Developmental and cross-linguistic variation in the infant vowel space: The case of Canadian English and Canadian French

Susan Rvachew, Karen Mattock, and Linda Polka
McGill University, Montreal, Quebec H3G 1A8, Canada

Lucie Ménard
Université du Québec à Montréali

(Received 13 September 2005; revised 11 July 2006; accepted 12 July 2006)

This article describes the results of two experiments. Experiment 1 was a cross-sectional study designed to explore developmental and cross-linguistic variation in the vowel space of 10- to 18-month-old infants, exposed to either Canadian English or Canadian French. Acoustic parameters of the infant vowel space were described (specifically the mean and standard deviation of the first and second formant frequencies) and then used to derive the grave, acute, compact, and diffuse features of the vowel space across age. A decline in mean $F1$ with age for French-learning infants and a decline in mean $F2$ with age for English-learning infants was observed. A developmental expansion of the vowel space into the high-front and high-back regions was also evident. In experiment 2, the Variable Linear Articulatory Model was used to model the infant vowel space taking into consideration vocal tract size and morphology. Two simulations were performed, one with full range of movement for all articulatory paramenters, and the other for movement of jaw and lip parameters only. These simulated vowel spaces were used to aid in the interpretation of the developmental changes and cross-linguistic influences on vowel production in experiment 1. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2266460]

PACS number(s): 43.70.Ep, 43.70.Fq, 43.70.Kv [BHS] Pages: 1–XXXX

I. INTRODUCTION

A. Rationale

The acoustic characteristics of vowels produced by English-learning infants have been described in a number of prior studies (Buhr, 1980; Gilbert *et al.*, 1997; Kent and Murray, 1982; Robb *et al.*, 1997; Rvachew *et al.* 1996; Sussman *et al.* 1999; 1996). These studies have revealed a strong preference for central vowels, with very little developmental change in the location of the center of the vowel space, and a very gradual expansion of the range of vowels produced with age. These characteristics of infant vowels are interpreted as reflecting the limitations imposed by the structure of the infant's vocal tract and immature speech motor control. The process by which the young child overcomes these physiological limitations to acquire the vowel system of the ambient language is not well understood however. The purpose of this study was to shed some light on this developmental process by shifting the focus from the universal characteristics of the infant vowel space to individual differences in infant vowel production. Specifically, we describe the vowels produced by a relatively large number of infants drawn from a broad age range and two language backgrounds.

B. Background

1. Acoustic characteristics of infant vowels

Kent and Murray (1982) measured the formant frequencies of vocalic utterances produced by 21 English-learning infants aged 3, 6, and 9 months. Mean first formant ($F1$) and second formant ($F2$) values remained relatively stable across the three age groups of infants, (approximately 900–1000 Hz for $F1$ and 3000 Hz for $F2$). However, the range of first and second formant frequencies progressively increased with age, indicating an expansion of the vowel space. A preference for midfront or central vowels was observed throughout the 3-to 9-month age range. Buhr (1980) reported similar findings for a single infant who demonstrated a gradual growth of the vowel space along the $F1-F2$ dimensions between 16 and 64 weeks of age. Reshaping of the vowel space was also observed with the acute region becoming more defined at an earlier age than the grave corner of the vowel space.

Rvachew *et al.* (1996) described the vowels produced by nine infants, followed longitudinally from 6 to 18 months of age. The mean and standard deviation of first formant frequencies were stable throughout this period. A large and steady increase in the range of $F2$ values was observed during the period of the study. The mean $F1$ observed by Rvachew *et al.* (1996) was similar to that reported by Kent and Murray (1982) for younger infants but the mean $F2$ was considerably lower at approximately 2400 Hz. A small reduction of the mean $F2$ was observed during the latter half of the observation period.

Robb *et al.* (1997) described the vowels produced by 20 children aged 4 to 25 months in a cross-sectional study. Contrary to expectations, no decrease with age in mean $F2$ or $F1$ was apparent. In a longitudinal study of four infants between the ages of 15 and 36 months, Gilbert *et al.* (1997) did find a significant lowering of $F1$ and $F2$, but only between 24 and 36 months.

2. Role of physiological limitations

Several factors can explain the observed acoustic shifts in vowel production across the infancy period: anatomical growth, motor control development, auditory (peripheral and central) abilities, and other cognitive factors. The structure of the human vocal tract is obviously an important determinant of the acoustic characteristics of speech sounds. Studies using magnetic resonance imaging (MRI) confirm the long-standing impression that vocal tract development does not involve a simple linear increase in vocal tract length (Fitch and Giedd, 1999). Major developmental changes in vocal tract structure that occur shortly after birth include the descent of the larynx, lengthening of the pharyngeal cavity, and a sharper angle between the oral and pharyngeal cavities (Kent and Vorperian 1995). Kent *et al.* (1999) described the growth of the supralaryngeal vocal tract in one infant who received repeated MRI scans between birth and 30 months of age. They found that changes in the size of vocal tract structures were generally coordinated, even during growth spurts at 1 and 4 months of age and 12 and 15 months of age. Most of the increase in vocal tract length in the infants' first year could be explained by the descent of the larynx and tongue, whereas the lengthening of the hard palate made a greater contribution to vocal tract growth during the second year of life.

The influence of the morphology of the vocal tract on the acoustic characteristics of infant vowels has been investigated in studies in which the Variable Linear Articulatory Model was used to synthesize vowels that would be produced by vocal tracts having the dimensions observed for different age groups, specifically a 4-week-old infant, 2-, 4-, 8-, and 12-years-old children, a 16-years-old adolescent, and a 21-years-old adult male (Ménard *et al.*, 2002; 2004). Listener judgments of the resulting vowels indicated that the infant's vocal tract anatomy does not prevent the production of the full range of vowels used in the ambient language. At the same time, infant vocal tract anatomy does at least partly explain infant production preferences: When the maximal vowel space is plotted for the infant and adult vocal tracts, a larger portion of the infant vowel space corresponds to vowels that would be perceived to be low or front vowels, when compared to the adult vowel space. It is also important to note that while it is possible to produce vowels with an infant vocal tract that are perceptually equivalent to adult vowel categories, in many cases the infant would need to employ different articulatory gestures than the adult to achieve the same perceptual outcome.

