Developmental Science

Developmental Science 11:4 (2008), pp 467–473 DOI: 10.1111/j.1467-7687.2008.00692.x

PAPER

The development of articulatory signatures in children

Latika Singh and Nandini C. Singh

National Brain Research Centre, Manesar, India

Abstract

The ability to perceive and produce sounds at multiple time scales is a skill necessary for the acquisition of language. Unlike speech perception, which develops early in life, the production of speech sounds starts at a few months and continues into late childhood with the development of speech-motor skills. Though there is detailed information available on early phonological development, there is very little information on when various articulatory features achieve adult-like maturity. We use modern spectral analysis to investigate the development of three language features associated with three different timescales in vocal utterances from typically developing children between 4 and 8 years. We make comparisons with adult speech and find age dependence in the appearance of these features. Results suggest that as children get older they exhibit increasingly more power in features associated with shorter time scales, thereby indicating the maturation of fine motor control in speech. Such data from typically developing children could provide milestones of speech production at different timescales. Since impairments in spoken language often provide the first warning signs of a language disorder we suggest that speech production could also be used to probe language disorders.

Introduction

Vocal learning critically depends on three features – the ability to perceive sounds, produce sounds and finally the skill to relate the two (Doupe & Kuhl, 1999). A deficit in any of these three features could result in language impairment. Research has shown that the temporal information in spoken language is at multiple time scales (Rosen, 1992). As a result language learning requires a child to not only perceive sounds at multiple time scales but to also develop speech-motor skills to produce them.

Extensive research on speech perception in infants has shown that the ability to perceive information at both long and short time scales develops early in life (Bertoncini, Bijeljac-Babic, Jusczyk, Kennedy & Mehler, 1988; Jusczyk & Bertoncini, 1988; Jusczyk, Cutler & Redanz, 1993; Kuhl, 2004). On the other hand, the development of spoken language skills continues into late childhood. Characteristic patterns seen in infant speech production indicate that during their first year, infants produce vowel-like sounds at 2–3 months and canonical syllables by 7–8 months (Oller, Eilers, Neal & Cobo-Lewis, 1998). This is followed by the production of sounds like stops and fricatives between 11 months and 2 years (Boysson-Bardies, 1996; Stoel-Gammon, 1992; Kuhl, Williams, Lacerda, Stevens & Lindblom, 1992; Vihman & Velleman, 2000). After children acquire sounds belonging to different phonetic categories, they then need to develop fine articulatory-motor maps wherein they learn to organize these articulatory gestures to produce fluent speech. This occurs between middle and late childhood, possibly during the process of sensori-motor integration. Surprisingly, only a limited number of studies have focused on language refinement in children between 3 and 10 years. These studies suggest that learning to coordinate the various gestures involved in producing speech with appropriate timed events is a difficult task that extends well into childhood (Nittrouer, 1995) and children learn separate aspects of speech production at different rates. Using voice onset and offset times (VOT), such studies have investigated developmental patterns in the production of voiceless and voiced word final-stops (Nittrouer, Estee, Lowenstein & Smith, 2003) and fricatives (Whiteside, Dobbin & Henry, 2003). Their results suggest improved articulatory skills with increasing age. A common observation that has emerged from all these studies is that children continue to refine their organization of articulatory gestures past the age of 7 years. However, most of these studies have focused on small sample sizes and have been restricted to a set of few words. Moreover, due to the tedious procedures involved in acoustic analysis they have focused on examining only a single feature associated with a specific time scale, like syllable durations (around hundreds of milliseconds) or place of articulation (tens of milliseconds). A detailed analysis of when various articulatory features begin to mature in children is clearly lacking. The present study is an effort to provide information on the maturation of articulatory features at both the long and the short time scale, in typically developing children. We develop a technique (Singh &

Address for correspondence: Nandini C. Singh, National Brain Research Centre, NH-8, Nainwal Mode, Manesar – 122 050, India; e-mail: nandini@nbrc.ac.in

Theunissen, 2003) based on spectral analysis wherein we extract spectro-temporal modulations that encode three articulatory signatures namely syllabicity, formant transitions and place of articulation, wherein each signature is associated with a different time scale. Syllabicity that is buried in temporal events at hundreds of milliseconds allows us to study development at the long time scale. At the short time scale, we study formant transitions encoded around 25–40 milliseconds and place of articulation, which is buried around 10–20 milliseconds (Stevens, 1980; Rosen, 1992).

