ISCA Archive
http://www.isca-speech.org/archive

Speech Prosody 2008
Campinas, Brazil
May 6-9, 2008

Octave Equivalence as an Aspect of Stimulus-Response Similarity During Nonword and Sentence Imitations in Young Children

Beate Peter, Carol Stoel-Gammon, and Daniella Kim

Department of Speech and Hearing Sciences
University of Washington, Seattle
bvpeter@u.washington.edu

Abstract

The ability to perceive similarities between musical notes one octave apart has been documented in all ages including infants. Whether the octave relationship as an aspect of similarity in speech stimuli is as robust in speech tokens is unknown. The purpose of this study was to examine whether children demonstrate an awareness of this octave relationship in the production of speech stimuli during speech imitation tasks. Eleven children, age 4;7 to 6;9, imitated nonwords and sentences presented by male adult voices, at pitch levels below young children's vocal ranges. Results show that the participants imitated the stimulus pitches one octave higher. The evidence was particularly robust in the nonword imitation task, where the low pitch levels in the stimulus required the children to adjust their conversational pitch levels downward to produce pitches at an octave above the stimulus. We conclude that young children are capable of perceiving the octave relationship as a salient aspect of similarity in speech, and that they utilize this aspect of similarity when imitating low-pitched speech stimuli below their vocal ranges to achieve a more authentic sounding reproduction of the stimulus.

1. Introduction

Musical tones that differ in frequency by a factor of 2 to 1 are perceived as being one octave apart in tonal pitch, and this octave relationship is easily perceived as one of close similarity. The perceived similarity between two tones an octave apart is so robust that all advanced musical cultures define scales by specifying intervals within the octave [1]. This similarity in perceived tones, called octave equivalence, has been demonstrated in humans of all ages including infants [2, 3].

Frequency is also encoded in the human voice during speech, where it is derived from the vibrating frequency of the vocal folds. This parameter is known as fundamental frequency, often referred to as F0 and perceived as vocal pitch. The rate of vibration is a function of vocal fold length, mass, and stiffness. To date, it is unknown whether vocal pitches produced during speech are processed in ways similar to musical pitches with respect to the perceptual octave equivalence observed in musical tones.

Because vocal fold anatomy changes across the lifespan and as a function of gender, F0 ranges vary across the population. Eguchi and Hirsh [4] measured typical fundamental vocal frequencies in 4-year-old boys and girls during sentences, arriving at a mean F0 of 286 Hz (SD = 26 Hz), and in 5-year-old children, the mean F0 was 289 Hz (SD = 23 Hz). Typical F0 mean values for male adults between 25 and 50 years of age in spontaneous speech and read speech range between 100 Hz and 120 Hz [5, 6]. A gross estimate of an F0 ratio between children's and adult male voices, hence, would approximately evaluate to 2.6.

Pitch matching between speakers in conversational contexts has been described in the discourse analysis literature [7, 8]. Referred to as pitch concord or tone concord, this phenomenon occurs when one speaker begins a response to an utterance by another speaker with the same relative pitch level found at the end of the preceding utterance. The concordant relative pitch is interpreted to mean that the second speaker wishes to convey agreement with what the first speaker just said, whereas disagreement is frequently encoded in a mismatched relative pitch ("concord breaking"). In discourse analysis, pitch levels are frequently assessed perceptually and relative to the speaker's voice range, using terms such as high key, mid key, and low key. The extent to which speakers attempt to match a given pitch when imitating another speaker, whether in the relative sense of high, mid, and low keys or in absolute measurements such as the Hertz scale, is unknown.

The purpose of the present study was to investigate the hypothesis that children show octave equivalence when imitating speech tokens that are below their vocal ranges, in an attempt to approximate the model as closely as possible. Finding evidence of octave equivalence in their productions would also imply that they perceive vocal pitches one octave apart as especially similar.

2. Method

Eleven children with typical development were enrolled in the study. Their ages ranged from 4 years, 7 months to 6 years, 9 months. The data were collected as part of the first author's dissertation study on subtypes in childhood speech disorders, where the participants presented here served as control participants with typical development [9, 10].

Each participant completed two imitation tasks. The first of these was based on the *Tennessee Test of Rhythm and Intonation Patterns* (T-TRIP) [11], a nonstandardized test of expressive prosody. During the <u>Rhythm</u> subtest, the participant hears prerecorded sequences of the syllable "ma" and imitates them. The sequences vary in length and stress patterns. For instance, Item 1 is MA.ma (where capital letters indicate a stressed syllable, lower-case letters indicate an unstressed syllable, and the period mark indicates a syllable boundary). Item 11 is MA.ma.ma.MA.ma, a much longer and metrically more complex sequence. Items 1 through 12 yielded 45 syllables for analysis, of which 18 were stressed and 27, unstressed. Appendix A shows the schematic for the presented items. The voice on the CD provided by the manufacturer is a male adult, and each item is presented for imitation twice in consecutive fashion. In the dissertation study examining subtypes of speech disorders, this task was chosen as a measure of prosodic accuracy in terms of number of syllables correct, proportionately timed vowel durations, and level of distinctiveness between stressed and unstressed vowels. For the present study, the Rhythm task served as a measure of vocal similarity between the adult model and the child's imitation in the absence of linguistic load.

