

# A reply to “Innate phonetic boundaries revisited” [J. Acoust. Soc. Am. 112, 1257–1260 (2002)] (L)

Susan Nittrouer<sup>a)</sup>

Center for Persons with Disabilities, Utah State University, Logan, Utah 84322

(Received 17 June 2002; accepted 20 June 2002)

[DOI: 10.1121/1.1501906]

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

Before launching into a response to the letter by Aslin, Werker, and Morgan (2002), I would like to make clear my belief that the work on infant speech perception has been well executed, and has been critical to the development of general theoretical perspectives on speech perception. I respect and appreciate the contributions of these authors, and consider my own work with children to be a natural extension of their and others' work on infant speech perception.

At the same time, the questions starting to be asked by speech researchers who are not focused solely on infant perception reflect a dramatic change in theoretical perspective, one that is not matched in magnitude by changes in perspective in the infant speech perception work. Many investigators in other areas of speech perception are no longer concerned with trying to identify the acoustic correlates of phonetic categories, or with asking at what age human listeners are able to classify those correlates, or with asking if nonhuman animals can classify those correlates. Nor are other investigators much concerned with distinctions between “auditory” and “phonetic” perception, a recurring theme in the letter of Aslin *et al.* A general paradigm shift is underway in studies of speech perception toward a focus on how listeners perceptually organize the many properties of the speech signal. For example, Robert Remez and his colleagues (e.g., Remez *et al.*, 1994) and Robert Shannon and his colleagues (e.g., Shannon *et al.*, 1995) have convincingly demonstrated that mature listeners can organize signals that are void of the detailed properties that we usually consider acoustic correlates (or “cues”) of phonetic categories in such a way that they still manage to arrive at a phonetically structured message. Investigators studying cross-linguistic speech perception are grappling with fundamental notions of what the objects of speech perception are, and what that means for our understanding of how the perceptual organization of the speech signal is shaped by experiences with a native language (Strange, 1995). What we consider to be the initial state of the human capacity for speech perception affects the further development of theories related to these topics, as well as many others.

The purpose of the brief article serving as the target of Aslin *et al.*'s letter was to state explicitly that the evidence is weak to support the position that infants are born able to discriminate all the phonemes of the world's languages. The authors say it best themselves on page 1259 of their letter, “Thus, at best, the infant literature supports a weak view of

categorical perception.” I believe it is imperative that we (i.e., those of us studying speech perception by immature listeners) make this point loudly and clearly to our colleagues studying related questions regarding speech perception. I think it is a fair representation of the field of infant speech perception to say that the collective message for several decades has been that infants are endowed with the “universal set” of phonetic categories (i.e., those that could occur in all the languages of the world), and that the ambient language functions to maintain those that will be used. According to this view, those that will not be used are lost, somehow. The Introduction of the target paper provides several examples of statements supporting this position, and they will not be repeated here. Many others exist, both in the writing of investigators doing the research and in the writing of others citing those investigators. For example, a popular text by Rhea Paul (2001) for students of speech-language pathology states “Speech perception, for example, is biologically programmed. We know this because infants as young as 4 weeks can distinguish between phonemes, even when they have no comprehension of language” (p. 399). In fact, that message has even made it into the popular media, with several PBS documentaries during the past decade or so featuring one or another infant speech perception researcher reporting that newborns are citizens of the world, able to discriminate all phonetic contrasts that can occur (i.e., the “universal set”). Those of us who have been around for a while recognize this position as the perceptual counterpart of an older view of infants' productions, which was that infants babble all the sounds of the world's languages; those not supported by the ambient language are lost. [Locke (1983) reviews this position, and the evidence against it.]

Assumptions about initial states profoundly impact research with children. In 1974, Isabelle Liberman and her colleagues (Liberman *et al.*, 1974) found that kindergartners were not as skilled as second graders at counting phonetic segments in words. Since the kindergartners were much better at counting syllabic segments (than they were at counting phonetic segments), I. Liberman *et al.* concluded that the poor results for phonetic segments did not have to do with task demands, but instead accurately reflected the level of structure the children could access. Of course, explicitly accessing phonetic structure in a counting task is different from discriminating pairs of syllables that differ by one phonetic segment. Nonetheless, both tasks inform us about linguistic organization in very young listeners. Thus, we are faced with a paradox. If infants readily discriminate syllables that differ

<sup>a)</sup>Electronic mail: nittrouer@cpd2.usu.edu

by one phonetic segment, why aren't they able to access that phonetic structure at five years of age?