The finding that the full range of vowel contrasts can be produced with the modeled infant vocal tract assumes adult-like levels of motor control, which is obviously not the case in natural speech. Green *et al.* (2000) measured temporal and spatial coupling of upper lip, lower lip, and jaw movements, during the production of [baba], [mama], and [papa], in 1-, 2-, and 6-year-old children and adults. Jaw movements were dominant, although poorly controlled with respect to force, in the 1-year-old children. By age 2, lip movements were more integrated with the jaw movement. Between ages 2 and 6 years, progressive differentiation of the rigid coupling of upper and lower lip movements was observed. The comparison of movement patterns for 6-years-old children and adults indicated continual refinements in movement control and coordination. Green *et al.* (2002) confirmed the developmental pattern of increased integration of lip movement control into a previously stabilized pattern of jaw movements between 1 and 2 years of age. These data are consistent with the prediction that the development of speech production involves an initial dominance of the "mandibular frame" followed by a progressive differentiation of articulator movements. However, the limitations imposed by immature speech motor control on the development of the infant's speech production abilities do not preclude a role for the auditory environment in shaping that nature of the infant's vocalizations.

3. Role of the auditory environment

Auditory input is clearly critical to the normal development of speech, right from birth. The canonical babbling stage is delayed or never achieved by infants with sensory-neural hearing impairment (see Oller, 2000 for a review of this literature) because hearing impairment interferes with the child's access to both self-produced and other-produced speech (Koopmans-van Beinum *et al.*, 2001). The specific phonetic content of adult speech may shape infant speech production patterns. Kuhl and Meltzoff (1996) manipulated the phonetic content of speech input to the infant in the laboratory by presenting one of three point vowels to different groups of infants aged 12, 16, or 20 weeks. Infants of all ages shifted the acoustic characteristics of their vowels toward the modeled vowel category.

Another strategy for examining the role of speech input is to study cross-linguistic variation in speech production. de Boysson-Bardies, *et al.* (1989) described the acoustic characteristics of the vowel space of 20 10-month-old infants being raised in monolingual French-, English-, Algerian-, and Cantonese-speaking families. They found support for the influence of the ambient language environment on the vowel formants, with variation in mean $F1$ and $F2$ frequencies being greater between language groups than within language groups. English-learning infants' mean $F2$ values were slightly higher than French-learning infants' mean $F2$, but mean $F1$ values were similar for the English- and French-learning infants. Their data suggest that there are systematic and language-specific differences in the articulatory movements produced by infants from different language backgrounds during the first year of life. However this study described only a single age group and a replication has not been published.

The purpose of the current study was to replicate these findings with infant learners of Canadian English (CE) or Canadian French (CF). Recent studies of the adult vowel productions of these languages indicate that CF and CE vowels are characterized by significant acoustic-phonetic differences even where there is phonological overlap (Escudero and Polka, 2003; LaCharite and Paradis, 1997; Martin, 2002). Specifically, the CF /i/ is more diffuse relative to CE /i/, and the CF /u/ is more grave relative to CE /u/. The CF /a/ is slightly less compact than CE /a/. The most acute vowel in CE is [æ], a vowel that is produced allophonically but not phonemically in CF. The acute corner of the CF vowel space appears to be less acute in comparison with CE. These data on vowels produced by adult speakers do not lead to specific predictions about the potential differences between the vowel spaces produced by infant learners of CE and CF because of the differences in the procedures used to obtain and describe speech samples produced by adult and infant talkers. None the less, the fact that there are significant differences in the acoustic-phonetic characteristics of the adult CE and CF vowel spaces supports the hypothesis that there may be cross-linguistic differences in the acoustic characteristics of vowels produced by infants who are exposed to one of these languages.

C. The current studies

The purpose of experiment 1 was to systematically examine developmental changes and cross-linguistic differences in the first and second formant frequencies of vowels. In this cross-sectional study, we recorded speech samples from 23 infants exposed to Canadian French and 20 infants exposed to Canadian English, aged between 10 and 18 months. Acoustic analyses were used to describe the frequency locations of the center and the corners of each infant's vowel space. Although phonetic transcriptions facilitate a direct comparison of infant and adult phonetic repertoires, this type of analysis was avoided. Oller (2000) has questioned the validity of phonetic transcription for the description of infant speech on a number of grounds, three of which are particularly relevant to this investigation. First, phonetic transcriptions of infant vowels are notoriously unreliable, especially for the identification of back vowels (e.g., Davis and MacNeilage 1995). Second, phonetic transcriptions are subject to listener biases that are particularly acute when listening to non-native speech sounds. Third, phonetic descriptions of infant speech imply, unrealistically, that infant vocalizations are composed of the same articulatory features that characterize adult produced phonemes. As Ménard *et al.* (2002, 2004) explained, listener perceptions of infant speech do not reliably point to the underlying articulatory gestures that produced the perceived vowel. Thus, for this study we describe vowels in terms of raw acoustic parameters, specifically the mean and standard deviation of the $F1$ and $F2$, and in terms of features that are simple linear combinations of the raw acoustic values, namely the acute-grave and compact-diffuse features. These features may be more closely associated with the goals of vowel production than the raw acoustic values, which vary significantly as a function of vocal tract size and shape. Other researchers (e.g., Kuhl *et al.*, 1997) have used these parameters to describe vowel production.

In experiment 2, vowel spaces were modeled on the basis of a simulation of the infant vocal tract at 6, 12, and 18 months of age. The vowel space that would be produced by these vocal tracts was derived in order to aid in the interpretation of the data obtained in experiment 1. This modeling study offered an opportunity to study the sole effect of vocal tract growth on acoustic data.

On the basis of the data reported by de Boysson-Bardies *et al.* (1989), we expected cross-linguistic differences in the infant's vowel productions. Specifically, a progressive divergence of the center of the vowel space for French and English infants, particularly with respect to the $F2$ dimension, was predicted. Changes in the first and second formant frequencies, for either language group, that are greater than would be predicted from simple growth of the vocal tract (as indicated by experiment 2) would lend further support to the hypothesis that the phonetic content of adult speech input has an influence on infant speech output during the first 18 months of life. Language-general changes in the vowel space were also expected, especially with respect to the overall size of the vowel space.

