Speech perception, word learning, and language acquisition in infancy: The voyage continues

Janet F. Werker

Department of Psychology

The University of British Columbia

This manuscript has been published in revised form in Applied Psycholinguistics [https://doi.org/10.1017/S0142716418000243].

In the Keynote Article, I presented an overview of research on the perceptual foundations of language acquisition, focusing primarily on research from my own lab, but also situating this research within the broader field. Specifically, I presented theory and data on the speech perception sensitivities infants have at birth; how those have already been shaped by both evolution and prenatal experience, and how they continue to evolve in response to experience over the first months of life; what the mechanisms of change might be; and how changing speech perception sensitivities interface with language acquisition proper. Consideration of development in both monolingual and bilingual infants was given. Along the way, I tried to illustrate how my thinking has evolved in response to emerging empirical data and evolving theoretical constructs. Finally, I highlighted two specific new foci my lab is currently addressing. First, I noted that speech perception in infancy occurs frequently in face-to-face interactions – with considerable information from the visible talking face, as well as from the infant's own oral-motor movements – and noted that much of our work now examines speech perception and early language development as multisensory rather than unisensory. Second, I introduced another growing focus of my research (again, informed by research in the field showing lexical development much earlier than once believed) where I am now increasingly looking not only at how phonological development guides word learning, but also at how word learning might inform native phonetic category development, beginning even in early infancy, and how this bidirectional process may be a major mechanism by which perceptual attunement occurs.

The commentators addressed virtually every issue raised in the keynote article, unpacked several issues more than I had done, introduced relevant research (and indeed even entire relevant topic areas I had omitted), and presented some thoughtful critiques and challenges. The

inclusion of these commentaries and the various literatures raised by the authors enormously enrich and expand the review I prepared. While I will not be able to address all the important points that were made, a number of themes emerged that I would like to engage.

In her commentary, Demuth raised the question of whether there might be the same kinds of individual differences in perceptual development as have been documented in speech production. She noted that many research groups, including ours, are now looking at correlations between performance on speech perception tasks at one point in time and later language processing, word recognition, and other grammatical achievements. She suggested that it would also be of interest to compare on a finer scale how performance at one point in time (for example, discrimination of a particular stop distinction) relates to more advanced linguistic use at a later point in time (for example, use of that same stop distinction in word learning), and tracking this at the level of the individual rather than the level of the group. The imprecision of perception measures in comparison to production measures makes this challenging, but it is a challenge worth implementing more broadly than it has been to date.

Related to individual difference in perceptual development are the differences in the environments infants occupy. No two infants grow up in exactly the same language environments – even if they are twins raised in the same home, there will be some differences in the amount and quality of language input received. In her commentary, Johnson pointed out the enormous variability in input speech within and across monolingual and bilingual-learning infants. Some monolingual-learning infants hear only a single accent of a single language in all aspects of their daily lives, whereas others are raised with multiple dialects, with accented speech, or with a single language embedded in a background of many other languages. Similarly, the bilingual experience is extremely variable – from ‘one-parent, one-language’ to a rich

mixture from a single speaker and from multiple speakers. Recent research Johnson cited confirms that this variability matters and can, in some cases, hinder infant speech perception development. Yet, on the other hand, in their commentary, Singh, Morini, Golinkoff, and Hirsh-Pasek cite data showing that variability can sharpen sensitivity – in phoneme categorization tasks, in word segmentation tasks, and in learning similar sounding words. Better characterizing variability, and understanding when it limits and interferes with acquisition versus when it helps, are clearly important research goals for the future.

One population that is quite distinct from the language-learning populations much of our work is focused on, is hearing impaired infants, including those who have been given cochlear implants (CIs) as an intervention to restore hearing. The variability in this population includes all of the variability described above in monolingual and bilingual learning infants, but also variability in the degree of hearing loss, the age of implantation, whether a unilateral or bilateral implant is used, and so on. Concordantly, the variability in outcome and developmental trajectory is much greater than that seen in hearing infants. Much of Wang and Houston's commentary focused on the insights that have been obtained from this population. What their work shows is that this kind of variability can be harnessed in experimental tasks to reveal some regularities. We now know, for example, that just as preference for speech over non-speech predicts later language in infants at risk for autism (Droucker, Curtin & Voluoumanos, 2013), it also predicts later word recognition in children with CIs (Wang, Bergeson, & Houston, 2017). We know that age of implantation does matter, with less variability and better outcome the younger an infant is implanted. We know as well from the work Wang and Houston reviewed that CI infants and children are more likely to look at the mouth when successfully resolving speech, leading to the conclusion that multisensory information might be even more important

than it is in a hearing child. And, Wang and Houston presented data that children with CIs – even those children for whom the CIs are very successful at ameliorating hearing difficulties – appear to rely even more heavily than do hearing children on the meaning of a word rather than its surface acoustic characteristics to help guide phonetic category learning (as discussed more below).