Assumptions about initial states impact research and intervention with children having trouble learning language. We know that children experiencing conditions such as poverty or chronic middle-ear effusion have greater difficulty than age-matched peers with speech perception and phonological awareness tasks (e.g., Eimas and Clarkson, 1986; Godfrey *et al.*, 1981; Mody *et al.*, 1999; Nittrouer, 1996). Those of us interested in these (and other) clinical populations would like to understand why such conditions produce speech perception problems, and what we can do about it. The way in which we conceptualize the speech perception capacities of the infant affects how we understand the constellation of findings across studies with infants, children, and children with risk factors for language problems. Consequently, I felt it was imperative to report my finding that not all infants demonstrate capacities for phonetic discrimination, and to emphasize that some other studies have supported the same conclusion. I could have kept my results to myself, and so avoided controversy, but I think such findings must be acknowledged in future theoretical development. In the remainder of this letter, I would like to address the major criticism made by Aslin *et al.* of the target paper.

But what *is* Aslin *et al.*'s major criticism? It is never clear. At some points in the letter the criticism seems to be that I was inaccurate in my assertion that infant speech perception investigators have said that infants are born able to discriminate every phonetic contrast that could occur in the world's languages, with maintenance or loss of those discriminative capacities determined by which contrasts are found in the ambient language (i.e., Universal Theory). The authors assert that I built a "straw man with no clothes." At other points they suggest that I failed to dispel this very notion with the evidence I presented because my assumptions and data collection techniques were flawed. Because there is no focus, I will deal with each theme in the letter separately.

Aslin *et al.* fret that the target paper credited Aslin and Pisoni (1980) with offering Universal Theory as a good description of infant speech perception capacity. Perhaps the first author of the letter wants to distance himself from this position. If that is the case, let me assure him that I never attributed such a position either to him or to David Pisoni. The target paper was clear in stating that the conclusion reached by most investigators doing early infant speech perception research has been that newborns can discriminate all phonetic contrasts that could occur in the world's languages, and that over the first few months of life those not supported by the ambient language are "lost." What the target paper states is that this view matches what Aslin and Pisoni termed Universal Theory. Specifically, the statement made was:

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). Experience listening to a native language during the first year of life, the theory holds, maintains those boundaries supported by the ambient language,

and causes those boundaries not supported by the ambient language to dissolve. This view of perceptual development is what Aslin and Pisoni (1980) call a "universal" theory. (Nittrouer, 2001, p. 1598)

Aslin *et al.* are also concerned that I misrepresented Universal Theory, as defined by Aslin and Pisoni (1980). The text above shows how it was described in the target paper. Aslin and Pisoni wrote:

Universal theory assumes that, at birth, infants are capable of discriminating all the possible phonetic contrasts that may be used phonologically in any natural language. According to this view, early experience functions to maintain the ability to discriminate phonologically relevant distinctions, those actually presented to the infant in the environment. However, the absence of phonologically irrelevant contrasts, which are obviously not presented to the infant, results in a selective loss of the abilities to discriminate those specific contrasts. (p. 79)

In fact, in discussing Aslin and Pisoni's chapter, Aslin *et al.* write:

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. (p. 1257)

I fail to see the alleged discrepancy between either the Aslin and Pisoni definition or that offered in the letter of Aslin *et al.* and my description of Universal Theory.

Similarly, Aslin *et al.* suggest that the target paper misrepresented other papers, but then present exactly the same interpretation of those papers as that provided in the target paper. For example, they write:

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. (p. 1258)

I believe that is exactly what the target paper said: Best (1994) is offered as an example of a paper espousing a (slightly modified) version of Universal Theory.

As this example shows, there are places in the letter of Aslin *et al.* where the rhetoric is strong, yet no clear blow is struck. There are also places in the letter where the target paper is criticized for failing to appreciate aspects of current thinking regarding infant speech perception that no one could ever be expected to appreciate. For example, Aslin

*et al.* write “Indeed, there is likely no one today who would describe the sensitivities shown by the newborn infant as straightforwardly ‘innate.’” (p. 1258) In this context, “no one” presumably means “no one doing infant speech perception research” since no one in any other area of investigation would describe the capacities of newborns as anything other than innate. “Present in newborns or at birth” is, in fact, the definition of the word “innate.”

Aslin *et al.* offer many reasons why a demonstration of innate abilities to discriminate phonetic contrasts in all the world’s languages should not be held to strict criteria. For example, they state that some speech contrasts are simply easier to discriminate than others, due to variation in acoustic salience. Given such a suggestion, it would have been nice if they had offered a hint as to how acoustic salience should be measured, but it doesn’t really matter because appeals to such notions have never helped in our understanding of human speech perception. For any given phonetic contrast, regardless of its acoustic salience (however defined), listeners vary in their abilities to discriminate it, depending on their native language experience. That is, the acoustic salience of a given contrast presumably remains constant across listeners of different languages, yet abilities to discriminate it vary across listeners. If indeed discriminative capacities are fully mature at the onset of experience, then we should expect infants’ discriminative capacities for phonetic contrasts to be as unaffected by variation in acoustic salience as those of adults. Besides, if the suggestion being made is that I failed to find evidence of discrimination for some contrasts because they were not acoustically salient, it should be recalled that the three contrasts used were ones that others have reported infants discriminate successfully (i.e., vowels, VOT, and sibilants).

