PERCEPTION AND IMITATION OF FINNISH OPEN VOWELS AMONG CHILDREN, NAÏVE ADULTS, AND TRAINED PHONETICIANS

Lotta Alivuotila[1,2], Jussi Hakokari[3,2], Janne Savela[3,2], Risto-Pekka Happonen[1], Olli Aaltonen[2]

University of Turku
[1]Department of Oral and Maxillofacial Surgery
[2]Department of Phonetics
[3]Department of Information Technology
lotta.alivuotila@utu.fi, jussi.hakokari@utu.fi

ABSTRACT

We have compared identification and imitation of a synthetic vowel continuum varying from [æ] to [ɑ] among Finnish speakers. Results indicate that special practice is needed for listeners to monitor only the sensory information in imitation and to bypass what is learned and stored in long-term memory. We had three kinds of participants: preschool children, naïve adults, and phoneticians. All the groups were able to identify the vowels systematically in the listening experiments, although individual differences were found in the location of the category boundary. Adults performed better than children in goodness rating. The experts rated goodness accurately. After the listening tests, the participants imitated the same stimuli. In this condition imitation proved to be categorical among children and naïve adults as the previous studies have suggested. Phoneticians could imitate gradually changing vowel qualities without any abrupt changes reflecting the way how the continuum was categorized into phonemes.

Keywords: imitation, categorization, identification, goodness rating.

1. INTRODUCTION

From the 1960's on, imitating synthetic vowels has been a method in examining how vowels are categorized into phonemes. Investigations conducted by Chistovich et al. [1] and followed by Kent [e.g. 3,4] showed that imitation was more difficult for ambiguous vowels (sounds that had no prominent phonemic identity) than for unambiguous ones. Also Schouten [9] reported imitation behavior of Dutch-English bilinguals to be categorical to some degree and affected by the vowel systems of both languages. Repp and Williams [8] replicated Kent's study and obtained similar results, although strictly categorical responses were scarce. These results support the idea that speech production and perception are closely linked in speech processing.

It has been suggested that children perform more poorly than adults in both speech perception and imitation (see Kent [6] referring to Eguchi [2]). Kent reported children to be more affected by their native vowel system than adults while imitating [6]. Investigating preschool children's imitation ability, he found out that the precision of vowel imitation is limited as much, if not more, by perceptual aspects as by imprecision in articulation. Therefore vowels produced by a speaker may sound the same to him regardless of the small acoustical differences between them. [5].

In many of the previous investigations the imitation aspect has been emphasized while the identification aspect has been given less attention. Chistovich et al. [1] and Kent [3] concluded that categories present in the participants' imitation responses were different from those emerged from categories of the identification responses. Obviously, this was due to the method used in perception research at that time. There was no way to reveal the internal structure of vowel categories, because listeners were usually asked to label the stimuli by fairly crude written identification tasks. The perception aspect was investigated more thoroughly in the study at hand by introducing forced choice identification and goodness-rating tests.

We compared perception and imitation of synthetic Finnish unrounded open vowels in the front-back continuum [æ-ɑ] among three participant groups of children, naïve adults and trained phoneticians. /æ/ and /ɑ/ are contrastive phonemes in Finnish with no other phoneme category acoustically in between. Consequently, it was possible to manipulate only the 2nd formant frequency (F2) that corresponds to the horizontal movement of the tongue, separating /æ/ from /ɑ/. Other formants, such as the 1st formant (F1) mainly defined by openness of the oral channel (vertical movement), were fixed throughout the continuum. The previous studies have manipulated more than one formant.

The primary objectives of this investigation were to examine the participants' ability to identify, rate goodness, and imitate synthetic vowel stimuli in which only a single dimension was manipulated. The ultimate objective of our project is to investigate the topical question of the link between perception and production in human speech processing.

