

Innate phonetic boundaries revisited (L)

Richard N. Aslin^{a)}

Department of Brain and Cognitive Sciences, University of Rochester, Rochester, New York 14627

Janet F. Werker

Department of Psychology, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

James L. Morgan

Department of Cognitive and Linguistic Sciences, Brown University, Providence, Rhode Island 02912

(Received 22 May 2002; accepted for publication 20 June 2002)

Nittrouer [J. Acoust. Soc. Am. **110**, 1598–1605 (2001)] raised several serious concerns with the so-called Universal Theory of phonetic category development that she characterized as the accepted wisdom in the field of infant speech perception. She then presented data from infants and children that were claimed to be inconsistent with Universal Theory and led her to question the entire notion of phonetic categories. Here we argue that Nittrouer not only misrepresented Universal Theory, but also provided no data to refute either this theory or the existence of phonetic categories. © 2002 Acoustical Society of America. [DOI: 10.1121/1.1501904]

PACS numbers: 43.71.An, 43.71.Ft, 43.71.Hw [KRK]

The study of infant speech perception was initiated in 1971 by the publication of Eimas *et al.*'s seminal paper on young infants' discrimination of voicing differences in synthetic consonant-vowel syllables varying in voice-onset-time (VOT). Since 1971, dozens of experiments have confirmed the initial observation that discrimination of the acoustic correlates of consonant contrasts is discontinuous. That is, not all physically equal steps along synthetic speech continua are equally easy or difficult for infants to discriminate. Rather, depending on the regions in which they fall, some steps are difficult to discriminate, much like the chance levels of performance obtained from adults for within-category differences, whereas other steps are relatively easy to discriminate, much like between-category performance obtained from adults.

The initial account of these findings suggested that infants process speech with mechanisms that are both innate and specifically linguistic. However, since 1971 there has been considerable debate in the literature about the proper level of analysis to invoke to account for infants' discrimination of speech signals. A *phonemic* account, initially attractive because infants' putative category boundaries closely matched those of their native language environment, was swiftly shown to be untenable by experiments revealing that (a) many non-native speech contrasts were discriminable by infants, (b) adults were often unable to discriminate (or had reduced sensitivity to) these non-native contrasts, and (c) infants' category boundaries did not always correspond to adult values in the native language. Moreover, many of the same discontinuities in speech discrimination observed in infants were also present in nonhuman animals (Kuhl and Miller, 1975, 1978; Kuhl and Padden, 1982, 1983), casting doubt on a *phonetic* account of speech sound discrimination.

One result of this initial wave of follow-ups to the Eimas *et al.* (1971) study was a perspective summarized by Aslin

and Pisoni (1980) in which speech discrimination by infants was accounted for largely by a general auditory rather than a phonetic mechanism. Consistent with the results of animal experiments, this perspective proposed that young infants were able to discriminate speech contrasts without recourse to a phonetic level of analysis by using general auditory mechanisms that were categorical in nature and independent of experience with any native language.¹ The role that experience with a native language played in speech perception was characterized by one of four mechanisms: (a) maturation—a gradual unfolding that was uninfluenced by experience, (b) induction—an unfolding that required specific experiences, (c) attunement—discriminative abilities that were partially present prior to the onset of experience and then shaped by that experience, or (d) maintenance—discriminative abilities that were fully mature at the onset of experience and were maintained or lost depending on the presence or absence of that experience. It is this last mechanism that Aslin and Pisoni described as a Universal Theory of the development of speech perception.

Nittrouer (2001) provides several challenges to what she claims are the major tenets of Universal Theory. First, she states, “Those early studies of infant speech perception led to the widely accepted view that infants are born with sensitivities to phonetic boundaries for all languages (i.e., the universal set)” (p. 1598). She then cites apparently contradictory findings from several studies of infants. One study (Holmberg *et al.*, 1977) found that infants have difficulty discriminating a fricative contrast, while another study (Lasky *et al.*, 1975) found that infants from a Spanish-speaking environment failed to discriminate a VOT contrast that is used in Spanish but succeeded in discriminating non-native VOT contrasts that are relevant in English and Thai. Nittrouer concludes, “In spite of these seemingly contradictory findings, however, the notion of innate phonetic boundaries has persisted” (p. 1599).

Ironically, Nittrouer (2001) cites the very results that Aslin and Pisoni (1980) used in support of Universal Theory.