Aslin *et al.* claim that the target paper was idiosyncratic in its view of categorical perception. They suggest that infant speech perception investigators have historically tested a version of categorical perception different from (i.e., weaker than) the one against which I was judging that work. However, all descriptions of paradigms being tested in the infant speech perception work are consistent with the classic notion of categorical speech perception, and there was nothing in the target paper to suggest that the view being discussed was different from (i.e., stronger than) the classic notion of categorical speech perception. In fact, Aslin *et al.* describe the collective finding of the infant speech perception work at the start of their letter by writing:

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 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. (p. 1257)

That account certainly sounds like a description of classic categorical speech perception, and it sounds like the claim is being made that the performance of infants demonstrates classic categorical speech perception. The purpose of the target paper was explicitly to state that collective results from infants do not support such a strong claim. And in fact, Aslin *et al.* contradict their own statement (just above) toward the end of their letter when they write “Thus, at best, the infant literature supports a weak view of categorical perception.”

Toward the end of the target paper, I raise the point that perhaps we should be asking questions about the concept of phonetic categories as objects of perception for all language users, not just for infants. And again, Aslin *et al.* seem to make a similar suggestion toward the end of their letter, citing evidence from other investigators showing that phonetic categories are not as solid as standard views of categorical speech perception suggest they would be. So, at this point in their letter (p. 1259), I am not sure where we disagree, and so am not sure what the source of their consternation is. In any event, worrying about the precise definition of categorical perception used is irrelevant, at least regarding the target paper. Notions of categorization differ in their views of boundary effects and of the basis of equivalence for members of a given category. The work reported in the target paper involved discrimination of natural tokens of phonetic categories. Any version of categorical perception would expect these tokens to be discriminated by listeners with representations of these categories.

Given that there actually seems to be agreement between the points made by the target paper and points made toward the end of Aslin *et al.*’s letter, it is unclear why my methods were so severely attacked. However, concerning that attack, let me say that such an approach to disagreement is disappointing among colleagues, and simply inappropriate. Although I spent several years tweaking the methods reported in the target paper, at no point did procedures differ drastically from those of others. And at no time did slight variations in methods alter outcomes: In addition to the 121 infants and children whose data were reported in the target paper, we tested roughly another 150 using slight variations of the methods reported there, and we never found strong evidence of discrimination. At all times methods were sound, reflecting my own background working with children and infants, and even doing some hearing testing with visual reinforcement methods. I had many well-respected colleagues at Boys Town National Research Hospital who were able to offer input concerning the use of visually reinforced head-turning procedures and psychophysical method, and I contacted several infant speech perception researchers across the country to obtain further guidance. Nonetheless, let us for a moment extend credence to the concerns of Aslin *et al.*, and address them.

The first criticism is that the use of an infant seat or highchair could negatively impact children’s responding, as most other investigators have had infants sit on parents’ laps. Although I really don’t think we want to consider too seriously the possibility that infant speech perception might be affected by where the child sits, it should be said that my work did not represent the first use of an infant seat for

testing. I refer readers to the figure on p. 245 of Kuhl (1985). Here we see an infant being tested while in an infant seat.

The next criticism is that my criteria of requiring 8 out of 10 correct responses to change trials, with no more than one false alarm to a no-change trial, was overly strict. By the standards of most psychophysical testing, that simply is not a stringent criterion. On the contrary, using the criterion of 8 out of 10 correct responses, regardless of whether they are change or no-change trials, as the authors suggest, is really a very lax criterion. Assuming that the ratio of change to no-change trials is 0.5, this means that the infant need only turn to 3 out of 5 change trials (i.e., 60%) to be judged as making the discrimination. The other 5 trials judged as correct can all be no-change trials, and it is not difficult to maintain an infant's attention well enough that the probability of a spontaneous headturn is quite low. Thus we are left with Aslin *et al.* suggesting that if one adopts a conceptualization of "weak" phonetic categories, and uses a lax criterion to judge that infants recognize contrasts among these categories, we may expect an infant to successfully discriminate exemplars of phonetic categories, as long as she is not in an infant seat and the contrast is acoustically salient. In sum, if we must accept all the caveats and concessions to infant speech perception research that Aslin *et al.* propose, there are as yet no strong conclusions to be reached concerning infants' capacities for speech perception, which was the message of Nittrouer (2001).