2. METHODS

2.1. Participants

The eight participants were all male, native South-Western Finnish speakers with no known history of hearing defects. The participants were grouped by age and phonetic experience. The three naïve adults were from 29 to 31 years old and had an academic education in natural sciences. The three children were all 5-year-olds and enrolled in the same nursery. One of the children had been diagnosed with mildly delayed linguistic development but no disordered speech. The two experts were trained phoneticians (28 and 68 years old) specifically experienced in pronunciation and transcription. Furthermore, they were familiar with the details of the experimental design, the synthetic stimuli included.

2.2. Stimuli

14 synthesized vowels in continuum [æ-ɑ] were used in the experiment. The isolated steady-state vowels were generated with Klatt synthesis software. The resulting stimuli were measured with Praat analysis software for future comparison. The continuum was created by varying the frequency values for F2 from 1995 Hz (1500 mels) to 969 Hz (980 mels) in equal steps of 40 mels. The frequency values for F1 and F3 were fixed at 756 Hz (830 mels) and 2517 Hz (1700 mels), respectively. The stimuli were 350 ms long, with fundamental frequency first rising from 100 Hz to 120 Hz by 120 ms, then falling down to 80 Hz by the end.

2.3. Procedure

The experiment was designed with PXLab software and was run on a laptop computer. A headset was used for audio playback and recording. The experiment consisted of two different tasks. First the participant was asked to perform an identification task and an imitation task was followed. A total of 70 synthetic vowels were randomly presented for both tasks, including five tokens of each vowel stimulus. The experiment lasted about 15 minutes. The experiments took place in various locations, but all adult participants worked privately in a room with no external diversions. The children worked under supervision.

2.4. Identification task

The participants heard a range of vowels forming a continuum of ambiguous vowels between two unambiguous endpoints. The participants were required to identify the randomly arranged stimuli as one or the other of the two endpoints (forced choice). Identification decision was made by pressing one of the two laptop keys labeled <Ä> (=/æ/) and <A> (=/ɑ/) according to the Finnish orthography. Reaction time was measured.

After the participants had labeled the heard stimulus, they were required to evaluate its "goodness" as a representative of the phonemic category. The adults rated the stimuli by pressing one of seven keys labeled from "very good" (7) to "very poor" (1). The children were asked to evaluate the stimuli by pressing one of two keys "good" (7) or "poor" (1) instead, as operating a scale of seven options was considered unnecessarily challenging for 5-year-olds.

2.5. Imitation task

The participants heard the stimuli in a random order. They were instructed to imitate each vowel as closely as possible. Having produced the imitation, the participants were asked to press the <completed> key to move on to the next stimulus. Response time was not limited. The participants encountered each stimulus five times during the experiment. The responses were recorded for later acoustical analysis.

2.6. Acoustical analysis

The responses were analyzed with Praat software (version 4.4.33). Since differences in the point of measurement can induce considerable differences in formant frequencies, mean values were extracted from as great an area of the vowel as possible.

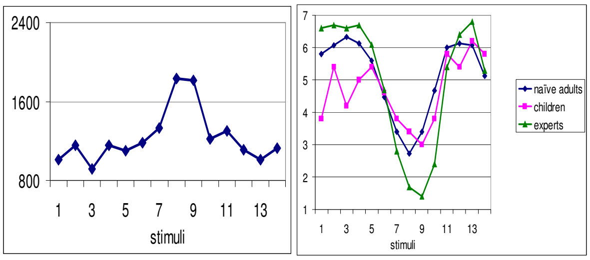
F1 and F2 values were measured from spectrograms. Other formant frequency values were overlooked as F2 was the only one manipulated in the continuum. The maximum formant was set to 5500 Hz for the children and to 5000 Hz for the adult males. Praat was configured to look for 5 formant peaks within the abovementioned bands. Occasionally, after serious consideration, it was necessary to measure the frequencies manually when the software could not track the formants reliably with the given configuration. That was carried out by carefully examining harmonic structure in a spectral slice.