We studied adult speech for the presence of these three articulatory signatures. We obtained speech productions from children between 4 and 8 years and examined them for the presence or absence of each signature in the population. An articulatory feature is designated an articulatory signature if the energy distribution of the spectro-temporal modulations produced by children is the same as that seen in adults. Thus the focus in this study is on determining when various articulatory features become articulatory signatures in children's speech. With advances in technology, speech production as a paradigm is relatively easy to implement since it involves only recording vocal utterances from children, which can later be analyzed.

The paper is organized as follows: In section II we describe the participants and database used in the current study. In section III, the methods of acoustic analysis are discussed. In section IV we present the results and in section V we discuss the conclusions and directions for future research.

Methods

Participants

The speech data were collected in English from 160 school-going children in the age group 4–8 years, with equal numbers of boys and girls. Data were classified into five age groups wherein GI consisted of 20 children of mean age (4.3 years ± 0.2), GII consisted of 32 children of mean age (5.3 years ± 0.3), GIII consisted of 40 children of mean age (6.4 years ± 0.3), GIV consisted of 31 children of mean age (7.4 years ± 0.3), GV consisted of 32 children of mean age (8.2 years ± 0.24) and 16 adults whose mean age was 25 years. None of the speakers had been diagnosed with a speech or language problem and all reported normal hearing. They volunteered to participate in the study with parental consent. This study was also approved by the Human Ethics Committee of the Institute.

Design and procedure

All speech samples were collected using an Acer Laptop computer with a high-quality microphone. The microphone was provided with head fittings, and was kept at an appropriate distance of about 5 cm from the mouth of the speaker. Two tasks were administered, picture naming and phrase repetition. The picture naming task consisted of 20 colored pictures, which included common objects, animals, vehicles (examples: pigeon, bus, square, pen, car, apple, ice cream, circle, television, cup, flower, pencil, triangle, plate, scooter, train). Participants were presented with each picture in turn and asked to name it. If a participant failed to name the picture, then no response was recorded. In the present study, the aim was to get the spoken utterances from the entire participant group rather than scoring whether or not the child recognized the picture or knew the words. For the phrase repetition task, two short phrases were used. They were 'cut the cake' and 'has a cup' and were selected because they were short and simple.

Prior to recording, the tasks were explained to subjects in order to familiarize them with the procedure. However, no specific instructions were given to the subjects regarding the manner of production. All samples were recorded in a quiet room. Recording was done at a sampling rate of 22050 Hz and 16-bit resolution, using standard recording software. After the data collection, each waveform file was manually examined by listening to the recorded speech. Five waveform files that were of very low recording quality were excluded from this study. Amplitude waveforms were normalized to –18 dB using Gold Wave (version 5.10).

Analysis

All the speech productions from each child were combined to form a single speech sample that was approximately 10 minutes in length. This was followed by a calculation of the modulation spectrum for each sample (Singh & Theunissen, 2003). To calculate the modulation spectrum, a spectrogram was first obtained by decomposing the speech sample into an ensemble of narrow-band signals. As seen in Figure 1(a), vocal utterances can be characterized by fluctuations in frequency and time, a visual display of which is offered in a spectrogram. Spectral modulations (ω_x) are energy fluctuations across a frequency spectrum at particular times, and temporal modulations (ω_t) are energy fluctuations at a particular frequency over time. A 2-D Fourier transform of spectrogram in terms of these spectro-temporal modulations gives the modulation spectrum (Singh & Theunissen, 2003). Figure 1(b) shows a typical modulation spectrum of a speech sample from an adult speaker. The central region, which we call the long time scale, is localized between 2 and 10 Hz and carries supra-segmental information. The side lobes, which correspond to the short time scale, are localized between 10 and 100 Hz and encode segmental information. The segmental information can be further divided into two regions, one between 25 and 40 Hz, which encodes slower formant transitions, and a second region between 50 and 100 Hz, which captures amplitude fluctuations arising due to place of articulation in stops

