PERCEPTUAL CATEGORIZATION OF SYNTHESIZED ENGLISH VOWELS FROM BIRTH TO ADULTHOOD

Lucie Ménard[1], Barbara Davis[2], and Louis-Jean Boë[3]

[1]Université du Québec à Montréal, Canada, [2]The University of Texas at Austin, USA,
[3]Institut de la Communication Parlée, Université Stendhal/INPG, France
menard.lucie@uqam.ca, babs@mail.utexas.edu, boe@icp.inpg.fr

ABSTRACT

The goal of this experiment is to determine the influence of non uniform vocal tract growth on the ability to reach acoustic-perceptual targets for English vowels. An articulatory-to-acoustic model integrating non uniform vocal tract growth was used to synthesize 342 5-formant vowels, covering maximal vowel spaces for speakers at 5 growth stages: newborn, 4 years, 10 years, 16 years, and 21 years old (adult stage). 37 American English speakers participated in a perceptual categorization experiment. Results indicated that the three cardinal vowels /i u æ/ can be perceived by speakers based on a newborn-like vocal tract. Articulatory-to-acoustic relationships for a given vowel may differ across growth stages.

Keywords: Speech acquisition, vowel perception

1. INTRODUCTION

It has been shown that vocal tract growth is non uniform ([3]). At birth, the infant's pharyngeal cavity is much shorter than the oral cavity, whereas the adult male's pharyngeal cavity is longer than the front cavity. The effects of non uniform cavity growth on vowel production in French speakers have been described in an earlier paper ([1]). Even though the ratio of pharyngeal length to front cavity length differs, the acoustic target for the three cardinal vowels /i u a/ can be reached in the infant vocal tract. However, in the French experiment, a larger proportion of stimuli were perceived as front and open vowels in the newborn vocal tract than in the adult male vocal tract. Thus, perceived front and open vowels by French speakers, corresponding to some of the most frequently produced vowels of the baby's early sound inventory ([2]), are favored in a newborn vocal tract. As the earlier result was language-specific, we investigated the generality of articulatory-to-acoustic relationships by analysis of English listener responses. The following experiment was designed to simulate the effects of vocal tract growth on perceived vowel targets in English.

2. METHOD

2.1. The articulatory model

We have used the *VLAM* growth model (*Variable Linear Articulatory Model*), developed by S. Maeda [6]. *VLAM* integrates knowledge acquired from previous models with currently available growth data. This model is controlled by seven articulatory parameters: protrusion and labial aperture, movement of the tongue body, dorsum and tip, jaw height, and larynx height. The growth process is simulated by modifying the longitudinal dimension of the vocal tract according to two scale factors; one for the anterior part of the vocal tract and the other for the pharynx, interpolating the zone in-between. The model is extensively described elsewhere ([1]). Simulated vocal tract lengths have been calibrated, month by month and year by year using the data provided by [7]. The *VLAM* model has been implemented and tested in an environment (*GROWTH*) originally developed based on an articulatory model for adults. Fundamental frequency values emerge following the growth data presented by [8]. The model is thus suitable for use in systematic simulation studies as well as for use in phonetics.

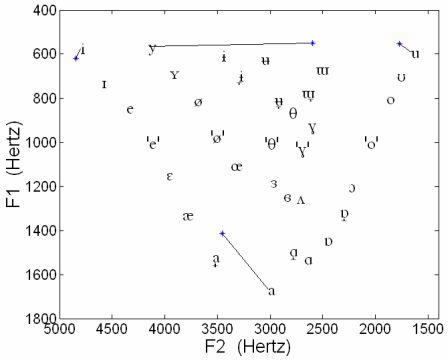
2.2. Stimuli

The stimuli used in the present study are similar to those used in the previous study of French listeners ([1]). First, maximal vowel spaces (hereafter MVS [9]) were generated by a uniform distribution of command parameters, for five growth stages: newborn, 4 years, 10 years, 16 years, and 21 years old. The following vocal tract lengths were obtained, for a neutral configuration (all parameters set to the null value): 7.70 cm (newborn), 10.67 cm (4 years old), 12.65 cm (10

years old), 13.56 cm (16 years old), and 17.45 cm (21 years old). A total of about 7,000 vowels for each age were modeled.