II. EXPERIMENT 1

A. Method

1. Participants

Forty-three typically developing infants from predominately middle-class families were recruited from birth registries for the Montréal region. Each infant was no younger than 300 days and no older than 570 days. All infants were reportedly born between 38 and 42 weeks gestation following uncomplicated pregnancies, with no known history of ear infections or hearing impairment, and were healthy on the day of testing. A parent questionnaire about language use in the home (e.g., by parent, siblings, television, radio), and in the speech directed to their infant from others (e.g., grandparents, babysitter, daycare worker) confirmed that 23 infants were being raised by monolingual CF speaking families, and 20 infants were being raised by monolingual CE speaking families. Thirty-two of the 43 infants passed several audiological screenings (tympanometry and otoacoustic emissions) performed by an audiologist beginning at 2–3 months of age (these infants were initially recruited for another study in our lab). The remaining infants passed a tympanometric screening on the day of the speech sample recording.

a. Speech sample recordings. Samples of the infants' vocalizations were recorded during a play session between mother and infant, either in a sound proof booth in the laboratory or in the infant's home. Mothers were instructed to interact with their infant in the usual manner using a set of quiet toys. Recording sessions continued until the infant produced 60 utterances perceived to meet the utterance selection criteria (described below), or until 30 minute had lapsed, whichever came first. The speech samples were obtained using a portable DAT recorder and a Sennheiser microphone affixed to the infant's clothing at the shoulder. Following recording, all speech utterances were digitized at 22 050 Hz using Time Frequency Response software (AVAAZ) installed on IBM PC hardware equipped with a Creative Labs Live Drive.

b. Acoustic analysis. Each utterance was assigned an Infraphonological code (i.e., canonical syllable, fully resonant vowel, quasiresonant vowel, marginal syllable, squeal, raspberry or growl, using the criteria described in detail by Oller, 1986). Isolated vowels and vowels contained within canonical syllables were selected for formant analysis if vowel or syllable duration was less than 500 ms, and if the vowel had normal phonation, full resonance, and at least two measurable formants. These utterance types comprised 25% of the sample. The remaining utterances (28% marginal syllables, 47% "other" including quasiresonant vowels, squeals, growls, and raspberries) were not submitted to formant analysis. Seven vowels were discarded from the data set because either $F1$ ($n=2$) or $F2$ ($n=5$) was three standard deviations greater than the mean value. A total of 1190 utterances (665 French, 525 English) met the criteria. Vowel formant analyses were performed blind to the age and language background of the infant. To determine $F1$ and $F2$ frequencies, a 20 ms segment at the middle of the steady state portion was submitted to linear predictive coding (LPC) autocorrelation analysis with a window size of 256 points, 50% overlap, 98% preemphasis, Hanning window, and model order of 12. Model order was increased or decreased accordingly to obtain reliable measurements of some vowels where the formants were difficult to measure. Formant locations for all vowels were confirmed with narrowband short-time FFT spectrograms (512 points). The Peterson and Barney (1952) norms were also referred to in order to confirm that the obtained frequency values roughly approximated the expected relationship between formants, given the perceived quality of the vowel (e.g., a vowel sounding to be /u/-like would be expected to yield $F1$ and $F2$ values that were close in frequency). Replicate acoustic analysis of 299 vowels (25% of full sample) were conducted by a second individual trained in speech acoustics who was blind to the age, language background of the infants, as well as the measurements obtained by the first coder. Vowels were reanalyzed using the same measurement parameters as the first coder (see above). Intraclass correlations between the independently identified formant frequencies were 0.96 and 0.94 for $F1$ and $F2$ respectively. All first and second formant frequencies in hertz were converted to the mel scale (Stevens *et al.* 1937) using the formula

$$F_{\text{mels}} = (1127.010\,481)\ln\left(1 + \frac{F_{\text{hertz}}}{700}\right).$$

c. Statistical analyses. Each infant's vowel space was described using the following eight summary statistics, all expressed in mels: (1) $MF1$—mean of the first formant frequencies; (2) SD $F1$—standard deviation of the first formant frequencies; (3) $MF2$—mean of the second formant frequencies; (4) SD $F2$—standard deviation of the second formant frequencies; (5) *Grave*—minimum value of $(F1+F2)/2$; (6) *Acute*—maximum value of $(F1+F2)/2$; (7) *Compact*—minimum value of $F2-F1$; and (8) *Diffuse*—maximum value of $F2-F1$. The extraction of these summary statistics from an infant's vowel space is illustrated in Fig. 1. The figure shows $F1$ and $F2$ coordinates for each vowel produced by the infant. Superimposed are two bars that indicate the location of the center vowel and the standard deviation of the first and second formant frequencies as a measure of dispersion of formant values around the center vowel. Arrows on the figure indicate the vowels that represent the most *grave, acute, compact* and *diffuse* values in the vowel space.

Regression analysis was used to examine the main effect of language group, the main effect of infant age, and the

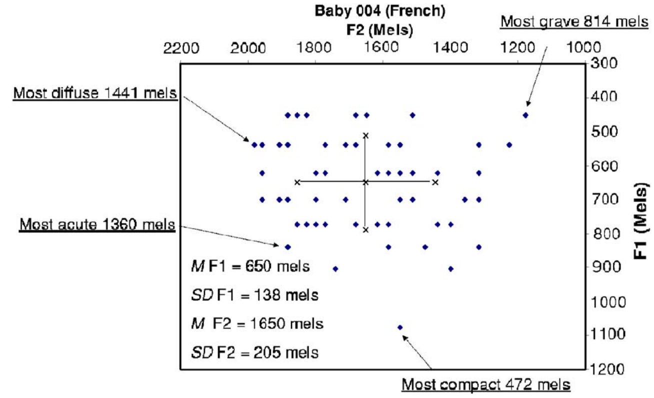

FIG. 1. The vowel space of one French infant. $F1$ and $F2$ coordinates of each vowel produced by the infant are plotted. The bars represent mean and standard deviation of the center vowel. The most grave, acute, compact, and diffuse vowels are indicated by arrows.

interaction of language and age on each summary statistic. These analyses revealed interaction effects for many of the summary statistics and consequently, simple regression analyses are reported for the effect of age on each summary statistic, independently for each language group.

B. Results

1. Parameters

Figure 2 (top left) shows a significant decline in $MF1$ from 962 to 730 mels for the French group [$B=-0.86$; SE $=0.30$; $F(1,21)=8.02$, $p=0.01$]. The smaller decline from 913 to 814 mels for the English group was not statistically significant [$B=-0.37$; SE$=0.27$; $F(1,18)=1.89$, $p=0.19$]. Figure 2 (top right) illustrates a small decline in SD $F1$ that was not significant for the French [$B=-0.10$; SE$=0.09$; $F(1,21)=1.39$, $p=0.25$] or the English [$B=-0.05$; SE$=0.08$; $F(1,18)=0.41$, $p=0.53$] group. Figure 2 (bottom left) depicts a significant decline in $MF2$, from 1714 to 1523 mels, for the English group [$B=-0.71$; SE$=0.24$; $F(1,18)=8.64$, $p=0.01$]. The much smaller decline for the French group, from 1667 to 1636 mels, was not statistically significant [$B=-0.11$; SE$=0.26$; $F(1,21)=0.18$, $p=0.68$]. Figure 2 (bottom right) illustrates a significant increase in SD $F2$, from 130 to 245 mels, for the English group [$B=0.43$; SE$=0.13$; $F(1,18)=10.91$, $p=0.00$]. The SD $F2$ for the French group was relatively stable throughout the age range, with a non-significant increase from 157 to 175 mels [$B=0.07$; SE $=0.14$; $F(1,21)=0.23$, $p=0.63$].

