1291b

Available online at www.sciencedirect.com
SCIENCE DIRECT

Journal of Phonetics 31 (2003) 613–618

Journal of Phonetics

www.elsevier.com/locate/phonetics

Peeling back the layers of time: integrating speech perception on the scales of stimulus time, experiential time, and developmental time

Catherine T. Best[a,b,*]

[a] *Wesleyan University, Department of Psychology, 218 Judd Hall, Middletown, CT 06459, USA*
[b] *Haskins Laboratories, 270 Crown Street, New Haven, CT 06511, USA*

Received 16 September 2002; accepted 8 April 2003

What is the role of temporal integration in the development of speech perception? The answer depends on how one construes "temporal integration." Over which time scale does the integration take place—the microscopic scale of stimulus time, the more extended scale of experiential time with a given language, or the macroscopic scale of developmental time? The answer also depends on what one thinks might be developing in "the development of speech perception." Naturally, the approach to both issues depends on one's theoretical perspective. This commentary will address the three time scales, within the context of selected perspectives on what, exactly, may be developing in speech perception. I begin by summarizing my own and several other key views regarding integration at the stimulus time scale, with special attention to Peter Jusczyk's WRAPSA model (e.g., Jusczyk, 1993, 1997). The invited papers' contributions to the issues are then considered, and possible points of contact between them are suggested. I conclude by raising some questions about developmental change in nonnative versus native speech perception as a reflection of long-term temporal integration.

One view of temporal integration, at the stimulus time scale, follows from a classic assumption that the intake of speech, or any other stimulus type, is a punctate, instant-by-instant affair of taking in disparate stimulus features or cues that must be linked together by processes *within the perceiver*. This view assumes that information distributed across even very brief time spans requires cognitive and/or neural integration. From this vantage point, the research goal is to determine the nature of the processes involved, and to identify how they may differ for diverse aspects of the speech signal (segmental, syllabic, prosodic, etc.) that have varying extents of temporal spread within an utterance.

*Department of Psychology, Wesleyan University, 218 Judd Hall, Middletown, CT 06459, USA. Tel.: +1-203-347-9411.

E-mail address: cbest@mail.wesleyan.edu (C.T. Best).

0095-4470/03/$ - see front matter
doi:10.1016/S0095-4470(03)00033-0

Another view of integration at the stimulus scale can be extrapolated from the motor theory of speech perception (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985, 1989), specifically that temporal integration is intrinsic to the specialized neuro-linguistic module that links speech perception and production. That is, the phonetic properties of the heard utterance are perceived via the neuro-motor commands by which the listener would produce the utterance. Thus, according to the motor theory, just as with the classic view of perception, temporal integrity is found in the listener.

A third view, which I favor along with Carol Fowler (e.g., Best, 1984, 1994, 1995; Fowler, 1986, 1989; Fowler & Dekle, 1991), is based on the ecological theory of perception developed by James and Eleanor Gibson (Gibson, 1966, 1979; Gibson & Gibson, 1955), in which perception is considered to be a dynamic act that extends across the time frame of the events being perceived. At the stimulus time scale, it assumes that speech events are dynamically coordinated gestures of the human vocal tract's articulators. The task of perception is to follow those distal events and extract information about them from the flow of stimulation. Temporal integration is not seen as the result of a process residing solely in the perceiver's mind or brain. Rather, it is the reflection of the active, continuous perceptual tracking of dynamic events that unfold over time. From the ecological perspective, temporal integrity cannot be attributed solely to the perceiver or to the perceived event—it is intrinsic to their interaction over time. Thus, temporal integration also pertains to the larger time scales of the perceiver's development and experience with a specific language. The ecological research goal is to identify the crucial parameters and inter-coordination of articulatory gestures, to determine how perceivers detect those gestures in speech, and to uncover how language experience and development change perceivers' detection of that information.