^{a)}Electronic mail: aslin@cvs.rochester.edu

The fact that infants in the Lasky *et al.* (1975) study discriminated the English and Thai boundaries but not the Spanish boundary, as well as evidence from Streeter (1976) that infants from a Kikuyu environment discriminated the English boundary that is not used by Kikuyu adults, led Aslin and Pisoni to posit that these discriminations were *not* the result of a phonetic level of analysis. Rather, these data, along with the results from animal experiments that showed similar VOT boundaries, were taken as evidence for a general auditory-based form of categorical perception in young infants. While Nittrouer summarizes this account of speech categories, she states, “These variations, however, fail to contradict the basic tenets of the universal theory” (p. 1599).

Missing from Nittrouer’s (2001) summary of Universal Theory was how Aslin and Pisoni (1980) placed this theory in context: “It is important to note here that we are *not* claiming that only one of these classes of theories [Universal, Attunement, Perceptual Learning] will uniquely account for the development of *all* speech contrasts. Rather, it may be the case that some hybrid of the theories provides the best description of the development of specific classes of speech-sound discrimination. In fact, this view of parallel developmental processes appears to be supported by current empirical findings” (pp. 79–80). Aslin and Pisoni go on to state, “The evidence on the development of voicing perception therefore appears to provide good support for the *attunement* theory” (p. 84) [emphasis added], not the Universal Theory. Finally, Aslin and Pisoni make their position clear: “From this brief summary of the perception of voicing contrasts in stop consonants, fricatives, vowels, and liquids, it should be apparent that the major roles of early experience that we outlined earlier cannot be uniformly invoked to account for the development of the abilities needed to discriminate *all* speech contrasts found in spoken language” (p. 88).

Nittrouer’s (2001) incomplete and misleading summary of Aslin and Pisoni’s (1980) model of infant speech perception is mirrored in her treatment of more recent models, most notably that of Best (1994), which outlines several different ways in which classes of speech sounds are affected (or not) by early experience via a process of perceptual assimilation. Although Best did not adopt the term Universal Theory in her model, a mechanism analogous to maintenance is embedded, in part, in her conceptualization. Similarly, Werker and Tees (1984), who provided the first time-course data for the change in discriminative performance between 6 and 12 months of age for two non-native consonant contrasts, made it clear that this mechanism of age-related change was not “loss” *per se* (Werker, 1995), likely occurred through the application of more general learning mechanisms (Lalonde and Werker, 1995), and was not applicable to all phonetic contrasts (see also Best *et al.*, 1988; Polka and Bohn, 1996). Moreover, other more recent treatments of the effects of early experience on speech perception by infants (Aslin *et al.*, 1998; Jusczyk, 1997; Werker and Tees, 1999) provide a much more nuanced view than the one ascribed by Nittrouer (2001) under the guise of Universal Theory. Indeed, there is likely no one today who would describe the sensitivities shown by the newborn infant as straightforwardly “innate.” Werker and Tees (1992, 1999) characterize the

mechanism accounting for newborn sensitivities as one of “probabilistic epigenesis,” whereas Jusczyk and Bertoni (1988) describe it as one of “innately guided learning.” Much of the current research is focused on identifying precisely what the early biases and experiential mechanisms are that bring about both initial discrimination performance and subsequent change, from Jusczyk’s WRAPSA model (1993) to Kuhl’s (1991) Perceptual Magnet Model to Maye *et al.*’s (2002) and Anderson and Morgan’s (2002) demonstrations of infants’ abilities to use statistical information to alter category structure and to acquire auditory equivalence classes (Kuhl, 1983). In short, Nittrouer’s straw man has no clothes.

The second challenge offered by Nittrouer (2001) is that infants are much worse at speech sound discrimination than one would expect from a reading of the previous literature. The irony here is that she claims on several occasions (including in the abstract) that her “results did not differ from those reported by others.”² Nevertheless, she states that “Overall these results fail to provide support for claims that universal phonetic boundaries are in place at birth. It is emphasized that the findings of this study do not really differ from those of others: success rates are similar across studies. What differs is the willingness of authors to use the results to support claims of innate phonetic boundaries” (p. 1603). What leads Nittrouer to a different conclusion?