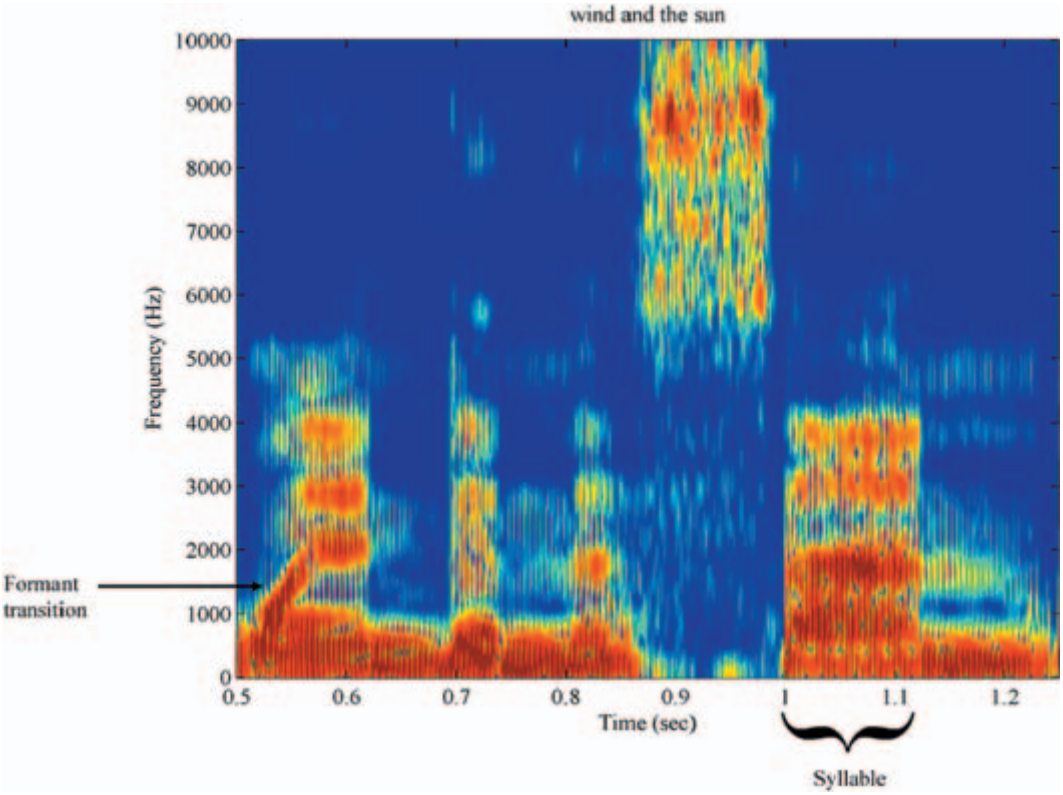

Figure 1(a) *Spectrographic representation of the utterance 'the wind and the sun' as uttered by a female speaker. Various articulatory features at different time scales are highlighted.*