This leads us back to our earlier question about developmental and experiential time scales: "What is perceptual development the development of?" An early proposal claimed that differential exposure to language-specific acoustic cues during a critical developmental period tunes up certain innate sensori-neural or psycho-acoustic processing mechanisms, such that sensitivities to the experienced acoustic features are maintained or refined, whereas sensitivities to nonexperienced features are lost (e.g., Aslin & Pisoni, 1980; Kuhl, 1993). That is, there is long-term integration of experienced stimuli within the nervous system. An alternative, more cognitive view has burgeoned in the past decade and a half, according to which developmental changes in speech perception result from experience-based establishment of native phonetic categories, or of other linguistic groupings, such as lexical, prosodic, or syntactic patterns (e.g., Jusczyk, 1993, 1994; Lalonde & Werker, 1995; Pegg & Werker, 1997). These representations, or memory traces, are posited to reflect the build-up of prototypes (Kuhl, 1993) or of the traces of a lifetime of experienced exemplars (e.g., Jusczyk, 1997). Again, there is long-term integration of stimulus experience, in this case in the building of mental representations. The motor theory of speech perception, on the other hand, assumes that experience in some way modifies the settings of the specialized speech module (e.g., Liberman & Mattingly, 1989). That is, experiential influences are integrated into the module's functioning.

The ecological view instead posits that perceptual learning is the attunement of skill in active extraction of higher-order invariants that are distributed across dynamic events. Complex events provide a wealth of information to perceivers, often including nested sub-events. For example, consider the temporally layered movements of a dance (sequenced movements of feet, of legs, of whole body; movement across the stage, etc.) or the temporally layered structure in the performance of a musical score (sequences of notes/chords, phrases, movements, etc.). Speech is no exception to complex temporal layering. But note that this does not mean all available information automatically forces itself on the perceiver. Rather, the richness is available but its crucial properties and organization must be discovered over development and through the progression of language experience. Becoming attuned to nested sub-events and their temporal inter-relationships takes developmental and experiential time. As Eleanor Gibson said, perceptual economy—that is, selective attention to crucial properties, and minimal distraction by irrelevant ones—improves through the discovery of ever-higher invariant properties of complex events (Gibson & Pick, 2002; Gibson & Gibson, 1955).

Peter Jusczyk's perspective on stimulus-level temporal integration in infant speech perception is especially relevant here. I will draw from his 1997 book, in particular his WRAPSA model (Word Recognition and Phonemic Structure Acquisition). It is interesting that Jusczyk noted several key beliefs in common with Eleanor Gibson, specifically that children systematically seek out structure in stimulation, extract recurring features, and increase perceptual economy over time by optimizing attention to crucial higher-order invariants. Where he departed from her view was in his assumptions that children accomplish this with the aid of innate child-internal processing biases, and that speech input is subjected to several processing stages: (1) detailed auditory analysis, (2) application of weighting schemes that have emerged from experience with distributional properties, (3) pattern extraction from fluent speech, including word segmentation and clause boundary detection, and (4) comparison of the processed input targets to stored linguistic representations.

Temporal integration plays at least two complementary roles in WRAPSA. On one hand, auditory analysis is posited to occur within very narrow time windows (segment-level) in young infants, due to their limited memory and processing resources. It follows, then, that the integration of information that is distributed across larger stimulus time windows (e.g., across syllables) would initially be rather poor, but would expand with development. This possibility has yet to be directly tested in very young infants, although sensitivity to coarticulatory effects between adjacent syllables is apparently in place by 4 months (Fowler, Best, & McRoberts, 1990).

On the other hand, Jusczyk also posited that young infants are especially sensitive to prosody, and he provided much evidence in support of this. Therein lies a deeper puzzle, however—by definition, prosody incorporates even more broadly extended variations in fundamental frequency, amplitude, and syllable durations. The paradox may disappear, however, if the narrow time window processes and the prosody-tracking processes handle complementary information. This possibility seems quite compatible with WRAPSA, and raises interesting counter-intuitive predictions regarding interactions between the two information types. For example, it suggests that young infants may better perceive phonetic information that is distributed across segments or syllables when the target items appear in a prosodic phrase rather than in isolation. Indeed, Jusczyk and his students have provided evidence supporting exactly that prediction (Mandel, Jusczyk, & Kemler Nelson, 1994).