Before leaving the characteristics of the infant learner behind, I should note that there are many other factors besides hearing loss that differ across infants. These include a broad range of factors from a history of middle ear infections to developmental disabilities that interfere with learning or even basic attention to speech (as in autism – a group studied by two of the commentators, Curtin and also Saffran). They also include differences in age of exposure as in infants born premature, or who move from one language community to another and who may be – e.g. in international adoption – completely then cut-off from their first language. There are also dietary (e.g. Innis, Gilley, & Werker, 2001) and pharmacological (e.g. Weikum et al., 2012) exposures that can change the timing of opening and closing of critical periods. It is within the background of all this variation that we search for regularities, explanations of change, and seek to understand how different developmental processes influence one another. Perhaps as data sharing becomes more and more standard (e.g. through projects such as ManyBabies, Frank et al., in press; Wordbank, http://wordbank.stanford.edu; and Databrary, https://databrary.org), and as data mining techniques continue to improve, we will be increasingly able to more methodically examine the influence of these variations.

Input speech provides infants with a signal from which to pull out statistical regularities. A number of commentators discussed statistical learning, but Singh, Morini, Golinkoff, and Hirsh-Pasek gave it perhaps the most extensive treatment. In so doing, they pointed out all the

ways in which statistical learning can support perceptual attunement, but also asked about its limitations. I couldn't agree more, and pointed out many of those same limitations – particularly of distributional learning – in my review article. Moreover, I noted those limitations as one of the reasons my students and I have continued to search for additional learning processes that could work together with distributional learning to help explain perceptual attunement, a notion to which I will return below. Singh and colleagues, however, situated their comments more within the context of whether laboratory experiments can fully account for how learning occurs in more naturalistic contexts – an important cautionary note many of us are trying to incorporate into our own research.

There was little disagreement concerning the importance of multisensory information in speech perception development. And indeed, many authors provided additional examples of the utilization of multisensory cues. One with which I was unfamiliar was discussed by Demuth. She raised the question of whether children who show greater sensitivity to multimodal cues will be advanced in speech production. My understanding from her literature review is that multisensory information may provide infants better access to their own means of production, which will not only support speech perception, but also their own productive vocabulary development. Related to this were the examples provided by Wang and Houston and from Johnson on how infants with CIs and infants raised bilingual pay more attention to the talking mouth when listening to speech.

The reader is also reminded by the commentators that both phonetic/phonological development and lexical development are only a part of the manifold of interacting aspects of speech perception and language development. Gerken's commentary reminds us that at the same time that they are establishing the phonetic and phonological categories of their native language, and beginning to learn words, infants are also establishing the foundations – at least in perception

– of syntax and morphosyntax. She noted, as well, that while it was once believed that syntax could only begin to develop once lexical knowledge was in place (and I might add, in particular once it could be seen in speech production), we now know infants are learning an enormous amount through simple bottom-up perception about the syntactic regularities of the native language. She pointed out, quite rightfully, that the remarkable advances made in understanding speech and language processing in infancy were made possible by the realization of just how much happens prior to the establishment of meaning, from Eimas and Juscyzk's transformative contributions of all the influences on phonetic and word form perception, (e.g. Eimas, Siqueland, Jusczyk, & Vigorito, 2091; Jusczyk, 1997; see Gerken & Aslin, 2005 for a review) to Gerken's and others' work on the perception of syntax and word order.[1]

While other linguistic features are one aspect of the broader context in which speech perception and word learning develop, so too are other communicative factors. As pointed out by Polka and Nazzi, it is just as important to infants – and perhaps more important early on – to figure out WHO is speaking as it is to attend to WHAT they are saying. Figuring out WHO is speaking, supports and guides figuring out WHAT they are saying (i.e. discriminating phonemes, learning words). Much of the 'who' information – along with other indexical cues, including gender, affect, and intonation – is carried on vowels. Polka and Nazzi's commentary reminded us that the course of vowel phonetic category acquisition is different than that of consonants, and that vowels appear to play a different role in early language acquisition than do consonants.