2. Features

Figure 3 (top left) shows an increase in the maximum value of the *diffuse* feature for both groups, specifically from 1101 to 1423 for the French group [$B=1.19$; SE$=0.58$; $F(1,21)=4.19$, $p=0.05$] and from 1115 to 1311 for the English group [$B=0.73$; SE$=0.59$; $F(1,18)=1.53$, $p=0.23$]. These data suggest a trend toward expansion of the vowel space into the acoustic area associated with widely spaced first and second formant frequencies, at least for the French group.

Figure 3 (top right) depicts a decrease in the *grave* fea-

PROOF COPY 023610JAS

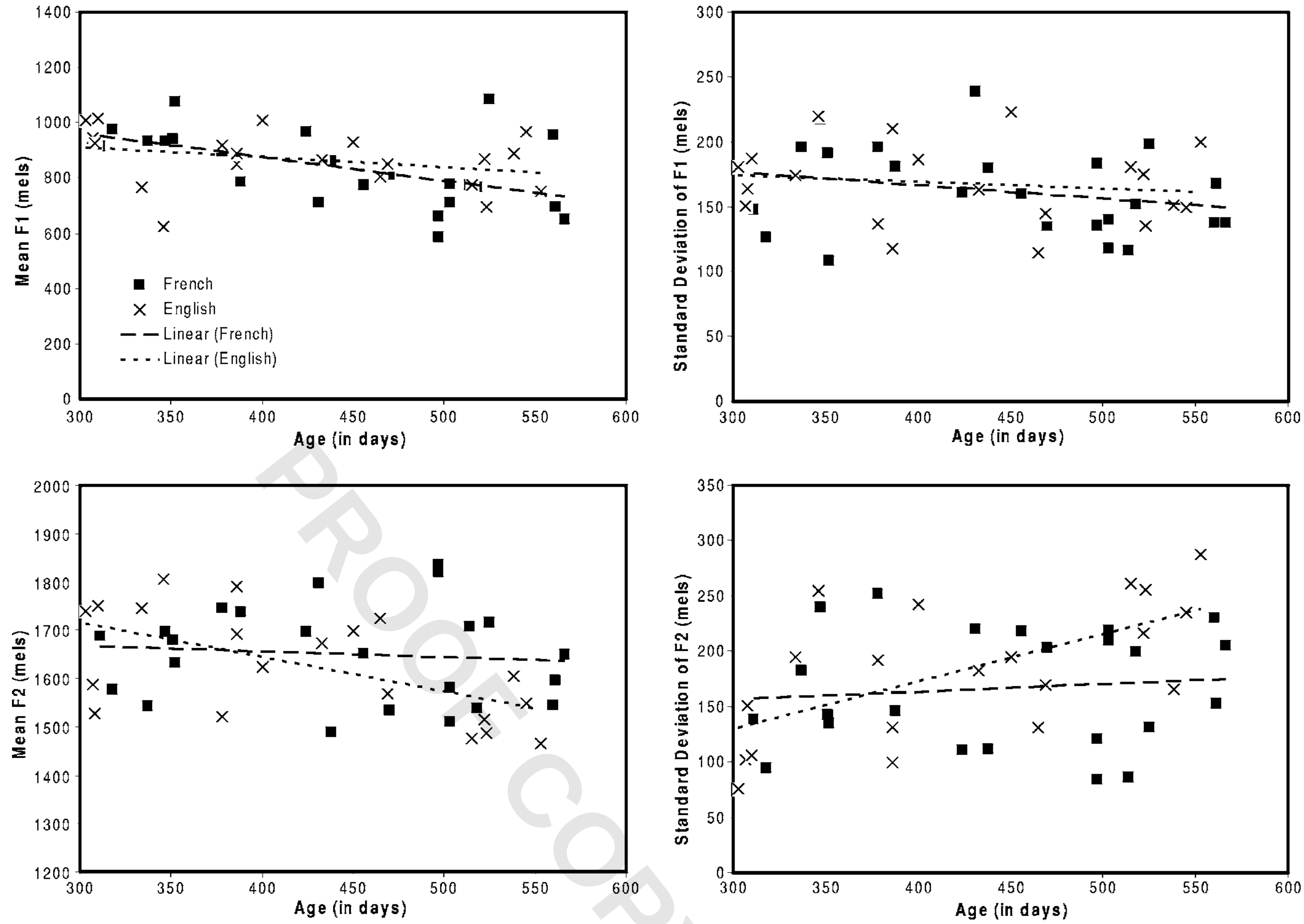

FIG. 2. Speech sample mean of each of four parameters (in mels) plotted for each infant as a function of age and language group, specifically M $F1$ (top left), SD $F1$ (top right), M $F2$ (bottom left), and SD $F2$ (bottom right).

ture that was from 1056 to 907 mels for the French group [$B=-0.55$; SE=0.27; $F(1,21)=4.27$, $p=0.05$] and from 1101 to 843 mels for the English group [$B=-0.96$; SE =0.29; $F(1,18)=10.67$, $p=0.00$]. These values indicate an age-related expansion of the vowel space into the acoustic area associated with low frequency and closely spaced first and second formant frequencies.

Figure 3 (bottom left) illustrates a significant decrease in the *acute* feature from 1545 to 1384 mels for the French group [$B=-0.60$; SE=0.19; $F(1,21)=9.65$, $p=0.01$] and a smaller decrease from 1516 to 1420 mels for the English group [$B=0.36$; SE=0.18; $F(1,18)=4.10$, $p=0.06$]. The decreasing values highlight an age-related compression of the vowel space in the acoustic area associated with relatively high and closely spaced first and second formant frequencies, at least for the French group.

Figure 3 (bottom right) suggests an interaction of age and language group for the *compact* feature. However, the increase from 334 to 440 for the French group was not significant [$B=0.39$; SE=0.55; $F(1,21)=0.51$, $p=0.48$]; the decline from 434 to 304 for the English group was not significant either [$B=-0.48$; SE=0.30; $F(1,18)=2.55$, $p=0.128$].

C. Discussion

Acoustic analyses of vowels produced by infants aged approximately 10 to 18 months indicated the presence of developmental changes that were common to both language groups as well as some significant differences across language groups, as illustrated in Fig. 4. Cross-linguistic variation was apparent in the frequency location of the center of the infants' vowel spaces. Specifically, the French-learning infants demonstrated a significant decline with age in the $MF1$, whereas the English-learning infants produced a significant decline with age in the $MF2$. The English-learning infants demonstrated a significant increase in the dispersion of second formant frequencies as age increased. The French-learning infants did not produce a reliable age-related change in SD $F2$ with age. Neither group showed age-related changes in SD $F1$.