Her primary evidence for rejecting the notion of innate phonetic boundaries, setting aside the aforementioned arguments by some of us that these categories are not initially phonetic *per se*, is that not all infants succeed in discriminating them. That is, Nittrouer apparently believes that universal means “in all cases” and “for all infants.” Both of these criteria are misguided. First, we have already made the argument that Nittrouer incorrectly ascribed Universal Theory to all speech contrasts, which was never the claim of this theory since it was proposed in 1980. Thus, the fact that some speech contrasts are easier to discriminate than others simply cannot be used as an argument against Universal Theory, particularly since there are obvious differences in the acoustic salience of different speech contrasts. Nittrouer argues that “success rates differed across contrasts” (p. 1603) and that “If procedures accounted for a large proportion of variance in success rates, we would have expected those rates to be similar across contrasts” (p. 1603). But that statement flies in the face of compelling evidence that performance varies with simple acoustic parameters (e.g., changes in tone frequency) in the conditioned headturning paradigm (e.g., Olsho, 1984; Sinnott and Aslin, 1985). This finding that some distinctions are more difficult than others holds across a variety of methodologies in addition to conditioned headturning. These include visual habituation, heart rate deceleration, and mismatch negativity (MMN) in the event related potential (ERP) recorded from scalp electrodes. Yet, importantly, in all of these tasks infants consistently perform better on between- than on within-category differences (Miller and Morse, 1976; Best, 1994; Dehaene-Lambertz and Baillet, 1998). In addition to differences in discrimination, infants are better able to categorize some stimuli than others. At 6–8 months of age they treat multiple instances from within a dental category as equivalent, and as different from multiple

instances from a retroflex category, whereas they fail to categorize instances that span the dental/retroflex category boundary (Werker and Lalonde, 1988). It is this relative pattern of performance, rather than absolute levels of discriminability, which is important.

Second, no technique used to assess any aspect of infant performance succeeds in obtaining precisely the same data from the same infants upon repeated testing. Infants are notoriously difficult to test. We cannot directly instruct infants on the nature of experimental tasks, and their short attention spans preclude gradual shaping of desired behaviors. Rather, at best, procedures can be set up to maximize the likelihood that most infants will be sufficiently engaged in the task at hand to form approximately correct inductions about its nature most of the time. Thus, one has to work very hard to establish optimal conditions to obtain even adequate performance. Nittrouer's placement of infants in a car seat or high-chair alters the typical context in which testing is conducted in the conditioned headturning procedure and could easily result in suboptimal performance. Similarly, use of a very few test trials sets a strict standard that only a few infants, even under optimal conditions, will meet. For example, Nittrouer used a fixed criterion of 8 out of 10 correct responses to change trials as a threshold for concluding that infants discriminated a specific speech contrast. She notes that Werker's lab has used this same criterion in the past. However, with the exception of Werker and Tees (1984), where the criterion was 8 out of 10 responses to change trials only, Werker's lab has more typically used a floating criterion of correct responses to both change and control trials within a total of 25 trials. Thus, Nittrouer's criterion for successful discrimination is considerably more stringent than that of Werker and other researchers in the field who use the conditioned headturning technique.

As a result of her experiments with infants, Nittrouer (2001) expresses surprise that only 65% of her subjects discriminated the vowel contrasts and only 35% discriminated the consonant contrasts. But why should we expect all infants to "pass" a test of discrimination, particularly with her testing setup and performance criterion? Surely an inattentive adult would show less than perfect performance on any task, and infants are not immune from such lapses of attention and motivation. Nittrouer argues that these task variables should be independent of the difficulty of the speech contrast. But that argument is simply not compelling: surely one would expect performance to vary with task difficulty even if there are general task variables (e.g., criterion effects) operating to reduce optimal performance. These issues are moot, however, because Nittrouer claims that her data are not that different from what others have reported in the literature. Why then does she come to such a different conclusion about innate phonetic categories?

A key difference between Nittrouer's interpretation of her data and that of others in the field is that she holds an idealized view of categorical perception. This view posits that labeling functions have sharp category boundaries and that within-category discrimination is at chance while between-category discrimination is at ceiling. The problem is that no data from infants have ever supported such a "strictly