[1] I would just qualify here that I was not proposing that we return to the kind of Shavkin and Trubetskoy notions criticized in Gerken's commentary, that meaning has to come first before it can guide perceptual uptake of phonetic and syntatctic information. Rather, I was trying to stress in the keynote article that the acquisition of meaning and phonetics can occur simultaneously, with bidirectional influences.

More specifically, Polka and Nazzi review the work showing that at the same time they are establishing the native vowel repertoire (with perceptual attunement of vowels often developmentally earlier than consonants), infants are also using the rhythmic characteristics of language that are carried primarily on vowels to bootstrap acquisition of syntax, and are increasingly – in particular after 6-months – prioritizing consonant distinctions in lexical acquisition (e.g. Nazzi, Poltrok, & Von Holzen, 2016).

The most consistent theme raised in the commentaries (Curtin & Graham; Gerken; Polka & Nazzi; Saffran; Singh, Morini, Hirsh-Pasek & Golinkoff; Wang & Houston) relate to the proposal that the early learning of word meanings might be a driver, along with other mechanisms such as statistical learning, in phonetic category attunement. The commentaries raised issues concerning the lack of stability of both phonological forms and word meanings in the first two years of life, and referenced the PRIMIR framework that Curtin and I developed (with a later extension to bilinguals with Byers-Heinlein). In citing the lack of stability of phonological forms, they pointed out the considerable research showing that when infants first learn words, they often fail to recognize them when spoken in a different accent, or by a different speaker, or with different affect in the voice – but that by 20-24 months, infants no longer regularly fail at these tasks (although even adults are slower at word recognition when indexical cues change). In citing the lack of stability of word meaning, Gerken noted how infants can fail to recognize the referent of a word if the background color of the cloth on which the object sits changes. Additionally, several commentators cited the work showing that even as late as 24 months, infants often fail to recognize a newly-learned word when retested on it – even in as brief a time as 2 minutes. This is all correct and relevant, and I agree entirely with the point that neither word forms nor word meanings are as stable at 12-14 months as they will be at 20-24

months. However, as elaborated below, I would argue that stability is not necessary for a bidirectional influence to operate between word forms and their referents. Indeed, the co-occurrence and temporal alignment could be factors that not only drive phonetic category attunement, but that also facilitate better stabilization of both word forms and word meaning.

At the time we proposed PRIMIR, we had shown (Stager & Werker, 1997; Werker et al., 2002), that when infants are first able to learn words in an associative learning task (specifically, in the Switch task), they fail if the words differ in only a single phonetic feature, even though infants that same age can discriminate the same phonetic difference when tested in a simple nonsense-syllable discrimination task. While this work was replicated by us and others many times, another line of work – primarily testing recognition of already-known words – revealed that infants this same age can use minimal phonetic differences to distinguish whether the correct or a mispronounced version of a word is used to refer to an object (e.g. Swingley, 2003). Our original work was thus challenged. We and others took up that challenge in a number of subsequent studies, and found that while infants of 14- (but no longer at 18-20) months do have difficulty in the standard Switch task, they can succeed at learning to associate two phonetically similar words with two different objects when the computational demands are minimized in a number of different ways – e.g. if minimal pair words differing in highly-salient features such as stress or salient vowel differences are used, and even if they are first pre-familiarized with the words or the objects. Curtin and I felt the answer lay not in debating whose empirical results were accurate, but instead in trying to understand why under some circumstances infants could succeed in using fine phonetic detail in word learning and word recognition, and why in others they would fail. Thus, we proposed PRIMIR (processing rich information from multidimensional, interacting representations) as a framework to unpack the relationship

between initial infant speech perception, perceptual attunement to the native phonetic repertoire, and word learning.

In PRIMIR (Werker & Curtin, 2005; Curtin & Werker, 2007; also see Curtin, Byers-Heinlein, & Werker, 2011), we note that perceptual biases functional by birth provide organization for the perception and differentiation of speech sounds according to both phonetic and indexical properties. We called this the *general perceptual plane*. These initial biases act along with experience to enable the infant to move from language-general to language-specific phonetic categories, primarily through bottom-up statistical learning (e.g. distributional learning) processes. We further proposed that as they reach the end of the first year of life, infants begin to recognize and represent word forms and to learn about multidimensional visual object categories. Given that our focus was on speech processing, we proposed a second plane, the *word form plane*. Importantly, word form representations include both phonetic and indexical information. We suggested that gradually – as infants begin to learn enough associations between word forms and visual objects, and to experience those associations across different speakers and different contexts – they are able to pull out, through statistical tracking, the regularities of the phonetic features, because these remain relatively constant whereas the indexical features vary as a function of speaker, accent, affect, and so on. We argued that this process is what enables the infant to pull out a stable phonological representation and a functional phonological category. We called this the *phoneme plane*. With the emergence of phonological categories, they assume priority in word learning and word recognition contexts. Hence, while the indexical information remains equally perceptible, it is no longer given the same weight, thus prioritizing phonological contrast and enabling the use of fine (native) phonetic detail across a wider variety of word learning and word recognition tasks.

