRoberts, G.W.; Swingley, D.: Infants' developing competence in recognizing and understanding words in fluent speech, in Weisborn, Hochele, Approaches to bootstrapping in early language acquisition (Benjamins, Amsterdam, in press).