Christophe and colleagues (Christophe, Gout, Peperkamp, & Morgan, 2003, in this volume) focus on such interactions between prosody and two other temporally distributed stimulus properties, one syntactic and the other lexical. Their findings also support the idea that prosodic information fosters infants' recognition of lexical units and syntactic phrase boundaries. This insight may shed light on Werker and colleagues' recent data (Stager & Werker, 1997; Werker, Cohen, Lloyd, Casasola, & Stager, 1998), which suggest that infants' phonetic representations of newly learned isolated words are quite under-specified prior to about 18 months. Perhaps isolated words do not provide infants with the context for extracting fuller phonetic information because such words are divorced from the prosodic patterns of connected speech, and because isolated word lists are not what infants typically hear. We could hypothesize, then, that more fully specified phonetic representations of words should be evident well before 18 months, so long as the words are embedded within prosodic phrases. This is exactly the context in which infants as young as 14 months have, in fact, displayed detailed phonetic specifications of familiar words (Swingley & Aslin, 2002; see also Fennell & Werker, in press).

Port's (2003) paper in this volume addresses another prosodic feature: rhythmic discreteness in speech, associated with vowel onsets (roughly, P-centers, see Fowler, 1979). Port's research focuses on production in adults, but he argues that such metrical skills are among the earliest acquired in child speech, a notion that mirrors Jusczyk's thoughts about infants' early responsiveness to prosody. In other words, the rhythmic patterns in speech may also aid the child's perception. Comparison of Port's and Christophe et al.'s papers brings to mind the following questions, then: How might the metrical patterns that Port observed in fluent speech contribute to the prosodic effects that Christophe and colleagues found in infants' perception of lexical and syntactic information? Could the perceptual effects be due, in part, to correspondence between rhythmic peaks and crucial locations in syntactic and lexical units? That is, might prosodic peaks aid infants' temporal integration by lining up with the narrow processing windows that Jusczyk proposed for young infants, thus focusing their attention on the correct segment-level phonetic information? This possibility suggests one means by which the metrical patterns that Port observed could be involved in Christophe et al.'s prosodic effects. Prosodic boundaries are usually marked by phrase-final syllable lengthening, especially of the rhyme, and by lengthening of the onset of the following prosodic phrase. Given that perceived beats, or P-centers, in speech are highly associated with vowel onsets, prosodic boundary markers may be perceived as exaggerated intervals between adjacent beats, or as a missing beat.[1]

I conclude with a few broader questions about the developmental and experiential time scales evident in language-specific phonetic learning. Jusczyk, Werker, Polka, Kuhl, and I, among others, have all reported developmental changes in nonnative speech perception that result from native language experience (e.g., Best, McRoberts, LaFleur, & Silver-Isenstadt, 1995; Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Polka, Colantonio, & Sundara, 2001; Werker & Lalonde, 1988). Should these developmental changes be understood as evidence of very long-term temporal integration? Certainly, long-term comparison is needed to learn and recognize a native word across sentences and paragraphs, and across intervals of days, weeks or months, or to maintain topical focus across larger verbal stretches than a sentence. If developmental changes are instantiations of temporal integration, how do they interact with the information that is distributed across the decreasingly shorter time windows of conversations, sentences, phrases (prosodic, syntactic), stress feet, syllables, and segments? If, on the other hand, long-term language experience effects are different in kind from shorter-term temporal integration, what is the crucial difference? And

[1] Thanks are due to an anonymous reviewer for this insightful idea.

either way, how do long-term developmental and experiential effects influence perceptual integrity for shorter-term temporal properties at the stimulus time scale? It remains for future research, with infants as well as adults, to address these questions.

References

Aslin, R. N., & Pisoni, D. B. (1980). Some developmental processes in speech perception. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology perception*, Vol. 2 (pp. 67–96). New York: Academic Press.