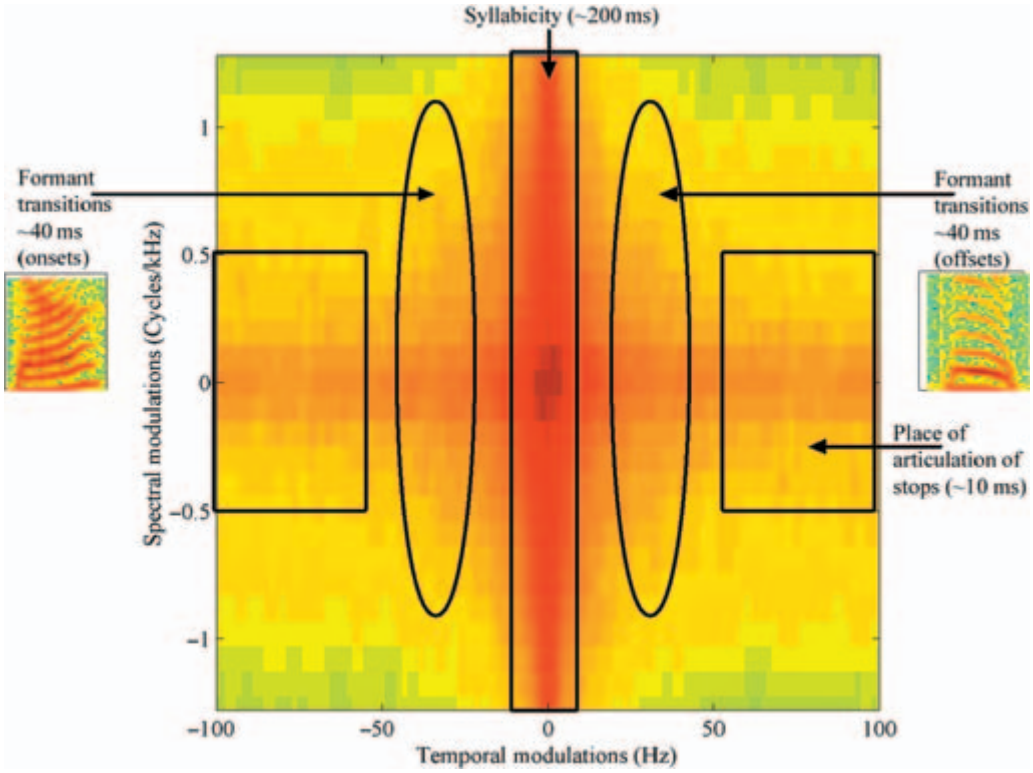

Figure 1(b) *The modulation spectrum of a typical adult speech utterance. Different regions of the modulation spectrum outlined in black encode specific articulatory features as depicted by a rectangle (syllabicity), oval (slower formant transitions) and a square (faster changes like place of articulation of stops/fricatives).*

and fricatives (Figure 1(b)). As we go from 1 to 100 Hz, we approach sounds whose amplitude fluctuations become faster and go from syllabic to vowel-like to plosive-like segments. Thus, based on the temporal scale the modulation spectrum can be classified into three regions:

Syllabicity (~100 to 500 milliseconds) (2–10 Hz)

A syllable is a unit of speech that is made of a syllable nucleus (most often a vowel) with one or more option phones (single sounds or 'phonetic segments'). Syllables are often considered the phonological 'building blocks' of words. They can influence the rhythm of a language, its prosody and its stress pattern. Syllabicity is the pattern of syllable formation in a particular language. For example, the mean syllabic duration for English is 190 ms; 60% of the syllables fall between 106 ms and 260 milliseconds, average rate is ~5 Hz (5 syllables/second) (Takayuki & Greenberg, 1997). So the temporal changes occurring at this rate encode information related to syllables; briefer temporal changes encode sub-syllabic properties like formant transitions.

Formant transitions: (25–40 milliseconds) (40–25 Hz)

Sounds such as fricatives and stops involve obstruction being made in the vocal tract. Although the articulatory gestures that make these obstructions are quite rapid, both closing and opening movements are evident even in spectrograms. Typical formant transition occurs around 40 ms (Tallal, Stark & Mellits, 1985). So, amplitude modulations with transition rates faster than 40 ms encode formant transitions and can be picked up at temporal modulations around 25 Hz. The left quadrant encodes up-sweeps (like formant onsets) whereas the down-sweeps (offsets) are represented in the right quadrant. Since higher formants are weaker, the spectro-temporal features associated with these have less power.

Place of articulation in stops (10–20 milliseconds) (50–100 Hz)

The acoustical properties of a language in different temporal windows can be described with different mechanisms, and the rapidity with which the power spectrum changes for different sounds is one such characteristic mechanism (Stevens, 1980). Since changes occur at different time scales, children learn to produce them at different points in development, and the two constraining factors are the maturity of the vocal system and the ability to process sounds at different time scales. On the one hand, there are the true consonants (p, d, m, s, z) for which a rapid change in spectrum occurs over a time interval in the range of ~20 ms, as the articulatory system moves from a constricted configuration to a vowel, or vice versa. On the other hand, for speech sounds such as vowels and glides, the spectrum changes much more slowly. In short, the property of interest that distinguishes the consonants from vowels and glides is not the shape of the spectrum at any particular instant of time but rather the rapidity with which the spectrum changes, somewhat independently of where in the frequency range the change occurs. Analogous to this, the place of articulation for stop consonants in syllable-initial position and fricatives is determined by acoustic events in the vicinity of the consonant release – probably within 10–20 ms of the release. These are rapid onsets and transient events, and generate high temporal rates. So modulations faster than 20 ms encode such transitions and are located between 50 and 100 Hz.

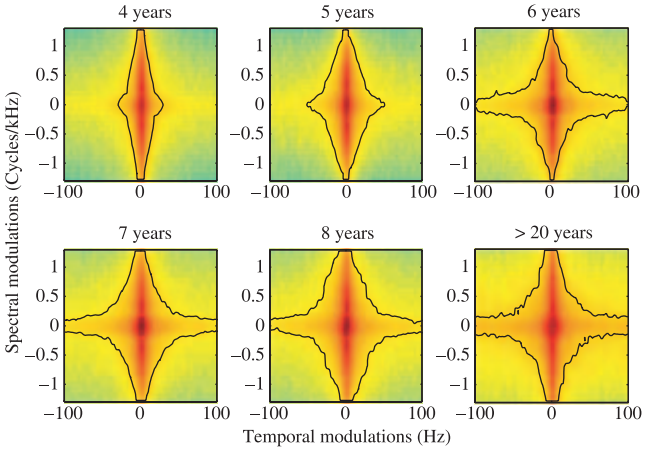

Figure 2 *Representative modulation spectrum of a child from each age group. Spectro-temporal reorganization of sounds along with adult-like intensity in shorter time scales with increasing age is seen.*