Both groups showed a developmental expansion of the size of the vowel space along the diffuse-grave dimension although expansion into the diffuse corner was greater for the French group and expansion into the grave corner was greater for the English group. These findings suggest a developmental expansion into the areas of the vowel space traditionally associated with tongue retraction and advancement in adult articulation. Compression of the vowel space in the acute corner, particularly marked for the French group, suggests less extreme jaw opening gestures with age.

The distribution of the infants' vowels within this global vowel production space appears to differ across the language groups. At the same time, English-learning and French-

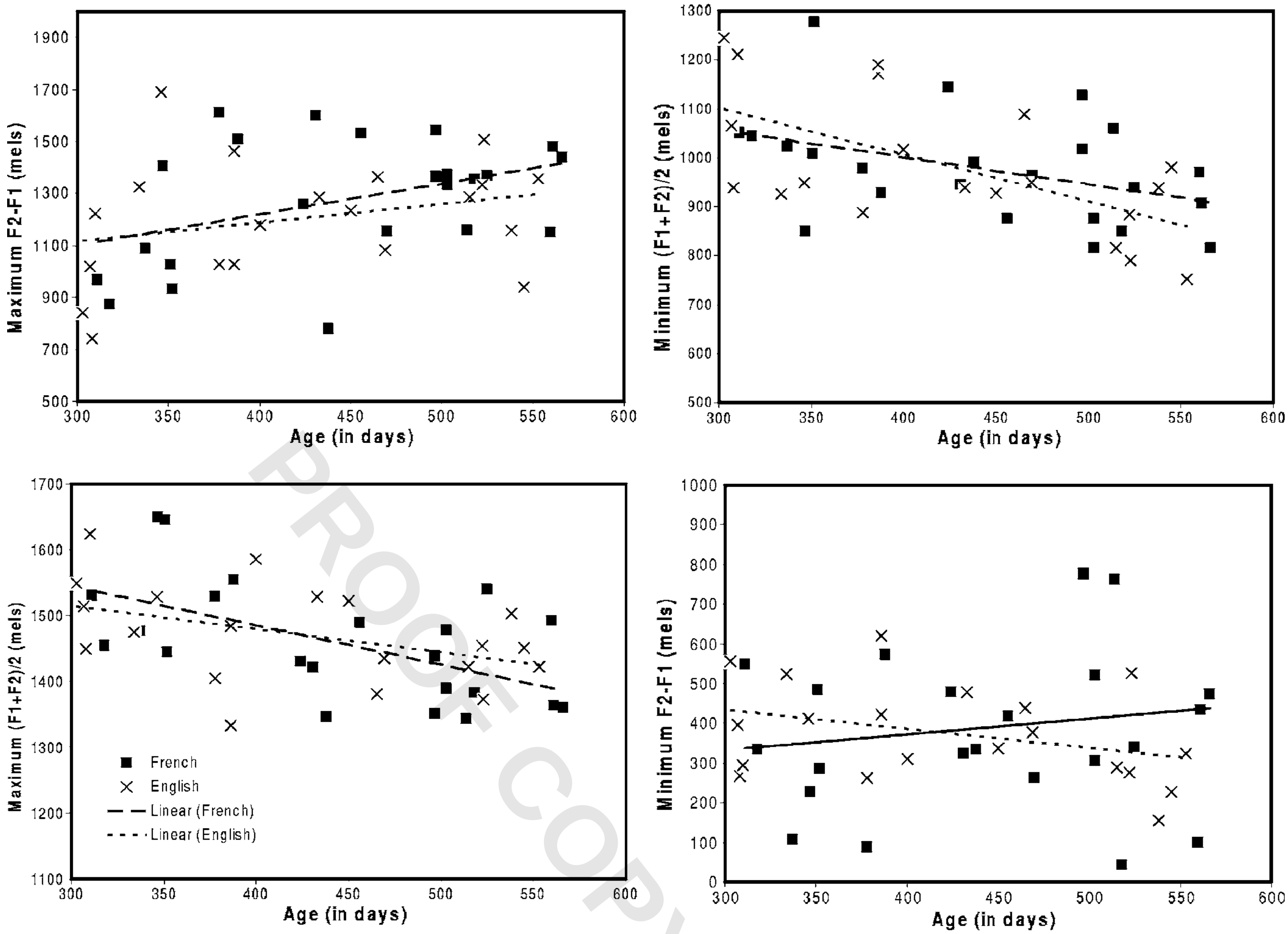

FIG. 3. Feature value for each infant's corner vowels plotted as a function of age and language group, specifically *diffuse* (top left), *grave* (top right), *acute* (bottom left), and *compact* (bottom right).

learning infants demonstrated expansion of the vowel space between 10 and 18 months. In order to interpret these developmental changes in relation to the growth of the vocal tract during this developmental period, a simulation experiment was conducted.

III. EXPERIMENT 2

A. Method

The maximal vowel spaces that could be produced by an infant, aged 6-, 12-, or 18-months of age, were modeled using the Variable Linear Articulatory Model (VLAM). This model integrates the growth data currently available (Goldstein 1980) into a previous model already existing for the adult (Maeda, 1979; 1990). The latter is based on a statistical analysis of 519 midsagittal cineradiographic images of a French speaker uttering ten sentences (Bothorel *et al.*, 1986). The analysis revealed that seven articulatory parameters (P_i, $i \in \{1, \ldots, 7\}$) could account for 88% of the variance of the tongue contours (Boë *et al.*, 1995): labial protrusion, labial aperture, tongue tip position, tongue body position, tongue dorsum position, jaw height, and larynx height. Each parameter is adjustable at a value in the range of ±3.5 standard deviations around the mean values for this articulator in the cineradiographic images. These parameters control the position of the articulators in the model, and hence the midsagittal contour. The cross-sectional area function is computed from the midsagittal contour following the Heinz and Stevens (1965) formula and the transfer function is calculated following the Badin and Fant (1984) model. VLAM integrates nonuniform vocal tract growth, in the longitudinal dimension, by two scaling factors: one for the oral cavity and another for the pharyngeal cavity, the zone in-between being interpolated. The values of the factors, from 0.3 to 1.2, were calibrated year by year and month by month based on Goldstein's (1980) length data.

Two sets of maximal vowel spaces were simulated, with the first set representing very limited articulatory movement (henceforth limited range simulation) and the second representing the full range of movement for all seven articulatory parameters (henceforth full range simulation). The limited range simulation was accomplished by generating the acoustic properties of all vowels that could be produced given the full range of variation in the jaw and lip height movements, while holding tongue tip position, tongue body position, and larynx height in the neutral position. The full range simulation was accomplished by generating the acoustic properties of all vowels that could be produced given the full range of variation in all seven articulatory parameters.