This process, by which weightings can change even once all planes are in place, is further explained by the operation of three dynamic filters: initial perceptual biases, task demands, and developmental level. These filters are like a changeable lens on a camera, which allows the infant to focus on and differentially weigh the information in the multiple interacting planes. Thus, were we to apply this framework to Nazzi and Polka's commentary, we would propose that if the task is one of knowing WHO is speaking – or if, given the developmental stage of the child, this information is more salient – then that is the lens used in perception. Focus on this particular aspect of the stimuli, then, could overwhelm the phonological category in a word recognition task, making the infant fail to recognize a recently-learned word if now produced by a different speaker or fail to distinguish that word from a phonetically similar words. If, however, even at 14 months, the task was set up in such a way that it was clear to the infant that it was one of referential word learning (e.g. by using short sentences or first showing the infant known word-object pairings, as in Fennell & Waxman, 2010), then the infant – even at this young age – could prioritize the phonetic information. By 20 months or so, when the phoneme plane has emerged, word learning tasks would function – given the child's developmental level – to highlight phonetic contrast over and above indexical differences. This, then, would result in the greater stability seen in phonological forms at this age.

I was very pleased that so many commentators raised PRIMIR more than I had done (and actually, almost scolded me for not talking about it more). In so doing, they helped me to reflect more deeply on what I think of PRIMIR, given the need to accommodate the new findings – from Tincoff and Juscyzk (1999), Bergelson and Swingley (2012), and increasingly many others – indicating that infants know many more words, and know them perhaps more deeply even if not fully referentially, than initially believed. When we proposed PRIMIR, it was widely

accepted that while infants might recognize an occasional highly frequent word, these were the exceptions – they were associatively-based representations that had no lexical depth, and that "real" referential understanding didn't likely come on-line until 18-20 months. While the evidence cited in so many of the commentaries makes it clear that in early infancy, words forms and word meanings are not as stable or robust by 18-20 months of age, there is increasing evidence that these early words do have greater lexical depth than believed even 6 months ago. For example, it is exciting to know that there is internal structure to the infant lexicon by 24 months of age (see Wojcik & Saffran, 2013), but Bergelson and Aslin's recent paper (2017) indicates that the beginnings of this structure are evident even in 6-month-olds. Specifically, when shown two objects that are either semantically related (e.g. juice and milk) or not (e.g. juice and foot) in an eye-tracking task, and presented with one of the words, 6-month-old infants look longer at the images if they are semantically unrelated than if they are related. This finding is taken as evidence that even very early word understanding has deeper meaning than a simple associative word-object link. In reminder, when Curtin and I first developed PRIMIR, no one had any inkling that infants might be beginning to learn the relation between word forms and their referents as early as 6 months of age. Thus, I think it represents a natural, data-informed evolution of PRIMIR to suggest that, for guiding the establishment of native categories, the co-occurrence of word forms and objects exists alongside the bottom-up statistical processes originally stressed.

Still, even if infants do understand some words from very early in development, the challenge remains: how might word meaning guide native phonetic category establishment if both the word forms and their meanings are unstable? I would suggest that aligning a word form with a referent even if each is unstable, rather than necessarily making the task more difficult,

could provide anchors by which the stabilization of each is greater (see also Saffran, 2014). We first provided empirical evidence for this in 2009 (Yeung & Werker) where we showed that at an age where they otherwise have stopped discriminating a non-native phonetic distinction, infants can show sensitivity to it if the two non-native sounds are paired with two different objects. In subsequent work, Yeung showed this effect to be even stronger if referential cues are given (Yeung, Chen, & Werker, 2013) and that the phonetic learning from word-object alignment generalizes (Yeung & Nazzi, 2014). Thus, even if this pairing is ephemeral and not well maintained, the temporary alignment may begin a process by which both the word form specification and that of the reference becomes more veridical with respect to the infants' language-learning environment. Indeed, in their 2017 paper, Bergelson and Aslin provide evidence that the words individual infants best recognize in the lab are those that correspond to common objects in their own homes. There is no reason to expect that these experiences are not similarly driving, then, the specification of the conventional, language-specific phonological form and hence contributing to the perceptual attunement that occurs in the first year of life.