The second task was based on the Recalling Sentences subtest from the *Clinical Evaluation of Language Fundamentals- Preschool* (CELF-P) [12]. In this task, the participant imitates a phrase or sentence that was presented in story format. This subtest contains three trial sentences, which were not analyzed, and eighteen stimulus sentences of varying length. The story was prerecorded by an adult male speaker in his 20s and presented to each participant via laptop computer and external speakers. In the dissertation study on subtypes of childhood speech disorders, this imitation task was chosen as a measure of expressive language ability as it captures, per test manual, length of utterance, syntactic complexity, and informational content. For the purposes of the present study, it was chosen as an imitation task with a substantial linguistic load, allowing us to measure vocal pitch in the context of a task requiring the child to reproduce sentences of increasing length and complexity.

In order to show that the children adjusted their F0 to double the male adult's target, producing the imitation at one octave higher, we measured each child's customary pitch from conversational speech samples collected during the study sessions.

All acoustic analyses were completed using *Praat Version 4.2.09* [13], measuring F0 in Hertz. Pitch input ranges were 75 Hz to 500 Hz for the child voices and 75 to 300 Hz for the male adult tokens. To estimate conversational F0 for each participant, between 5 and 12 utterances were isolated from the recordings and mean F0 for each utterance was calculated and averaged. For the nonword imitation task, vowel pitches were measured in a 50 msec time window beginning 20 msec after vowel onset. Since each item was presented and imitated twice in consecutive fashion, the second imitation was selected for acoustic analysis unless the first imitation contained a more accurate syllable pattern. For the sentence imitation task, pitch was measured as an average across the sentence. For each analyzed item, an imitation-to-stimulus ratio was calculated and averaged individually for each participant. To show how the children's F0 during the imitation tasks differed from their conversational F0, ratios between conversational F0 and the F0 measured during each of the imitation tasks was calculated.

The data were checked for reliability by remeasuring approximately 15% of the tokens. For the nonword imitation task, the average absolute difference between the two sets of measurements was 2 Hz, and the analogous absolute inter-rater for the sentence imitation task was 8.5 Hz. F0 measurements for the conversational samples differed on average by 3.4 Hz. These values were judged to be acceptable for the purposes of this study.

3. Results

3.1 Mean Conversational F0

During conversational speech, the participants' mean F0 ranged from 214 Hz to 330 Hz, with an average F0 of 292 Hz for the entire group. These values are consistent with expectations for the children's age range. Table 1 shows the mean conversational F0 and relative standard deviation for each participant, along with age and gender information.

Table 1: *Participant code, age (years;months), gender, mean conversational F0, and relative standard deviation*

Participant Code	Age (y;m)	Gender	Mean Conversational F0 (Hz)	RSD
T01	6;2	M	283	0.05
T02	5;1	M	312	0.06
T03	6;0	M	325	0.15
T04	5;1	M	279	0.08
T05	6;9	M	265	0.07
T06	5;1	M	303	0.14
T07	6;1	F	229	0.18
T08	4;10	M	330	0.08
T09	4;10	F	353	0.13
T10	4;11	F	316	0.20
T11	6;4	M	214	0.10
		Mean	292	0.11

3.2 Nonword Imitation

During the nonword imitation task, the adult male voice showed an average F0 of 116 Hz. As a group, the children produced the vowels with a mean F0 of 245 Hz, indicating a lower pitch than the mean F0 of 292 Hz observed in their conversational speech. Vowel ratios calculated for each child-produced vowel relative to the adult model showed an overall average of 2.03 for the group, ranging from 1.67 for participant T09 to 2.38 for participant T08. Figure 1 shows the distribution of mean ratios across the participants.