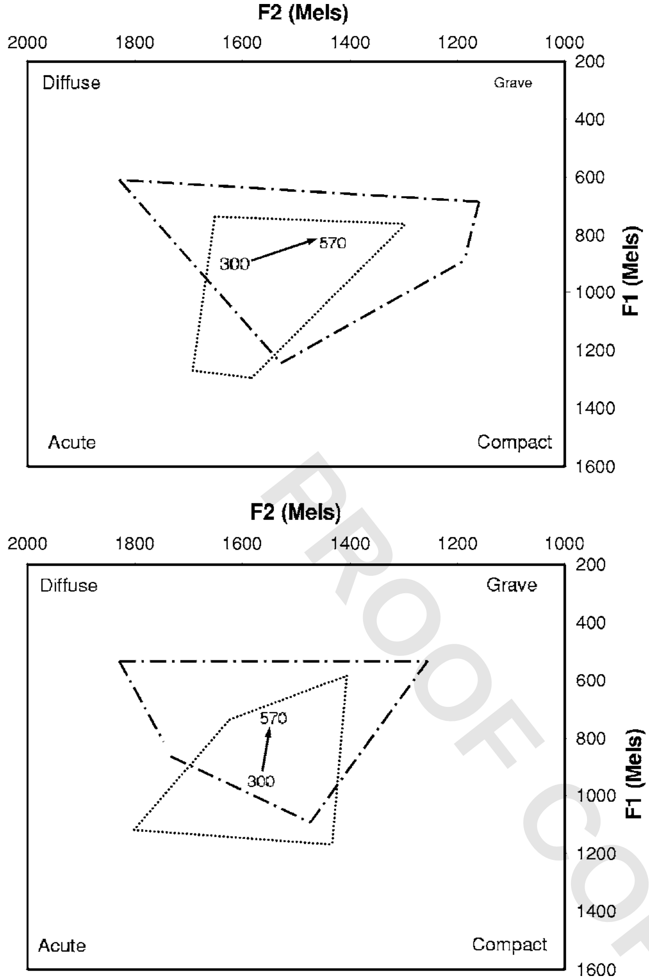

FIG. 4. Graphic summary of the findings for the English-learning infants (top) and French-learning infants (bottom). On both charts the arrow indicates the movement of the center of the vowel space as age increases from 300 to 570 days, the dotted-line quadrilaterals trace the periphery of the vowel space at 300 days, and the dashed-line quadrilaterals trace the periphery of the vowel space at 570 days.

B. Results and discussion

The resulting vowel spaces for the 6-, 12-, and 18 -month vocal tracts were described using the same procedures outlined above for experiment 1. These values for the simulated vowel spaces are shown in Table I, with the limited range simulation represented in the upper half and the full range simulation represented in the lower half of the table. Changes in these values with age, especially as shown for the limited range simulation, largely reflect increasing length of the vocal tract. As would be expected, $MF1$ and $MF2$ decrease with age although the decreases shown are quite small, less than 50 mels on average. Changes in the corners of the vowel space with age are also quite small, decreasing less than 50 mels for each feature. Most change occurred to the grave and acute features. The compact feature shows the smallest degree of age related change in the simulation as it did for the acoustic measures reported in experiment 1. Changes in the acoustic characteristics of the vowel space from the limited range simulation to the full range simulation reflect the increase in the range of vowels that the infant could produce, given full range of movement

TABLE I. Summary statistics for the limited range and full range simulated vowel space parameters for the 6-, 12-, and 18-month-old vocal tract.

Summary statistic	6 months	12 months	18 months
Limited range simulation			
Mean F1	431.00	299.24	383.62
SD F1	15.50	12.09	13.71
Mean F2	960.84	940.98	918.05
SD F2	35.97	42.20	46.38
Grave	669.02	641.43	624.29
Acute	725.26	700.62	683.82
Compact	488.15	491.54	484.38
Diffuse	671.82	680.12	684.47
Full range simulation			
Mean F1	407.73	384.18	373.04
SD F1	30.32	24.55	26.57
Mean F2	914.13	882.02	847.60
SD F2	105.36	92.08	103.97
Grave	526.63	534.88	489.84
Acute	746.93	722.16	707.09
Compact	276.89	283.10	258.86
Diffuse	765.93	722.56	724.04

of all articulators. As shown in the lower part of Table I, small decreases in $Mf1$ and $MF2$ occurred although a large increase in SD $F2$ is shown. Substantial decreases in the grave and compact features and a small increase in the diffuse feature are also apparent when comparing the limited range with the full range simulation. Thus these simulations modeled the expected expansion of the infant vowel space and suggest that this expansion occurs as a consequence of improved speech motor control.

IV. GENERAL DISCUSSION

A. Developmental changes

The most obvious developmental change for the English and French infants was an expansion of the vowel space, especially with respect to the grave and diffuse features. In addition to being consistent with previous findings (e.g., Buhr, 1980; Gilbert *et al.*, 1997; Kent and Murray, 1982; Robb *et al.*, 1997; Rvachew *et al.*, 1996), these changes are a predictable consequence of developmental changes in the infant's ability to control tongue tip, tongue body, and tongue dorsum position, independently of jaw height, as indicated by the simulations reported in experiment 2 and shown in Table I, when comparing the limited and full range simulations.

Improved control of the jaw during the infant period should manifest itself as less extreme jaw displacement during the opening and closing phases of syllable production (Green *et al.*, 2000). More control of the jaw in the midopen position should result in less extreme acuteness values (i.e., reductions in maximum $F1+F2/2$) and greater compactness values (i.e., decreases in minimum $F2-F1$). Figure 3 (bottom left) confirms a statistically significant reduction in acuteness values. No clear developmental effects on compactness values were observed however.

B. Cross-linguistic differences

A priori predictions about likely cross-linguistic differences were much more difficult to formulate because no previous studies have compared vowels produced by Canadian-English- and Canadian-French-learning infants. de Boysson-Bardies *et al.* (1989) reported a similar mean $F1$ but a slightly higher mean $F2$ for the vowel spaces of 10-month-old infants exposed to British English compared to the vowel spaces of infants exposed to Parisian French. The 10-month-old infants enrolled in this study showed a similar pattern of differences in mean formant frequencies. We expected to see a linear divergence of the center of the vowel spaces with age; however, an unexpected interaction of age and language group was observed: $MF2$ decreased in the Canadian-English group while the $MF2$ remained stable across age for the Canadian-French group. Cross-linguistic differences were also observed in $MF1$ as the children grew older: $MF1$ decreased in both groups but the decrease was much greater for the Canadian-French infants than for the Canadian-English group. Although some decrease in $MF1$ and $MF2$ is expected as the infant's vocal tract lengthens, the decrease in $MF1$ observed for the Canadian-French group and the decrease in $MF2$ observed for the Canadian-English group were each much larger than would be predicted on the basis of vocal tract growth alone (see Table I, experiment 2). It is possible that the observed interaction of age and language group on $MF2$ might be due to differences in vowel inventory. English has fewer front vowels and no rounded front vowels, in contrast to French which has three rounded front vowels and three unrounded front vowels, with the front vowels having higher $F2$ than back vowels. The $F2$ decrease in the CE group might be due to the combined effects of vocal tract growth and increased frequency of back vowels, thus decreasing $F2$. While a similar decrease in $F2$ may occur in the CF group for the same reasons, the effect may be balanced by a greater frequency of front vowels, resulting in a stable $MF2$ across age in this language group.