In summary, the commentators raised a number of important issues, challenges, and caveats that augment and more broadly situate the content of my keynote article. I appreciated the opportunity to consider and address their comments. I particularly enjoyed the chance to revisit PRIMIR in more depth, given both the new data and the challenges raised. It is apparent from the lively discussion that ensued that although we have learned an enormous amount about infant speech perception and its relation to broader language acquisition, each new discovery raises new questions and opens new areas for research. I can only imagine what a keynote article and its responses might look like in another 10 years!

Acknowledgments:

I would like to acknowledge the support of funding from both the Social Sciences and Humanities Research Council of Canada (435-2014-0917) and the Natural Sciences and Engineering Research Council of Canada (RGPIN-2015-03967) in preparing this commentary, as well as support from the Canadian Institutes for Advanced Research (CIFAR) and the Canada Research Chairs Program. I would also like to thank my lab members for their discussion with me of the commentaries, and Savannah Nijeboer, my Research Manager, for her editorial comments and suggestions

References:

Bergelson, E., and Aslin, R. N. (2017). Nature and origins of the lexicon in 6-month-olds. *Proceedings of the National Academy of Sciences of the United States of America, 114*(49), 12916-12921.

Bergelson, E. & Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Science,* 113, 12397-12402.

Curtin, S., Byers-Heinlein, K., & Werker, J. F. (2011). Bilingual beginnings as a lens for theory development: PRIMIR in focus. *Journal of Phonetics, 39,* 492-504.

Curtin, S., & Werker, J. F. (2007). The perceptual foundations of phonological development. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 579-599). Oxford University Press.

Droucker, D., Curtin, S., & Vouloumanos, A. (2013). Linking infant-directed speech and face preferences to language outcomes in infants at risk for autism spectrum disorder. *Journal of Speech, Language, and Hearing Research, 56*, 567-576.

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

Fennell, C. T. & Waxman, S. R. (2010). What paradox? Referential cues allow for infants use of phonetic detail in word learning. *Child Development, 81*, 1376-1383.

Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., Hamlin, J. K., Hannon, E. E., Kline, M., Levelt, C., Lew-Williams, C., Nazzi, T., Panneton, R., Rabagliati, H., Soderstrom, M., Sullivan, J., Waxman, S., Yurovsky, D. (in press). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. *Infancy.*

Gerken, L. A. & Aslin, R. N. (2005). Thirty years of research on infant speech perception: The legacy of Peter W. Jusczyk. *Language Learning and Development, 1*, 5-21.

Innis, S. M., Gilley, J., & Werker, J. F. (2001). Are human milk long-chain polyunsaturated fatty acids growth related to visual and neural development in breast-fed term infants? *Journal of Pediatrics, 139*(4), 532-538.

Jusczyk, P. (1997). *The Discovery of Spoken Language*. MIT Press, Cambridge, MA.

Nazzi, T., Poltrock, S., & Von Holzen, K. (2016). The developmental origins of the consonant bias in lexical processing. *Current Directions in Psychological Science, 25*(4), 291-296.

Saffran J. (2014). Sounds and meanings working together: Word learning as a collaborative effort. *Language Learning. 1*;64 (Suppl 2):106-120.

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word learning tasks. *Nature, 388*(6640), 381-382.

Swingley, D. (2003). Phonetic detail in the developing lexicon. *Language and Speech, 46*, 265-294.

Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science, 10*(2), 172-175.

Wang, Y., Bergeson, T., & Houston, D. M. (in press). Infant-directed speech enhances attention to speech in deaf infants with cochlear implants. *Journal of Speech, Language, and Hearing Research*.

Weikum, W. M., Oberlander, T. F., Hensch, T. K., & Werker, J. F. (2012). Prenatal exposure to antidepressants and depressed maternal mood alter trajectory of infant speech perception. *Proceedings of the National Academy of Sciences, 109*(2), 17221-17227.

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, *1*(2), 197-234.

Werker, J. F., Fennell, C. T., Corcoran, K., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, *3*(1), 1-30.

Wojcik, E. H. & Saffran, J. R. (2013). The ontogeny of lexical networks: Toddlers encode the relationships amongst referents when learning novel words. *Psychological Science. 24*(10), 1898-1905.

Yeung, H. H., Chen, L. M., & Werker, J. F. (2014). Referential labeling can facilitate phonetic learning in infancy. *Child Development, 85*(3), 1036-1049.

Yeung, H. H., & Nazzi, T. (2014). Object labeling influences infant phonetic learning and generalization. *Cognition*, *132*(2), 151-163.

.