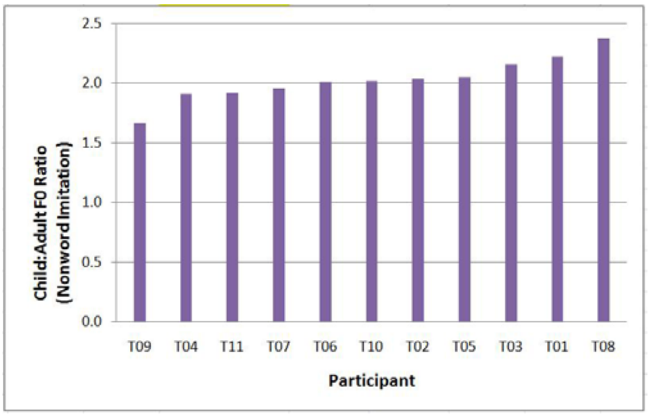

Figure 1: *Mean F0 ratios (child:adult) during the nonword imitation task.*

3.3 Sentence Imitation

During the sentence imitation task, the male adult's voice averaged an F0 of 156 Hz, which is slightly higher than expected for a man in his 20s. As a group, the children imitated the sentences at a mean F0 of 295 Hz, which is close to their mean F0 observed during conversational speech. When the F0 in each imitated sentence was measured and a ratio was calculated between each child's production and the adult's model, as a group the children showed an average ratio of 1.9. Figure 2 shows the distribution of ratios across the participants.

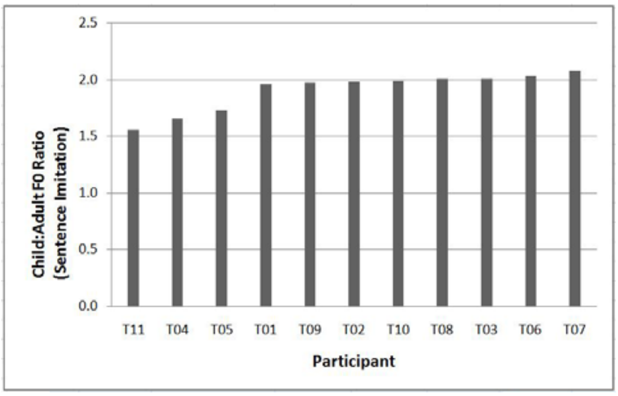

Figure 2: *Mean F0 ratios (child:adult) during the sentence imitation task.*

3.3 Adjustments from Conversational F0 in the Imitations

To determine the extent to which the children adjusted their F0 to match the male voice at one octave above the stimulus, ratios were calculated individually for each child, to compare the mean conversational F0 for each child and her/his mean F0 during each of the imitation tasks. A ratio evaluating to 1 indicates no adjustment of customary F0, while larger or smaller ratios indicate a deviation from customary F0. For the nonword imitation task, conversational/imitation F0 ratios deviated from 1, averaging 1.19 for the group (SD=0.13). For the sentence imitation task, the ratios evaluated to a mean of 0.99 (SD=0.12). Table 2 summarizes these ratios.

Table 2: *Ratios for conversational speech F0 relative to nonword and sentence imitations.*

Participant Code	Conversational F0 / Nonword Imitation F0	Conversational F0 / Sentence Imitation F0
T01	1.07	0.92
T02	1.32	1.03
T03	1.30	1.05
T04	1.25	1.09
T05	1.12	0.99
T06	1.28	0.97
T07	0.98	0.71
T08	1.17	1.06
T09	1.25	1.16
T10	1.36	1.02
T11	0.97	0.88
Mean	1.19	0.99

4. Discussion

The purpose of this study was to address the question whether children imitate stimuli at a vocal pitch one octave above the model when the model is below their vocal range, thus showing octave equivalence in their imitations. Adjusting their voice to twice the F0 in the stimulus would show an awareness of the similarity in the octave relationship with respect to vocal pitch, and would suggest that the children were attempting to use this aspect of their productions to achieve an overall more similar-sounding imitation of the model.

As a group, the participants imitated the stimuli at F0 levels close to one octave above the model in the nonword imitation task and in the sentence imitation task. In both tasks, the group mean for the calculated F0 ratios between the imitations and the model approximated 2.0.

For the sentence imitation task, the male speaker's F0 averaged 156 Hz, which is higher than that expected for adult males during conversational speech. One possible explanation for this pitch level is that the speaker was reading a children's story, where the stimuli were dialogue lines spoken for the mother and children in the story. The speaker may have added the elevated pitch for a more natural sounding rendition of the story. Because the children's conversational mean F0 values averaged 292 Hz, as is consistant with published norms for their ages, it could be argued that imitating the sentences in the story at twice the F0 produced by the male voice approximates their mean conversational F0. When the F0 measurements in the child productions during the sentence imitation task were compared to those in conversational speech, no substantial differences were noted. For this reason, conclusions from these data alone are ambiguous. The results from the nonword imitation task, however, clearly show that the children adjusted their conversational F0 to imitate the stimuli at an F0 one octave higher than the stimuli. The male voice in this task averaged 116 Hz, much lower than half the F0 in the children's conversational samples. The F0 measurements in this task differed substantially from those in conversational speech, indicating that the participants lowered their conversational F0 to imitate the stimuli one octave above the F0 in the model.