Results

Modulation spectra for all 155 children in the age group 4–8 years were obtained. Figure 2 shows a representative modulation spectrum of a typical child belonging to the particular age group. As seen in Figure 2, the modulation spectra differ across ages. We examined the modulation spectrum for each child, for the presence of three articulatory signatures namely syllabicity, formant transitions and place of articulation. As explained earlier, for an articulatory feature to qualify as an articulatory signature the energy distribution of the spectro-temporal modulation profile in children's speech must be similar to that seen in adults.

Figure 3 presents a representative comparison for each articulatory feature across different age groups. As seen in Figure 3(a), the *power* exhibited by children between 4 and 8 years in temporal modulations between 2 and 10 Hz which encodes syllabicity is the same as that seen in adults. Temporal modulations between 25 and 40 Hz encode formant transitions. As seen in Figure 3(b), this feature is evident with adult-like power at around 5 years and thereafter. Adult-like power associated with place of articulation information, encoded between 50 and 100 Hz, is evident at around 6 years.

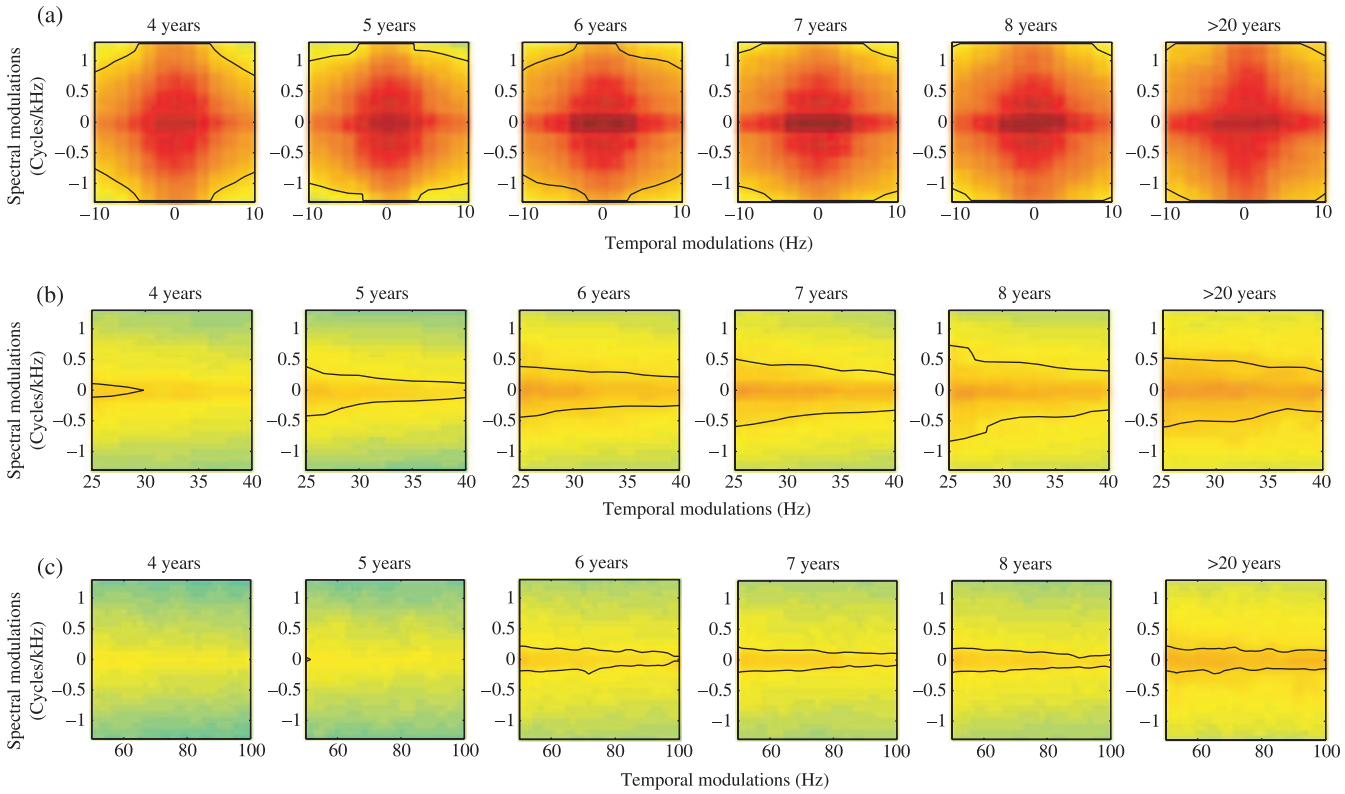

Figure 3 *Representative spectro-temporal energy distribution for each articulatory feature from children belonging to different age groups. (a) The spectro-temporal energy distribution between 2 and 10 Hz which encodes syllabicity, shows that children 4 years old and above exhibit adult-like intensity. (b) The articulatory signature for formant transitions are encoded between 25 and 40 Hz and are evident in speech produced by children 5 years of age and older. Figure 3(c) shows the spectro-temporal energy distribution between 50 and 100 Hz which encodes place of articulation information. Adult-like energy distribution associated with this time scale is evident around 6 years of age.*

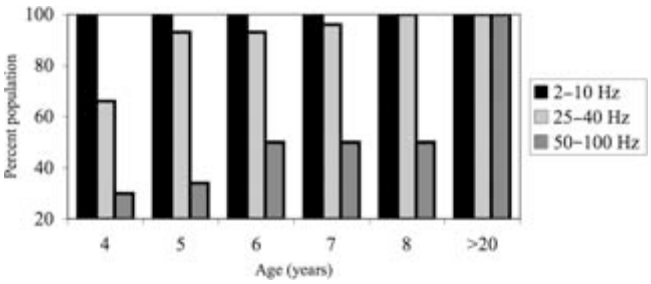

Figure 4 *Bar chart showing the percentage of children exhibiting various time scales as a function of age. As children get older, they exhibit the presence of shorter time scales.*