Our findings of a decrease in formant frequencies with age is not consistent with the earlier findings of Gilbert *et al.* (1997) and Robb *et al.* (1997) who found that $F1$ and $F2$ remained stable across age. This difference in findings across studies may be accounted for by utterance selection. In their papers, Gilbert *et al.* (1997) and Robb *et al.* (1997) acknowledge that possible changes in $F1$ and $F2$ may have been obscured by nasal resonance in the younger children's vocalizations. In the present study we controlled for this by only analyzing vowels with normal phonation and full resonance.

A cross-linguistic difference was also observed for the SD $F2$, with the Canadian-English infants showing an increase with age in the range of $F2$ values (as has been reported in other studies; e.g., Robb *et al.*, 1997; Rvachew *et al.*, 1996). The Canadian-French infants did not show this pattern of change for SD $F2$ however.

C. Future directions

These results demonstrate significant developmental changes in the shape of the infant vowel space as well as significant impacts of the auditory environment on the frequency location of the center of the infant vowel space. These patterns of developmental change and cross-linguistic differences appear to emerge after 12 months of age but are clearly evident before 18 months of age. The observed individual differences in vowel production are undoubtedly explained by a complex interaction of factors, including changing vocal tract morphology, developing speech motor control, and the child's intake of self- and other-produced speech. More research is required to understand how these factors determine infant speech output.

Recent technological advances, such as magnetic resonance imaging and computational modeling, allow us to make predictions about the impact of changing vocal tract morphology on the acoustic characteristics of speech output. More direct observation of infant articulatory movements are required however, in order to better model the impact of limited speech motor control along with the limitations imposed by the size and shape of the infant's vocal tract. Kinematic studies of jaw and lip movements indicate that maturation of speech motor control is not a linear process. For example, after examining the correlation between the spatiotemporal trajectories for adult and child jaw movements during bisyllable productions, Green *et al.* (2002) concluded that jaw movements appear to be more adultlike at the end of the first year, than at the end of the second year. Reduced stability of jaw movements at the later age may be due to the challenge of developing independent control of other articulators or the challenge of producing speech for communicative purposes.

The cross-linguistic differences that were observed in this study are difficult to interpret. Presumably, the speech that is heard by the infant provides targets for speech production that shape the specific characteristics of the infant's speech output. The exact nature of these targets is unknown. Although the acoustic characteristics of adult-produced Canadian-English and Canadian-French vowels have been described (Escudero and Polka, 2003), these kinds of descriptions are not well suited to the task of understanding the target for infant speech production. First, these descriptions are based on adult-directed speech, and it has been shown that the acoustic-phonetic nature of infant-directed speech is significantly different from that of adult-directed speech (Kuhl *et al.*, 1997). In particular, the talker's vowel space when addressing an infant is larger, with more extreme point vowels, in comparison with the vowel space produced in an adult-directed register. Second, adult-produced speech is usually described in relation to specific phonetic targets. For example, Escudero and Polka (2003) found that the Canadian-English [u] is considerably less grave than the Canadian-French [u], while the Canadian-English [æ] is more acute that the Canadian-French [æ]. However, since it is not possible to ask infants to produce a specific vowel, the infant vowel space is always described in terms of more global characteristics as we have done here. We are currently engaged in an effort to describe infant-directed speech from Canadian-French- and Canadian-English-speaking parents, in terms of the center and corners of their vowels spaces. This kind of information may facilitate the development of

PROOF COPY 023610JAS

specific predictions about cross-linguistic differences in infant vowel production across different language groups.

The infant's access to self-produced speech must also be considered. The development of speech motor control requires that the infant develop a mapping between auditory-perceptual targets, articulatory gestures, and the acoustic-phonetic product of those articulatory movements (Callan *et al.*, 2000). This model of speech development highlights the importance of feedback of the infant's own speech. A better understanding of how the infant processes this feedback is necessary if we are to predict patterns of developmental change in speech production. Investigating the role of visual speech (e.g., the visual cues for production of different vowels) is another avenue for further research.

V. CONCLUSIONS

In this study we described the vowel spaces produced by infants in terms of the center and the corners of the vowel space. The infants were drawn from two language groups, Canadian English and Canadian French, and covered a broad age range, from 10 to 18 months. The findings were interpreted in relation to simulations of vowel production given differences in vocal tract length and speech motor control. Some individual differences in the vowel spaces, such as an expansion of the vowel space into the diffuse and grave regions, were associated with ageing of the infant. These developmental changes appear to reflect maturation of the vocal tract and speech motor control. Other differences were associated with the infant's ambient language environment. Infants exposed to Canadian French demonstrated a decline in mean first formant frequencies whereas infants exposed to Canadian English showed a decline in mean second formant frequencies with age. The divergence of the vowel spaces between the two language groups emerged between 12 and 18 months of age. In order to understand the mechanism by which the ambient speech environment influences infant speech production, future research should attempt to link the characteristics of infant vowels to the infant's perception of both adult- and self-produced speech.

ACKNOWLEDGMENTS

This research was supported by research grants from the Canadian Language and Literacy Research Network and the Natural Sciences and Engineering Research Council of Canada to the first author and a postdoctoral fellowship to the second author from the Centre for Research in Language, Mind and Brain. The authors thank the infants and their families for their participation, Voula Tsagaroulis for audiometric testing, Jade Heilmann for recruiting and data collection, Pi-Yu Chiang, Jessica Whittley, Shani Abada, and Heather McKinnon for digitizing the speech samples, and Marie Desmarteau and Sara Turner for acoustic analysis.

Badin, P., and Fant, G. (**1984**). "Notes on vocal tract computation," *Speech Transmission Laboratory—Quarterly Progress Status Report*, Vol. **2–3**, 53–108.

Boë, L.-J., Gabioud, B., and Perrier, P. (**1995**). "The SMIP: An interactive articulatory-acoustic software for speech production studies," Bulletin de la Communication Parlée **3**, 137–154.