The Rhythm subtest in the *Tennessee Test of Rhythm and Intonation Patterns* is designed to assess prosody in the absence of word meaning, where the accuracy of the imitation is measured in terms of the number of syllables and stress patterns compared to the presented model. The sentence imitation task was designed as a measure of expressive language, where performance is assessed in terms of lexical and grammatical accuracy. The participants were instructed to imitate the model and no reference was made to the model's voice. Pitch matching at twice the model's F0 likely occurred without a conscious effort by the children. The observed octave effect in the vocal pitch of the imitations adds an implicit dimension of stimulus-response similarity to these types of tasks not previously observed and discussed in the literature.

Future studies should address the role of linguistic aspects in the robustness of octave equivalence. In this study, we show the octave effect in the presence and in the absence of meaningful linguistic content, although artifacts could not be completely ruled out in the sentence imitation task because the elevated F0 in the model's voice. Replication with lower F0 levels will shed light on the influence of linguistic meaning on octave equivalence in the imitation. In addition, future studies should investigate to what extent children with a variety of communication disorders show the octave equivalence effect described here in a sample of participants with typical development.

5. Conclusions

The results from this study are consistent with the hypothesis that children utilize the perceptual similarity between pitches one octave apart when imitating spoken stimuli, thus achieving an imitation that resembles the stimulus in the additional dimension of pitch levels. While the similarity aspect of octave relationships have been well established in the perception of musical stimuli, this study shows that it is just as salient in speech-related tasks and can be utilized as an added measure of similarity in imitation tasks.

6. Acknowledgments

For their help with the data analysis, we thank James Alyson, Jillian Barrera, Armi Cagingin, Jessica Eickhoff, Gary Hershey, Laura Humphreys, and Tara Miranda Larkin. Special thanks to the children who participated in this study.

The tasks described in this study have been reported elsewhere, but the type of analysis and the results presented here are unique to this paper.

7. References

[1] Burns, E. (1999). Intervals, scales, and tuning. In *The Psychology of Music*, D. Deutsch (ed). San Diego: Academic Press, 215-264.

[2] Demany, L., Armand, F. (1984). The perceptual reality of tone chroma in early infancy. *Journal of the Acoustical Society of America*, 76, 57-66.

[3] Trainor, L.J. (1997). Effect of frequency ratio on infants' and adults' discrimination of simultaneous intervals. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 1427-1438.

[4] Eguchi, I. S., & Hirsh, I.(1969). Development of speech sounds in children. *Acta Otolaryngologica, 2,* 57(Suppl.).

[5] Mysak, E.D., & Hanley, T.D. (1959). Vocal aging. *Geriatrics*, *14*, 652-656.

[6] Hollien, H., & Shipp, T. (1972). Speaking fundamental frequency and chronologic age in males. *Journal of Speech and Hearing Research, 15,* 155-159.

[7] Brazil, D., Coulthard, M., & Johns, C. (1985). *Discourse intonation and language teaching. Applied Linguistics and Language Study.* London: Longman.

[8] Wennerstrom, A.K. (2001). *The music of everyday speech. Prosody and discourse analysis.* Oxford: Oxford University Press.

[9] Peter, B., & Stoel-Gammon, C. (2005). Timing errors in two children with suspected childhood apraxia of speech (sCAS) during speech and music-related tasks. *Clinical Linguistics & Phonetics, 19*(2), 67-87.

[10] Peter, B., & Stoel-Gammon, C. (in press). Central timing deficits in children with primary speech disorders. *Clinical Linguistics & Phonetics.*

[11] Koike, K.J., & Asp, C.W. (1981). Tennessee test of rhythm and intonation patterns. *Journal of Speech and Hearing Disorders, 46,* 81-86.

[12] Wiig, E.H., Secord, W., & Semel, E. (1992). *Clinical Evaluation of Language Fundamentals – Preschool*. San Antonio: The Psychological Corporation.

[13] Boersma, P., & Weenink, D. (2004). *Praat Version 4.2.09*. Institute of Phonetic Sciences, Amsterdam, The Netherlands.

Appendix A. Schematics of Nonword Stimuli

Note: Lowercase letters indicate unstressed syllables, uppercase letters indicate stressed syllables.

Item Number	Rhythm Pattern
1	MA.ma
2	ma.MA
3	MA.ma.MA
4	ma.MA.ma
5	MA.ma.ma
6	MA.ma.MA.ma
7	ma.MA.MA.ma
8	ma.ma.MA.ma
9	ma.MA.ma.ma
10	ma.MA.ma.MA
11	MA.ma.ma.MA.ma.ma
12	ma.MA.ma.ma.MA.ma