The reliability of the spectrographic measurements was confirmed by having another phonetician reanalyze a part of the imitation data and make spot checks throughout the material. Greater measurement errors were expected for children than for adults. The error in estimation of formant frequency is considered to be directly related to the fundamental frequency of the speaker's voice; i.e. a quarter of the speaker's fundamental frequency (Kent [6] referring to Lindblom [7]). Based on this relationship, the expected error for the adult males was about 30 Hz (F_0 ~110 Hz). The expected error for the children was about 60 Hz (F_0 ~240 Hz). Separate means and standard deviations for the F2 data were calculated for five imitations of each of the 14 synthetic stimuli. The imitation results are presented as scatter plots showing F2 of individual vowels.

3. RESULTS AND DISCUSSION

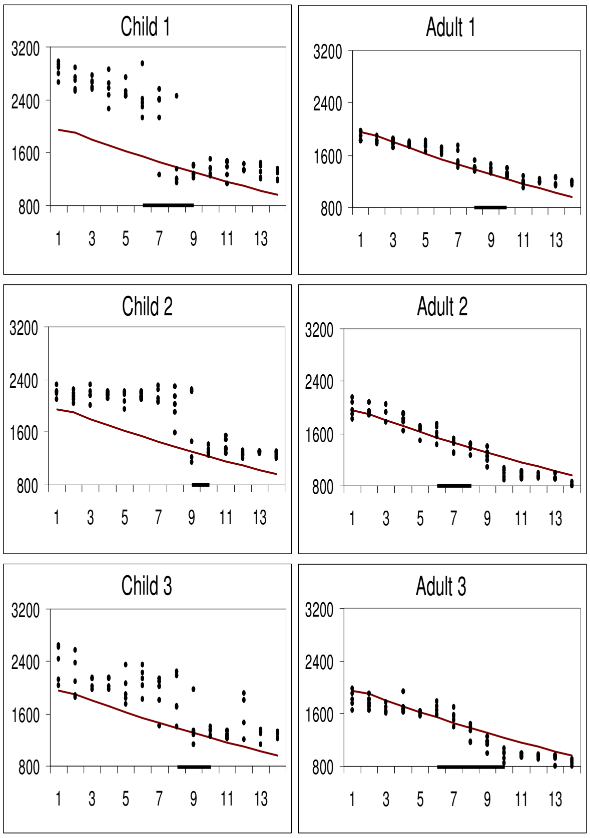
All participants performed the identification task easily. The phoneme boundary was located between the stimuli 6 and 11 in all subjects, but there was great individual variation (from 6-10 to 9-10). All but one participant's (Child 2) reaction time roughly doubled at the category boundary [fig. 1].

Figure 1: Mean reaction time of identification (in ms, left) and goodness rating (scale of 1-7, right) of all participants.



Children did not perform as well in goodness rating. Substantial individual differences occurred in children: Child 1 had no difficulty in rating goodness, Child 2 performed the task with uncertainty, and Child 3 was not able to rate goodness at all. The naïve adults rated the stimuli with no difficulty although they were less precise than the experts [fig. 1].

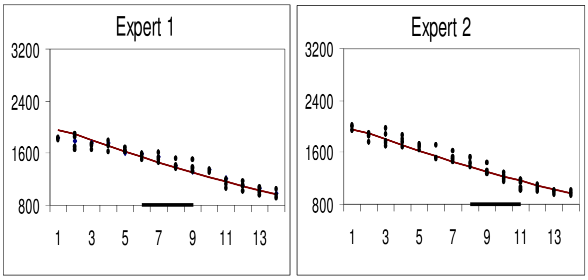
Figure 2: Children's (left) and naïve adults' (right) imitation results of the 14 stimuli (F2, in Hz). Perceptual category boundaries are marked with a bold line on the x-axis.



Imitation proved to be categorical among children and naïve adults. Children's performance [fig. 2] was the most strongly affected by their native vowel system. Variation in their pronunciation increased heavily at the category boundary, and they clearly produced either a front or a back vowel but were not able to produce the central vowels in between (~[a]). Child 3, while having normal speech, suffered from mildly delayed linguistic development and performed poorly compared to his peers.