I end this letter as I began it, by stating that the work of investigators studying infant speech perception is laudable, and has been instrumental in the development of our general theories of speech perception. However, it is time to abandon our old notions of phonetic categories as objects of perception for either infants or adults. We need both new conceptual frameworks and new methods. I do not have a solid suggestion (yet) for what we should set in place of phonetic categories as the objects of speech perception. But that does not mean we should cling to our old notions of categories and boundaries distributed along theoretical continua never observed in natural speech—particularly when these notions no longer advance our understanding of basic human speech perception, of what goes wrong when children fail to acquire speech normally, or of the best methods for intervention with such children. Most models of speech perception used for research and clinical practice involve (either explicitly or implicitly) some version of the phonetic segment as the basic unit of speech organization, the "building block" of language, as it is described in textbooks. Accordingly, the common view of speech perception is of listeners extracting cues (or features) and compiling them into phonemes from which they then create higher levels of linguistic structure. An alternative view suggested by a diverse range of studies (e.g., Browman and Goldstein, 1990; Nittrouer and Crowther,

2001; Remez *et al.*, 1994; Shannon *et al.*, 1995; Surprenant and Goldstein, 1998) is of listeners organizing the various properties of an acoustic signal into a precisely timed, language-specific perceptual structure from which linguistic (including phonetic) forms somehow emerge. Although not yet fully framed, this new view would fundamentally change our research paradigms and our clinical practice.

## ACKNOWLEDGMENTS

This work was supported by research Grant No. R01 DC00633 from the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health.

- Aslin, R. N., and Pisoni, D. B. (1980). "Some developmental processes in speech perception," in *Child Phonology*, edited by G. H. Yeni-Komshian, J. F. Kavanagh, and C. A. Ferguson (Academic, New York), Vol. 2, pp. 67–96.
- Aslin, R. N., Werker, J. F., and Morgan, J. L. (2002). "Innate phonetic boundaries revisited," *J. Acoust. Soc. Am.* **112**, 1257–1260.
- 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.
- Browman, C. P., and Goldstein, L. (1990). "Representation and reality: Physical systems and phonological structure," *J. Phonetics* **18**, 411–424.
- Eimas, P. D., and Clarkson, R. L. (1986). "Speech perception in children: Are there effects of otitis media?" in *Otitis Media and Child Development*, edited by J. F. Kavanagh (York, Parkton, MD), pp. 139–159.
- Godfrey, J. J., Syrdal-Lasky, A. K., Millay, K. K., and Knox, C. M. (1981). "Performance of dyslexic children on speech perception tests," *J. Exp. Child Psychol.* **32**, 401–424.
- Kuhl, P. K. (1985). "Methods in the study of infant speech perception," in *Measurement of Audition and Vision in the First Year of Postnatal Life*, edited by G. Gottlieb and N. A. Krasnegor (Ablex, Norwood, NJ), pp. 223–251.
- Lieberman, I. Y., Shankweiler, D. P., Fischer, F. W., and Carter, B. (1974). "Explicit syllable and phoneme segmentation in the young child," *J. Exp. Child Psychol.* **18**, 201–212.
- Locke, J. L. (1983). *Phonological Acquisition and Change* (Academic, New York).
- Mody, M., Schwartz, R. G., Gravel, J. S., and Ruben, R. J. (1999). "Speech perception and verbal memory in children with and without histories of otitis media," *J. Speech Lang. Hear. Res.* **42**, 1069–1079.
- Nittrouer, S. (1996). "The relation between speech perception and phonemic awareness: Evidence from low-SES children and children with chronic OM," *J. Speech Hear. Res.* **39**, 1059–1070.
- Nittrouer, S. (2001). "Challenging the notion of innate phonetic boundaries," *J. Acoust. Soc. Am.* **110**, 1598–1605.
- Nittrouer, S., and Crowther, C. S., (2001). "Coherence in children's speech perception," *J. Acoust. Soc. Am.* **110**, 2129–2140.
- Paul, R. (2001). *Language Disorders: From Infancy through Adolescence* (Mosby, St. Louis).
- Remez, R. E., Ruben, P. E., Berns, S. M., Pardo, J. S., and Lang, J. M. (1994). "On the perceptual organization of speech," *Psychol. Rev.* **101**, 129–156.
- Shannon, R. V., Zeng, F. G., Wygonski, J., Kamath, V., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303–304.
- Strange, W. (1995). *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (York, Baltimore).
- Surprenant, A. M., and Goldstein, L. (1998). "The perception of speech gestures," *J. Acoust. Soc. Am.* **104**, 518–529.