Then, 5-formant vowels were generated, corresponding to the 38 monophthong oral vowels of the world's languages, reported in UPSID (UCLA Phonological Segments Inventory Database), developed by I. Maddieson and described in [10]. The UPSID vowels were an appropriate sample covering the entire MVS, while ensuring articulatory coherence, as well as a reasonable number of stimuli. For each growth stage, the 38 vowels were situated within the MVS following criteria inspired from the dispersion-focalization theory ([10]). In this theory, it is assumed that vowel systems are shaped by both dispersion constraints increasing mean formant distances between vowels, and focalization constraints increasing the trend to have focal vowels in the system, that is, vowels with close F1 and F2, F2 and F3, or F3 and F4. Once established, the optimal acoustic F1-F2-F3 values were related to their underlying articulatory values (command parameters in VLAM) by an inversion procedure exploiting the pseudo-inverse of the Jacobian matrix (details can be found in [1]**)**. These vowel prototypes will be referred to as "acoustic prototypes". Second, vowel prototypes based on the adult's command parameters were generated for each of the four other growth stages: newborn, 4 year, 10 year, and 16 year olds. These stimuli will be referred to as "articulatory prototypes". Thus, 38 vowel stimuli were synthesized for the adult vocal tract, whereas 76 prototypes (acoustic and articulatory prototypes) were generated for the other four vocal tracts (for a total of 342 stimuli).

Figure 1: Formant values, in the F1/F2 plane, of acoustic prototypes, for the newborn vocal tract. For the sake of clarity, only /i y u a/'s articulatory prototypes are represented by stars, and linked to the same vowel's acoustic prototype.



Fundamental frequency values of 436 Hz, 337 Hz, 229 Hz, 155 Hz, and 112 Hz were associated, respectively, with the newborn, the 4-year-old, the 10-year-old, the 16-year-old, and the 21-year-old vocal tracts. Fig. 1 shows the values of the 38 acoustic prototypes synthesized in the newborn vocal tract, in the F1 vs. F2 and F2 vs. F3 acoustic spaces. Articulatory prototypes are also depicted for /i y u a/.

2.3. Procedure

One occurrence of each of the 342 stimuli was presented binaurally via high-quality headphones, to a group of 37 American English listeners, age 18 to 25. The stimuli were grouped into 5 blocks of 76 (or 38 for the adult stage) randomized items, each token of a single age being presented in one block. Each block was followed by a brief pause. Listeners were undergraduates without a phonetics background. They had the task of identifying, using a button press, the perceived vowel among the 12 American English oral vowels /i ɪ e ɛ æ ɑ ɔ o ʊ u ʌ ɚ/. Each vowel was represented by a monosyllabic word of the form [hVd]: "heed" ([hiːd]), "hid" ([hɪd]),"hayed" ([heɪd]), "head" ([hɛd]), "had" ([hæd]), "hod" ([hɑd]), "hawed" ([hɔd]), "hoed" ([hod]), "hood" ([hʊd]), "who'd" ([hud]), "hud" ([hʌd]), "heard" ([hɚd]). The test lasted about forty minutes and took place in a quiet room. Informed consent was obtained from all participants.

3. Results

3.1. Dominantly perceived vowels

Stimuli were grouped according to their dominantly perceived vowel category; the category triggering at least 50% agreement among listeners. For each stimulus, formant values and F0, in Hertz, were converted into a Bark scale, using the following conversion formula: $F_{bark} = 7*\text{asinh}(F_{Hz}/650)$. Fig. 2 depicts the dispersion ellipses of each perceived vowel category (considering dominantly perceived vowels) in the newborn vocal-tract, the 4-year-old vocal-tract and the adult vocal tract, with a radius of ±1.5 standard deviation around the mean, in the F1 vs. F2 space.

First, the corner vowels /i u æ/ were perceived based on a newborn-like vocal tract, as well as in

the more mature vocal tracts. Thus, the relatively short pharyngeal cavity of the infant does not prevent listeners from perceiving perceptual targets related to the extreme positions of the vowel space.

Figure 2: Dispersion ellipses of dominantly perceived English vowels by adult subjects, for stimuli produced by a newborn, a 4-year-old, and an adult male vocal tract.