Bothorel, A., Simon, P., Wioland, F., and Zerling, J. P. (**1986**). *Cinéradiographie des Voyelles et Consonnes du Français [Cineradiographic study of French vowels and consonants]*, Institut de Phonétique de Strasbourg, Strasbourg, France.

Buhr, R. D. (**1980**). "The emergence of vowels in an infant," J. Speech Lang. Hear. Res. **23**, 73–94.

Callan, D. E., Kent, R. D., Guenther, F. H., and Vorperian, H. K. (**2000**). "An auditory-feedback-based neural network model of speech production that is robust to developmental changes in the size and shape of the articulatory system," J. Speech Lang. Hear. Res. **43**, 721–738.

Davis, B. L., and MacNeilage, P. F. (**1995**). "The articulatory basis of babbling," J. Speech Lang. Hear. Res. **38**(6), 1199–1211.

de Boysson-Bardies, B., Halle, P., Sagart, L., and Durand, C. (**1989**). "A crosslinguistic investigation of vowel formants in babbling," J. Child Lang **16**, 1–17.

Escudero, P., and Polka, L. (**2003**). "A cross-language study of vowel categorization and vowel acoustics: Canadian English versus Canadian French," 15th ICPhS, Barcelona (■, ■), pp. 836–864.

Fitch, W. T., and Giedd, J. (**1999**). "Morphology and development of the human vocal tract: A study using magnetic resonance imaging," J. Acoust. Soc. Am. **106**(3), 1511–1522.

Gilbert, H. R., Robb, M. P., and Chen, Y. (**1997**). "Formant frequency development: 15 to 36 months," J. Voice **11**(3), 260–266.

Goldstein, U. G. (**1980**). "An articulatory model for the vocal tract of growing children," Doctoral dissertation, Massachusetts Institute of Technology, Boston.

Green, J. R., Moore, C. A., Higashikawa, M., and Steeve, R. W. (**2000**). "The physiologic development of speech motor control: Lip and jaw coordination," J. Speech Lang. Hear. Res. **43**(1), 239–255.

Green, J. R., Moore, C. A., and Reilly, K. J. (**2002**). "The sequential development of jaw and lip control for speech," J. Speech Lang. Hear. Res. **45**(1), 66–79.

Heinz, J. M., and Stevens, K. N. (**1965**). "On the relations between lateral cineradiographs, area functions, and acoustic spectra of speech," *Proceedings of the 5th International Congress of Acoustics*, Vol. **1**, (■, ■), p. A44.

Kent, R. D., and Murray, A. D. (**1982**). "Acoustic features of infant vocalic utterances at 3, 6, and 9 months," J. Acoust. Soc. Am. **72**, 353–365.

Kent, R. D., and Vorperian, H. K. (**1995**). "Anatomic development of the craniofacial-oral laryngeal systems: A review," J. Med. Speech-Language Pathology **3**, 145–190.

Kent, R. D., Vorperian, H. K., Gentry, L. R., and Yandell, B. S. (**1999**). "Magnetic resonance imaging procedures to study the concurrent anatomic development of vocal tract structures: preliminary results," Int. J. Gynecol. Pathol. **49**, 197–206.

Koopmans-van Beinum, F. J., Clement, C. J., and van den Dikkenberg-Pot, I. (**2001**). "Babbling and the lack of auditory speech perception: a matter of coordination?" Developmental Science **4**(1), 61–70.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., and Lacerda, F. (**1997**). "Cross-language analysis of phonetic units in language addressed to infants," Science **277**, 684–686.

Kuhl, P. K., and Meltzoff, A. N. (**1996**). "Infant vocalizations in response to speech: Vocal imitation and developmental change," J. Acoust. Soc. Am. **100**(4), 2425–2438.

LaCharite, D., and Paradis, C. (**1997**). "Category preservation and proximity versus phonetic approximation in loanword adaptation," Linguistic Inquiry **36**, 223–258.

Maeda, S. (**1979**). "An articulatory model of the tongue based on a statistical analysis," J. Acoust. Soc. Am. **65**, S22.

Maeda, S. (**1990**). "Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal tract shapes using an articulatory model," *Speech Production and Speech Modelling*, edited by W. L. Hardcastle and A. Marchal, (Kluwer Academic, Dodrecht, The Netherlands), 131–149.

Martin, P. (**2002**). "Le système vocalique du français du Québec. De l'acoustique à la phonologie," La Linguistique **38**(2), 71–88.

Ménard, L., Schwartz, J., and Boë, L. (**2002**). "Auditory normalization of French vowels synthesized by an articulatory model simulating growth from birth to adulthood," J. Acoust. Soc. Am. **111**(4), 1892–1905.

Ménard, L., Schwartz, J., and Boë, L. (**2004**). "Role of vocal tract morphology in speech development: Perceptual targets and sensorimotor maps for synthesized vowels from birth to adulthood," J. Speech Lang. Hear. Res. **47**, 1059–1080.

Oller, D. K. (**1986**). "Metaphonology and infant vocalizations," in *Precur-*

PROOF COPY 023610JAS

sors of Early Speech edited by B. Lindblom and R. Zetterstrom (Stockton New York), pp. 21–35.
Oller, D. K. (**2000**). *The Emergence of the Speech Capacity* (Lawrence Erlbaum Associates, Mahwah, New Jersey).
Peterson, G. E., and Barney, H. L. (**1952**). “Control methods used in a study of the vowels,” J. Acoust. Soc. Am. **24**, pp. 175–184.
Robb, M. P., Chen, Y., and Gilbert, H. R. (**1997**). “Developmental aspects of formant frequency and bandwidth in infants and toddlers,” Folia Phoniatr Logop **49**, 88–95.
Rvachew, S., Slawinski, E. B., Williams, M., and Green, C. L. (**1996**). “Formant frequencies of vowels produced by infants with and without early onset otitis media,” Can. Acoust. **24**, 19–28.
Stevens, S. S., Volkmann, E. B., and Newman, J. (**1937**). “The mel scale equates the magnitude of perceived differences in pitch at different frequencies,” J. Acoust. Soc. Am. **8**, 185–■.
Sussman, H. M., Duder, C., Dalston, E., and Caciatore, A. (**1999**). “An acoustic analysis of the development of CV coarticulation: A case study,” J. Speech Lang. Hear. Res. **42**(5), 1080–1096.
Sussman, H. M., Minifie, F. D., Buder, E. H., Stoel-Gammon, C., and Smith, J. (**1996**). “Consonant-vowel interdependencies in babbling and early words: Preliminary examination of a locus equation approach,” J. Speech Lang. Hear. Res. **39**, 424–433.

PROOF COPY 023610JAS