Since speech development in children is variable, it is now necessary to ascertain when all of these features are evident in the population. For each age group, we counted the number of children that exhibit a particular articulatory signature. We observe a developmental trend in that as children get older they exhibit articulatory signatures associated with shorter time scales (Figure 4), indicating improvement in speech motor skills. If an articulatory signature is evident in the speech produced by more than 70% of the children belonging to that age group we label it as a signature evident in the population. We found that all the children (100%) between 4 and 8 years exhibited adult-like syllabicity (Figure 4). We therefore suggest that this is the first signature acquired and is probably evident in children's speech even before 4 years of age. Signatures associated with formant transitions (25–40 Hz) are evident in 93% of the population at 5 years, though 60% of the children exhibit this signature even at 4 years (Figure 4). Therefore, signatures associated with formant transitions are suggested to be evident in the population at around 5 years. The energy encoded at the shortest time scale, namely between 50 and 100 Hz, that encodes information related to place of articulation, is exhibited by 50% of the 6-year-olds, 30% of the 5-year-olds and in 25% of the 4-year-olds. Surprisingly, however, at 7 and 8 years we do not find an increase in the population exhibiting place of articulation information, which suggests that maturation of speech-motor skills extends beyond 8 years. We therefore conclude that as children get older, they exhibit articulatory features associated with shorter time scales.

Discussion

The present study was carried out to investigate the developmental changes in three articulatory features in language produced by children between early and middle

childhood, namely between 4 and 8 years. The features examined were syllabicity, which is associated with the long time scale; and formant transitions and place of articulation for fricatives and stops, both associated with the short time scale. This period of 4–8 years corresponds to the period of sensori-motor integration when speech production skills mature. As indicated earlier, an articulatory feature becomes an articulatory signature when it is produced by children with the same spectro-temporal energy distribution as that seen in adults. Our results clearly indicate age effects and an improvement in speech-motor skills. As children advance in age they exhibit articulatory signatures that are associated with shorter time scales. A point of interest here is that the articulatory features associated with shorter time scales may have been evident in children's speech even at a younger age but with a much smaller intensity and therefore do not qualify as articulatory signatures.