categorical" view of speech perception. In fact, there are no labeling data at all from infants to complement their discontinuous discrimination data. Thus, at best, the infant literature supports a weak view of categorical perception. Interestingly, the adult literature supports a similar view, with reaction time (Pisoni and Tash, 1974), goodness ratings (Miller, 1994), and same-different judgments (Carney *et al.*, 1977) supporting a less definitive view of categorical perception than originally attributed to Liberman *et al.* (1967). But of course this does not imply, as suggested by Nittrouer, that "Perhaps we should not even be asking if infants have well-formed phonetic categories, separated by boundaries, but rather if any language users do. In other words, the very concept of categories, and even more so of boundaries, needs to be reconsidered" (p. 1604). Clearly, there is an interpretive agenda here that goes well beyond the infant data that Nittrouer reports and compares to the existing infant data in the literature: "we have no evidence that boundaries exist in the natural world, or any account of how or why they might have evolved by natural selection. To extend to them any degree of psychological reality is unsupportable, and deleterious to efforts to understand how phonetic structure is indeed instantiated and retrieved from the speech signal" (p. 1604). One can only imagine what Nittrouer believes is a sufficient replacement for the notion of speech categories (and their logically necessary boundaries) since she offers no such account.

In summary, Nittrouer (2001) has provided a naive and incomplete view of Universal Theory and of categorical perception. She used methods commonly employed in the infant speech perception literature, though in a strikingly suboptimal manner. As a result, her findings on speech discrimination are at the low end of those reported previously. This combination of factors—weak data and an idealized view of Universal Theory and categorical perception—led her to conclude that infants have no phonetic categories or boundaries between them. We believe these conclusions are misleading and do not fairly represent the current thinking in the field of infant speech perception. While we may disagree about some of the subtle aspects of infant speech perception, Nittrouer's article throws the entire "baby" out with a very muddled perspective on the "bath water." As such, Nittrouer's challenge to the field of infant speech perception fails to measure up.

¹Other past accounts, such as that of Werker and Pegg (1992), proposed three separate patterns: one corresponding to general auditory sensitivities, one to language-general phonetic differences, and one to language-specific meaning-based phonemic differences.

²In fact, the studies that Nittrouer cites as comparisons tested 2-month-olds in habituation procedures, rather than 6- to 14-month-olds with conditioned headturning. As discussed below, Nittrouer adopts an excessively stringent criterion; as might be expected, the proportion of subjects failing to reach criterion in Nittrouer's study was much higher than that of other studies testing infants of the same age with the same technique (e.g., Morgan, 1994; Werker and Tees, 1984).

Anderson, J. L., and Morgan, J. L. (2002). "A statistical basis for speech sound discrimination," poster presentation at the International Conference on Infant Studies, Toronto, Canada.

Aslin, R. N., and Pisoni, D. B. (1980). "Some developmental processes in speech perception," in *Child Phonology, Vol. 2, Perception*, edited by