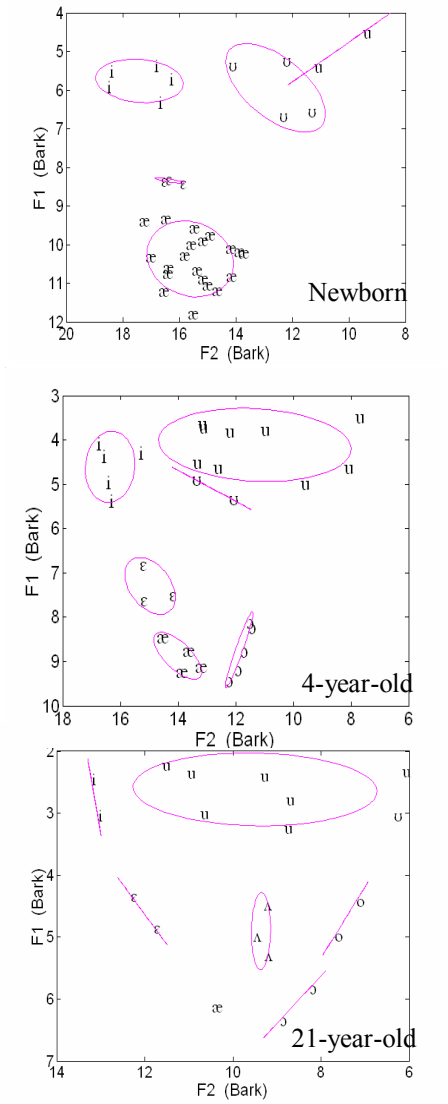
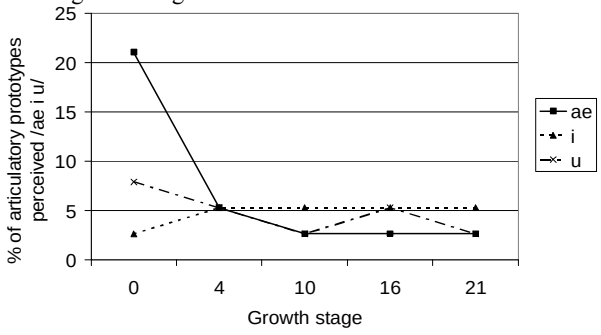


Fig. 2 also reveals that not all vowel categories can be consistently perceived when produced by vocal tracts representative of various growth stages. For instance, only 5 vowel categories (/i ʊ u ɛ æ/) are perceived by at least 50% of the listeners in the newborn-like vocal tract whereas 6 to 8 vowel categories (including /ɔ o ʌ/) are perceived at older growth stages. These differences may be related to the acoustic differences between less intelligible high-pitched voices (like the newborn) and more intelligible low-pitched voices ([1]). In addition, durational cues, not modeled in the present study, may be more important for vowel perception in newborn-like vocal tracts compared to older vocal tracts. Further analyses are conducted in order to investigate these hypotheses. Furthermore, an examination of Fig. 2 suggests different organization of the acoustic vowel space across growth stages. Indeed, the dispersion ellipses associated with the low vowel /æ/ and the high vowels /i ʊ/ enclose much broader acoustic areas in the newborn vocal tract, compared to the adult vocal tract. On the contrary, the high tense vowel /u/ is related to a larger acoustic area for the adult male than for the newborn. Thus, larger acoustic areas are related to perceived low and high lax vowels in the infant's vocal tract, compared to the adult male vocal tract.

3.2. Articulatory-to-acoustic relationships

The evolving acoustic-to-perceptual mapping during growth suggested in the previous section likely involves different perception of a given articulatory maneuver for the five growth stages. As a result, for the newborn and the adult male, similar underlying commands (corresponding to the articulatory prototypes) would be related to different perceptual targets. In order to assess this hypothesis, the number of dominantly perceived articulatory prototypes was calculated for each growth stage. Recall that the "articulatory" prototypes are the 38 stimuli generated in each vocal tract with similar underlying articulatory gestures. Results are depicted in Fig. 3 for /i u æ/. A larger proportion of articulatory prototypes were dominantly perceived as /æ/ in the newborn vocal tract, compared to the other growth stages. A similar trend is noticeable for /u/ but to a lesser extent.

Figure 3: Percentage of articulatory prototypes dominantly perceived /æ i u/ by English listeners, for each growth stage.



4. Discussion

These results suggest that length differences between a newborn-like vocal tract and an adult male vocal tract involve different articulatory-acoustic-perceptual relationships.