Children at 4 years exhibit adult-like syllabicity. Though syllabicity as an articulatory feature may have been acquired earlier, it is not yet known when it becomes an articulatory signature. Based on the present study, we can conclude that it is adult-like at 4 years and is evident in 100% of the children we examined and hence is now an articulatory signature. Typical development milestones indicate that speech at 4 years is fairly intelligible. In fact, studies of speech intelligibility have shown that temporal modulations between 2 and 8 Hz are the most crucial for speech comprehension (Drullman & Plomp, 1994). Since syllabicity is encoded between 2 and 10 Hz, this is in agreement with the fact that at 4 years, children's productions are comprehensible. Speech intelligibility studies have also suggested how intelligibility improves with the information from increased amplitude envelope information (Drullman & Plomp, 1994). The articulatory features associated with shorter time scales become articulatory signatures with increase in age. Thus, adult-like formant transitions are evident only at 5 years while articulatory signatures associated with place of articulation information are achieved at 6 years. However, even at 6 years this is evident in only 50% of the population and does not show a significant increase at 7 or 8 years. Thus by 7–8 years, only 50% of children exhibit all the articulatory signatures seen in adult speech. We therefore suggest that children's speech continues to mature well past 8 years, as has also been observed in other studies. A study by Nittrouer *et al.* (2003), carried out with children and adult samples of consonant–vowel–stops, suggested that learning to coordinate the various gestures involved in producing speech with appropriately timed events is a difficult task that extends well into childhood. In another study, which examined place of articulation using VOTs, Whiteside *et al.* (2003) showed that though variability in VOTs reduced with age, children had not reached adult-like maturity even by 7 years, and they suggested that motor speech skills continue to mature as children approach adolescence. Speech production studies carried out by Smith and Kenney (1999) for word and syllable durations in children suggest a number of factors as being responsible for children's speech motor control capabilities maturing/improving with increased age, and they attribute this maturation partly to basic, neurological development (e.g. greater myelination of the nervous system, increases in neural interconnections within the brain, etc.) along with more efficient and economical strategies of articulatory-motor control due to sensori-motor development. Our study suggests that reorganization of energy into sounds at different time scales is probably also part of this sensori-motor development. These results are therefore significant for two reasons. This is the first study that illustrates the spectro-temporal reorganization of speech sounds in children during sensori-motor integration to achieve adult-like speech. As children advance in age they demonstrate better speech motor skills, producing sounds at short time scales in a more adult-like manner. Our study also shows for the first time that this might be achieved by reorganizing energy in different spectro-temporal modulations.

Language development requires speech perception, speech production and the ability to relate them both. Psychophysical studies with infants have strongly suggested that the ability to perceive sounds at different scales is in place from a very young age (Werker & Tees, 1999; Jusczyk & Bertoncini, 1988; Bertoncini *et al.*, 1988; Eimas, 1985). Research has shown that infants at 3–6 months can discriminate languages based on syllabic rhythm (Jusczyk *et al.*, 1993) which is of the order of 100–200 ms and by 10–12 months can distinguish between [b] and [d], which is of the order of 10–20 milliseconds (Werker & Tees, 2005). However speech production, which is strongly dependent on motor maturation, develops and matures over a much longer period. In this context it is therefore interesting to note that though infants can *perceive* short time changes of the order of tens of milliseconds in auditory stimuli with the same precision as adults before they are 1 year old, they are able to *produce* the same sounds with adult-like maturity at only 5–6 years. We suspect that at around 4 years, when children start going to school, as speech motor skills mature it is the ability to relate perception and production that is refined and practiced to achieve adult-like language skills.

Clinically, speech *productions* are often the earliest indicators of language impairment. Studies from Tallal *et al.* (1985) have shown that dysphasic children exhibit deficits in both speech perception and speech production. We therefore propose that speech production could also be used to probe language disorders. An inability to produce language features at either the short or the long time scale could provide insights into the nature of the impairment. Given the incidence of increase in language disorders among school-going children, such analysis could assist clinicians to develop better diagnostic criteria to differentiate between normal and disordered speech in child populations. Studies from Tallal *et al.*

(1985) have shown that dysphasic children exhibit deficits in both speech perception and production. A large body of research has also shown that children with specific language impairment (SLI) (Tallal & Gaab, 2006) and dyslexia (Goswami, 2002) exhibit impairments in processing brief, rapidly changing stimuli around tens of milliseconds. Based on these studies we hypothesize that children with SLI and dyslexia may exhibit deficits in the production of sounds associated with formant transitions and place of articulation for stops. On the other hand, children with developmental disorders like autism show deficits in the perception/production of prosody encoded around hundreds of milliseconds and may exhibit deficits in the production of syllabic rhythm. We therefore suggest that such analysis could also be explored to study language disorders.