- G. H. Yeni-Komshian, J. F. Kavanagh, and C. A. Ferguson (Academic, New York), pp. 67–96.
- Aslin, R. N., Jusczyk, P. W., and Pisoni, D. B. (1998). "Speech and auditory processing during infancy: Constraints on and precursors to language," in *Handbook of Child Psychology, Vol. 2: Cognition, Perception & Language*, 5th ed., edited by D. Kuhn and R. Siegler (Wiley, New York), pp. 147–198.
- Best, C. T. (1994). "The emergence of native-language phonological influences in infants: A perceptual assimilation model," in *The Development of Speech Perception*, edited by J. C. Goodman and H. C. Nusbaum (MIT, Cambridge, MA), pp. 167–224.
- Best, C. T., McRoberts, G. W., and Sithole, N. M. (1988). "Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants," *J. Exp. Psychol. Hum. Percept. Perform.* **4**, 45–60.
- Carney, A. E., Widin, G. P., and Viemeister, N. F. (1977). "Noncategorical perception of stop consonants differing in VOT," *J. Acoust. Soc. Am.* **62**, 961–970.
- Dehaene-Lambertz, G. L., and Baillet, S. (1998). "A phonological representation in the infant brain," *NeuroReport* **9**, 1885–1888.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P. W., and Vigorito, J. (1971). "Speech perception in early infancy," *Science* **171**, 304–306.
- Holmberg, T. L., Morgan, K. A., and Kuhl, P. K. (1977). "Speech perception in early infancy: Discrimination of fricative consonants," *J. Acoust. Soc. Am. Suppl. 1* **62**, S99.
- Jusczyk, P. W. (1993). "From general to language specific capacities: The WRAPSA model of how speech perception develops," *J. Phonetics* **21**, 3–28.
- Jusczyk, P. W. (1997). *The Discovery of Spoken Language* (MIT, Cambridge, MA).
- Jusczyk, P. W., and Bertoncini, J. (1988). "Viewing the development of speech perception as an innately guided learning process," *Lang Speech* **31**, 217–235.
- Kuhl, P. K. (1983). "Perception of auditory equivalence classes for speech in early infancy," *Infant Behav. Dev.* **6**, 263–285.
- Kuhl, P. K. (1991). "Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not," *Percept. Psychophys.* **50**, 93–107.
- Kuhl, P. K., and Miller, J. D. (1975). "Speech perception by the chinchilla: Voiced-voiceless distinction in the alveolar-plosive consonants," *Science* **190**, 69–72.
- Kuhl, P. K., and Miller, J. D. (1978). "Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli," *J. Acoust. Soc. Am.* **63**, 905–917.
- Kuhl, P. K., and Padden, D. M. (1982). "Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques," *Percept. Psychophys.* **32**, 542–550.
- Kuhl, P. K., and Padden, C. (1983). "Enhanced discriminability at the phonetic boundaries for the place feature in macaques," *J. Acoust. Soc. Am.* **73**, 1003–1010.
- Lalonde, C., and Werker, J. F. (1995). "Cognitive influences on cross-language speech perception in infancy," *Infant Behav. Dev.* **18**, 459–476.
- Lasky, R. E., Syrdal-Lasky, A., and Klein, R. E. (1975). "VOT discrimination by four to six and a half month old infants from Spanish environments," *J. Exp. Child Psychol.* **20**, 215–225.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. G. (1967). "Perception of the speech code," *Psychol. Rev.* **74**, 431–461.
- Maye, J., Werker, J. F., and Gerken, L. A. (2002). "Infant sensitivity to distributional information can affect phonetic discrimination," *Cognition* **82**, B101–B111.
- Miller, C. L., and Morse, P. A. (1976). "The 'heart' of categorical speech discrimination in young infants," *J. Speech Hear. Res.* **19**, 578–589.
- Miller, J. L. (1994). "On the internal structure of phonetic categories: A progress report," *Cognition* **50**, 271–285.
- Morgan, J. L. (1994). "Converging measures of speech segmentation in preverbal infants," *Infant Behav. Dev.* **17**, 389–403.
- Nittrouer, S. (2001). "Challenging the notion of innate phonetic boundaries," *J. Acoust. Soc. Am.* **110**, 1598–1605.
- Olsho, L. W. (1984). "Infant frequency discrimination," *Infant Behav. Dev.* **7**, 27–35.
- Pisoni, D. B., and Tash, J. (1974). "Reaction times to comparisons within and across phonetic categories," *Percept. Psychophys.* **15**, 285–290.
- Polka, L., and Bohn, O-S. (1996). "A cross-language comparison of vowel perception in English-learning and German-learning infants," *J. Acoust. Soc. Am.* **100**, 577–592.
- Sinnott, J. M., and Aslin, R. N. (1985). "Frequency and intensity discrimination in human infants and adults," *J. Acoust. Soc. Am.* **78**, 1986–1992.
- Streeter, L. A. (1976). "Language perception of 2-month-old infants shows effects of both innate mechanisms and experience," *Nature (London)* **259**, 39–41.
- Werker, J. F. (1995). "Exploring developmental changes in cross-language speech perception," in *An Invitation to Cognitive Science. Part I: Language*, edited by L. Gleitman and M. Liberman (MIT, Cambridge, MA), pp. 87–106.
- Werker, J. F., and Lalonde, C. E. (1988). "Cross-language speech perception: Initial capabilities and developmental change," *Dev. Psychol.* **24**, 672–683.
- Werker, J. F., and Pegg, J. E. (1992). "Infant speech perception and phonological acquisition," in *Phonological Development: Models, Research, and Implications*, edited by C. Ferguson, L. Menn, and C. Stoel-Gammon (York, Parkton, MD), pp. 285–311.
- Werker, J. F., and Tees, R. C. (1984). "Cross-language speech perception: evidence for perceptual reorganization during the first year of life," *Infant Behav. Dev.* **7**, 49–63.
- Werker, J. F., and Tees, R. C. (1992). "The organization and reorganization of human speech perception," *Annu. Rev. Neurosci.* **15**, 377–402.
- Werker, J. F., and Tees, R. C. (1999). "Influences on infant speech processing: Toward a new synthesis," *Annu. Rev. Psychol.* **50**, 509–535.