Also naïve adults [fig. 2] were clearly guided by experience-based phonemic models. They had somewhat more variation around the category boundary, but not as much as the children. They did produce central vowels, but not as reliably as the experts.

Figure 3: Experts' imitation results of the 14 stimuli (F2, in Hz). Perceptual category boundaries are marked with a bold line on the x-axis.



Conversely, the experts [fig. 3] were able to ignore phonemic models while imitating.

We have shown that children, naïve adults and phonetically trained experts categorize easily both unambiguous and ambiguous vowels in which only F2 was manipulated. Age or phonetic experience did not seem to affect identification, which suggests that categorical perception of vowels matures early and does not develop much at later age. Goodness rating and imitation was found to sharpen with age and phonetic experience. Children were strongly affected by their native vowel system in the imitation task. Also naïve adults were clearly affected by acquired phoneme categories; their productions were somewhat uniform at the unambiguous endpoints. Trained phoneticians, however, were able to exclude the influence of phonemic models.

It is notable that while children are generally considered adept at acquiring a new language, the monolingual 5-year-olds in the study could not imitate ambiguous vowels. They have obviously acquired their native vowel system, but their cognitive and motor skills, on the other hand, are immature. Furthermore, they have not conceptualized their language the way literate adults have. Intensive training, analogous to exposure to a second language, might still produce good results.

The groups had difficulties in maintaining a constant F1 in similar proportions as they had difficulties with F2, but no uniform boundary effect was found. Some of the participants produced lower F1 towards the front end of the continuum ([æ]), which is a typical feature of Finnish speakers' vowel space. The Pearson correlation between mean F2 of the synthetic and the imitated vowels (all $p<0.001$) were .902 - .944 for children, .961 - .988 for adults, and .983 and .991 for the experts. Average standard deviation of mean for the children's F2 varied from 118 to 196 Hz (mean 168 Hz). The corresponding figures for adults were 54 - 83 Hz (mean 73 Hz) and for experts 51 Hz and 58 Hz.

4. CONCLUSION

We conclude that the groups had different imitation strategies. Children made use of acquired language-dependent phonemic models activated by the stimuli. Adults were inhibited by phonemic models to a lesser degree. Experts, however, could distinguish the ambiguous vowels in the continuum when imitating.

5. REFERENCES

[1] Chistovich, L., Fant, G., Serpa-Leitao, A., Tjernlund, P. 1966. Mimicking of synthetic vowels. *Speech Transmission Laboratory - Quarterly Progess and Status Report*.2., Stockholm, 2, 1-18.

[2] Eguchi, S. 1976. Difference limens for the formant frequencies: Normal adult values and their development in children. *J. Am. Audiol. Soc.*, 1, 145-149.

[3] Kent, R.D. 1973. The imitation of synthetic vowels and some implications for speech memory. *Phonetica* 28, 1-25.

[4] Kent, R.D. 1974. Auditory-motor formant tracking: a study of speech imitation. *J. Speech Hear. Res.* 17, 203-222.

[5] Kent, R.D. 1978. The imitation of synthesized vowels by preschool children. *J. Acoust. Soc. Am.* 63, 1193-1198.

[6] Kent, R.D. 1979. The imitation of synthesized English and nonEnglish vowels by children and adults. *J. Psychol. Res.* 8, 43-60.

[7] Lindblom, B. 1972. Comments on Paper 15 – Development of speech sounds in children, by S. Eguchi and I. Hirch. In Fant, G. (ed.) *International Symposium on Speech Communication Ability and Deafness*, Washington D.C.

[8] Repp, B.H., Williams, D.R. 1985. Categorical tendencies in vowel imitation: preliminary observations from a replication experiment. *Speech Communication* 4, 105-120.

[9] Schouten, M.E.H. 1977. Imitation of synthetic vowels by bilinguals. *J. Phonetics* 5, 273-283.