Finally, this study demonstrates a method that could be used to track the development of a sound system for language. Abnormal development of this sound system could provide early indications of children at risk for possible speech and language disorders.

Acknowledgements

This research was funded by the National Brain Research Centre and a research grant from the Department of Information Technology, Government of India. We also wish to thank all the children who participated in the study.

References

Bertoncini, J., Bijeljac-Babic, R., Jusczyk, P.W., Kennedy, L., & Mehler, J. (1988). An investigation of young infants' perceptual representations of speech sounds. *Journal of Experimental Psychology: General*, **117**, 21–33.

Boysson-Bardies de, B. (1996). *Comment la parole vient aux enfants*. Paris: Odile Jacob.

Doupe, A.J., & Kuhl, P.K. (1999). Birdsong and human speech. *Annual Review of Neuroscience*, **22**, 567–631.

Drullman, R., & Plomp, R. (1994). Effect of temporal envelope smearing on speech reception. *Journal of the Acoustical Society of America*, **95**, 1053–1064.

Eimas, P.D. (1985). The perception of speech in early infancy. *Scientific American*, **252**, 46–52.

Goswami, U. (2002). Amplitude envelope onsets and developmental dyslexia: a new hypothesis. *Proceedings of the National Academy of Sciences USA*, **99**, 10911–10916.

Jusczyk, P.W., & Bertoncini, J. (1988). Viewing the development of speech perception as innately guided learning process. *Language and Speech*, **31**, 217–238.

Jusczyk, P.W., Cutler, A., & Redanz, N.J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, **64**, 675–687.

Kuhl, P.K. (2004). Early language acquisition: cracking the speech code. *Nature Neuroscience Review*, **5**, 831–843.

Kuhl, P., Williams, K.A., Lacerda, F., Stevens, K.N., & Lindblom, B.L. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, **255** (5044), 606–608.

Nittrouer, S. (1995). Children learn separate aspects of speech production at different rates: evidence from spectral moments, *Journal of the Acoustical Society of America*, **97** (1), 520–530.

Nittrouer, S., Estee, S., Lowenstein, J.H., & Smith, J. (2003). The emergence of mature gestural patterns in the production of voiced and voiceless word-final stops. *Journal of the Acoustical Society of America*, **117** (1), 351–364.

Oller, D.K., Eilers, R.E., Neal, A.R., & Cobo-Lewis, A.B. (1998). Late onset canonical babbling: a possible early marker of abnormal development. *American Journal on Mental Retardation*, **103** (3), 249–263.

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philosophical Transactions: Biological Sciences*, **336**, 367–373.

Singh, N.C., & Theunissen, F.E. (2003). Modulation spectra of natural sounds and ethological theories of auditory processing. *Journal of the Acoustical Society of America*, **114**, 3394–3411.

Smith, L.B., & Kenney, M.K. (1999). A longitudinal study of the development of temporal properties of speech production: data from 4 children. *Phonetica*, **56**, 73–102.

Stevens, K.N. (1980). Acoustic correlates of some phonetic categories. *Journal of the Acoustical Society of America*, **68** (3), 836–842.

Stoel-Gammon, C. (1992). Pre-linguistic vocal development: measurement and predictions. In C.A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 439–456). Timmonium, MD: York Press.

Takayuki, A., & Greenberg, S. (1997). The temporal properties of spoken Japanese are similar to those of English. *Proceedings of Eurospeech, Rhodes, Greece* (pp. 1011–1014).

Tallal, P., & Gaab, N. (2006). Dynamic auditory processing, musical experience and language development. *Trends in Neurosciences*, **29** (7), 382–390.

Tallal, P., Stark, R.E., & Mellits, E.D. (1985). Identification of language-impaired children on the basis of rapid perception and production skills. *Brain and Language*, **25**, 314–322.

Vihman, M.M., & Velleman, S.L. (2000). The construction of a first phonology. *Phonetica*, **57**, 255–266.

Werker, F. Janet, & Tees, Richard C. (1999). Influences on infant speech processing: toward a new synthesis. *Annual Review of Psychology*, **50**, 509–535.

Werker, F. Janet, & Tees, Richard C. (2005). Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Developmental Psychobiology*, **46**, 233–251.

Whiteside, S.P., Dobbin, R., & Henry, L. (2003). Patterns of variability in voice onset time: a developmental study of speech motor skills in humans. *Neuroscience Letters*, **347**, 29–32.

Received: 2 November 2006
Accepted: 14 June 2007