Best, C. T. (1984). Discovering messages in the medium: speech and the prelinguistic infant. In H. E. Fitzgerald, B. Lester, & M. Yogman (Eds.), *Advances in pediatric psychology* (pp. 97–145). New York: Plenum.

Best, C. T. (1994). Learning to perceive the sound pattern of English. In C. Rovee-Collier, & L. Lipsitt (Eds.), *Advances in infancy research*, Vol. 8 (pp. 217–304). Hillsdale, NJ: Ablex Publishers.

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange, & J. J. Jenkins (Eds.), *Cross-language speech perception* (pp. 171–204). Timonium, MD: York Press.

Best, C. T., McRoberts, G. W., LaFleur, R., & Silver-Isenstadt, J. (1995). Divergent developmental patterns for infants' perception of two non-native consonant contrasts. *Infant Behavior and Development*, *18*, 339–350.

Christophe, A., Gout, A., Peperkamp, S., & Morgan, J. (2003). Discovering words in the continuous speech stream: The role of prosody. *Journal of Phonetics*, *31*, doi:10.1016/S0095-4470(03)00040-8.

Fennell, C., & Werker, J. F. (in press). Early word learners' ability to access phonetic detail in well-known words. *Language & Speech*.

Fowler, C. A. (1979). Perceptual centers in speech production and perception. *Perception and Psychophysics*, *25*, 375–388.

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, *14*, 3–28.

Fowler, C. A. (1989). Real objects of speech perception: A commentary on Diehl and Kluender. *Ecological Psychology*, *1*(2), 145–160.

Fowler, C. A., Best, C. T., & McRoberts, G. W. (1990). Young infants' perception of liquid coarticulatory influences on following stop consonants. *Perception and Psychophysics*, *48*(6), 559–570.

Fowler, C. A., & Dekle, D. J. (1991). Listening with eye and hand: Cross-modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *17*(3), 816–828.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review*, *62*(1), 32–41.

Gibson, E. J., & Pick, A. (2002). *An ecological approach to perceptual learning and development*. New York: Oxford University Press.

Jusczyk, P. W. (1993). From general to language-specific capacities: The WRAPSA model of how speech perception develops. *Journal of Phonetics*, *21*, 3–28.

Jusczyk, P. W. (1994). Infant speech perception and the development of the mental lexicon. In J. Goodman, & H. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 227–270). Cambridge, MA: MIT Press.

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Jusczyk, P. W., Friederici, A. D., Wessels, J. M. I., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, *32*, 402–420.

Kuhl, P. K. (1993). Innate predispositions and the effects of experience in speech perception: the native language magnet theory. In B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. McNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 259–274). Dordrecht: Kluwer Academic Publishers.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*, 606–608.

Lalonde, C. E., & Werker, J. F. (1995). Cognitive influences on cross-language speech perception in infancy. *Infant Behavior & Development*, *18*, 459–475.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1–36.

Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech perception. *Science*, *243*, 489–494.

Mandel, D. R., Jusczyk, P. W., & Kemler Nelson, D. (1994). Does sentential prosody help infants organize and remember speech information? *Cognition*, *53*, 155–180.

Pegg, J. E., & Werker, J. F. (1997). Adult and infant perception of two English phones. *Journal of the Acoustical Society of America*, *102*, 3742–3753.

Polka, L., Colantonio, C., & Sundara, M. (2001). A cross-language comparison of /d/-/D/perception: Evidence for a new developmental pattern. *Journal of the Acoustical Society of America*, *109*, 2190–2201.

Port, R. F. (2003). Meter and speech. *Journal of Phonetics*, *31*, doi:10.1016/j.wocn.2003.08.001.

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

Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods in the word-form representations of 14-month-olds. *Psychological Science*, *13*, 480–484.

Werker, J. F., Cohen, L. B., Lloyd, V. L., Casasola, M., & Stager, C. L. (1998). Acquisition of word-object associations by 14-month-old infants. *Developmental Psychology*, *34*, 1289–1309.

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, *24*(5), 672–683.