At the acoustic-to-perceptual level, it has been shown that perceptual targets corresponding to the corner vowels /i u æ/ can be reached when produced by a speaker with a relatively shorter pharyngeal cavity compared to the oral cavity. Hence, the lack of some vowels, in infant's early vocalizations, can be attributed to immature motor control capacities, which prevent the newborn from achieving such perceptual goals, as suggested by [2] and [4], but not to vocal tract length or configuration. However, a comparison of the size of the dispersion ellipses reveals that the perceived ellipses corresponding to /i ɛ æ ʊ/ enclose a much broader acoustic area in the newborn-like vocal tract than in the adult male vocal tract. Thus, these results suggest that front and low vowels are favored in the newborn's vocal tract configuration. Vocal tract length can thus constrain the infant's early vowel inventory. This result for English listeners supports the results of our previous study for French listeners ([1]).

At the articulatory-to-acoustic level, an examination of the proportion of dominantly perceived articulatory prototypes reveals different relationships according to vocal tract length and configuration. This was especially true for the perceived vowel /æ/: at the newborn stage, more than 20% of the articulatory prototypes were associated with this vowel category, whereas less than 5% are perceived as the same category, at the adult stage. Thus, 15% of the stimuli related to the target /æ/ for the newborn are no longer perceived as /æ/ across vocal tract growth stages. This result suggests that the articulatory-to-perceptual map acquired at the earliest stage of vocalizations is modified with growth.

5. Conclusions

This study aimed at assessing the influence of non uniform vocal tract growth on the ability to reach acoustic-perceptual targets corresponding to English monophthong oral vowels. 342 synthesized vowels generated in vocal tract lengths representative of a newborn, a 4-year, a 10-year, a 16-year, and an adult male were submitted as a perceptual categorization test to 37 American English listeners. Results show that length differences do not prevent listeners from perceiving the young speaker as reaching acoustic-perceptual targets related to the corner vowels /i æ u/. However, articulatory-to-acoustic-to-perceptual relationships differ across growth stages. These results can be used to support the generality of results previously found in a cohort of French listeners.

6. Acknowledgements

This work was supported by the Social Sciences and Humanities Research Council of Canada, and by the FQRSC (Quebec).

7. References

[1] Beck, J. M.: "Organic variation of the vocal apparatus", in Hardcastle, W. J. et Laver, J. (eds), *Handbook of Phonetic Sciences*, 256-297, 1996.

[2] Boë, L.-J., Perrier, P., Guérin, B., and Schwartz, J.L., "Maximal Vowel Space", *Eurospeech 89*, 281-284, 1989.

[3] Boë, L.-J. and Maeda, S., "Modélisation de la croissance du conduit vocal. Espace vocalique des nouveaux-nés et des adultes. Conséquences pour l'ontogenèse et la phylogenèse", *Journées d'Études Linguistiques : « La Voyelle dans Tous ces États »*, Nantes, 98-105, 1997.

[4] Davis, B. L., and MacNeilage, P. F., "The Articulatory Basis of Babbling", *Journal of Speech, Language, and Hearing Research*, 38: 1199-1211, 1995.

[5] Goldstein, U. G, *An articulatory model for the vocal tract of the growing children*, Thesis of Doctor of Science, MIT, Cambridge, Massachusett, 1980.

[6] MacNeilage, P.F. and Davis, B. L., "Acquisition of speech production : Frames then content", in Jannerod, M. (ed), *Attention and Performance XIII : Motor Representation and Control*, Hillsdale (NJ), Lawrence Erlbaum, 453-475, 1990.

[7] Maddieson, I., *Patterns of Sounds*, 2nd edition, Cambridge University Press, Cambridge, 1986.

[8] Ménard, L., Schwartz, J.-L., and Boë, L.-J., "Role of Vocal Tract Morphology in Speech Development: Perceptual Targets and Sensorimotor Maps for Synthesized French Vowels From Birth to Adulthood", *Journal of Speech, Language, and Hearing Research*, 47(5):1059-1080, 2004.

[9] Schwartz, J.-L., Boë, L.-J., Vallée, N. and Abry, C., "The Dispersion-Focalization Theory of vowel systems", *Journal of Phonetics*, 25, 255-286, 1997.

[10] Vorperian, H. K., Kent, R. D., Lindstrom, M. J., Kalina, C. M., Gentry, L. R., and Yandell, B. S. (2005): "Development of vocal tract length during early childhood: A magnetic resonance imaging study", *Journal of the Acoustical Society of America*, 117, 338-350.