Octave-shifted pitch matching in nonword imitations: The effects of lexical stress and speech sound disorder (L)^{a)}

Beate Peter,^{b)} Tara Larkin, and Carol Stoel-Gammon

Department of Speech and Hearing Sciences, University of Washington, P.O. Box 354875, Seattle, Washington 98105

(Received 2 April 2009; revised 10 July 2009; accepted 21 July 2009)

Perceptual similarities of musical tones separated by octave intervals are known as octave equivalence (OE). Peter *et al.* [(2008). Proceedings of the Fourth Conference on Speech Prosody, edited S. Maduerira, C. Reis, and P. Barbosa, Luso-Brazilian Association of Speech Sciences, Campinas, pp. 731–734] found evidence of octave-shifted pitch matching (OSPM) in children during verbal imitation tasks, implying OE in speech tokens. This study evaluated the role of lexical stress and speech sound disorder (SSD) in OSPM. Eleven children with SSD and 11 controls imitated low-pitched nonwords. Stimulus/response f_0 ratios were computed. OSPM was expressed preferentially in stressed vowels. SSD was associated with reduced expression of OSPM in unstressed vowels only. Results are consistent with the psycholinguistic prominence of lexical stress and prosodic deficits in SSD. © 2009 Acoustical Society of America. [DOI: 10.1121/1.3203993]

PACS number(s): 43.66.Hg, 43.70.Fq, 43.70.Dn, 43.66.Lj [DD]

Pages: 1663–1666

I. INTRODUCTION

The perceptual similarity between two musical tones separated by one or more octave intervals is referred to as octave equivalence (OE). Musical tuning systems in different cultures around the world vary widely in the way scales are arranged, yet all utilize octave-based perceptual similarities (Burns and Ward, 1978). Psychophysical evidence for OE has been found in a small number of studies in adults and infants, and even animals, although results have been somewhat controversial (Burns, 1999).

The biological basis for perceptual OE in musical tones is unclear. Several theoretical approaches (reviewed in Burns, 1999) have attempted to account for OE, including the place code theory, which posits that octave-separated tones activate the same place on the basilar membrane in the cochlea because they share harmonic frequencies. Langner (2007) pointed out that neurons in the inferior colliculus respond not only to characteristic frequencies but also to their integer multiples.

While OE has been documented in the music literature, it is unknown whether it exists in speech signals as well. Evidence of pitch matching at an octave interval in a spoken imitation task would imply the presence of perceptual similarity in the speech tokens. To evaluate pitch matching in imitation tasks, the vocal pitch levels in the target speaker and the respondent must be considered because these vary by age and gender. In 4- and 5-year-old children, mean fundamental frequency (f_0) levels were estimated at 286 and 289 Hz, respectively (Eguchi and Hirsh, 1969). Typical f_0 mean values for male and female adults between 25 and 50 years

range between 100 and 120 Hz and between 240 and 245 Hz, respectively; with advancing age, f_0 tends to rise in men and to drop slightly in women (Hollien and Shipp, 1972; Mysak and Hanley, 1959; Saxman and Burk, 1967).

In conversational speech, f_0 adjustments have been described as a function of discourse mechanics, for instance, in the form of pitch matching between speakers to convey agreement (Brazil *et al.*, 1985; Wennerstrom, 2001). Pitch is usually described relative to a speaker's vocal range (e.g., high, mid, or low key), not physical scales. In the context of an imitation task, Peter *et al.* (2008) unexpectedly found evidence of octave-shifted pitch matching (OSPM). They analyzed pitch in a study designed originally to evaluate the role of temporal aspects of speech and hand tasks in children (Peter and Stoel-Gammon, 2008). In two imitation tasks, nonwords and sentences, where the target voice was an adult male's, the participants tended to imitate the tokens one octave above the target level.

The purpose of the present study was to evaluate the role of lexical stress and speech sound disorder (SSD) in the expression of OSPM in a nonword imitation task. Stressed syllables appear to be associated with psycholinguistic prominence. Newborn infants demonstrate an awareness of language-specific stress patterns (reviewed in Echols *et al.*, 1997) and the ability to use stressed syllables as cues to segment words (Jusczyk *et al.*, 1993). Young children tend to omit unstressed syllables more frequently than stressed syllables (Ingram, 1974; Kehoe and Stoel-Gammon, 1997). In individuals who stutter, stressed syllables are affected predominantly (Hahn, 1942). Some children with SSD have difficulty speaking with appropriate prosody, which includes not only temporal aspects such as vowel durations and intensity levels but also aspects related to vocal pitch (Shriberg *et al.*, 2003). Lower expression of OSPM is therefore predicted in unstressed syllables and in participants with SSD.

^{a)} Aspects of this study were presented at the Fourth Conference on Speech Prosody, Campinas, Brazil, May 6–9, 2008 and at the Child Phonology Conference, Purdue University, June 2–3, 2008.

^{b)} Author to whom correspondence should be addressed. Also at Department of Medicine, Division of Medical Genetics, University of Washington, P.O. Box 357720, Seattle, WA 98195. Electronic mail: bvpeter@u.washington.edu

TABLE I. Schematics of stimuli in the nonword imitation task. Note: Capital letters indicate stressed syllables, lower-case letters indicate unstressed syllables, and periods mark syllable boundaries. Items 13 and 14 contain pauses and were not administered.

Item no.	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
13	MA.[pause]MA.ma
14	MA.ma.[pause]MA

II. PARTICIPANTS, PROCEDURES, AND RELIABILITY

This study was conducted with the approval of the University of Washington Human Subjects Division and in compliance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Eleven children, ages 4 years and 7 months to 6 years and 6 months, with a diagnosis of a moderate to severe SSD of unknown origin and 11 age- and gender-matched controls with typical development (TD) participated in this study. All participants were required to fulfill the following criteria: age 4 years and 6 months to 7 years; monolingual English home environment; hearing screen passed at 25 dB sound pressure level (SPL); no history of any developmental or acquired disorder interfering with hearing, speech fluency, or neuropsychologic functioning. The participants with SSD additionally were required to have a diagnosis of moderate to severe speech disorder of unknown origin, documented with standardized speech test (articulation and/or phonology) scores of 1.5 standard deviations below the mean or lower.

The imitation task was based on the “Tennessee test of rhythm and intonation patterns” (T-TRIP; Koike and Asp, 1981), a nonstandardized test of prosody. During the Rhythm subtest, the participant hears 14 prerecorded sequences of the syllable “ma” and imitates them. Items 13 and 14 were excluded from analysis because they contain pauses. Items 1–12 contain a total of 45 syllables, of which 43, 18 stressed and 25 unstressed, were available for analysis; a reliable pitch measurement could not be obtained for two of the modeled syllables. The voice on the CD provided by the manufacturer was a male adult, with f_0 in the stressed vowels averaging 133 Hz and, in the unstressed vowels, 105 Hz (overall average f_0 =116 Hz). Table I shows the metric structures of the test items.

Customary f_0 measures were measured in 5–12 utterances obtained from conversational speech samples collected during the study sessions. Acoustic analyses were performed by a team consisting of the first author and seven under-

graduate or postbaccalaureate students in the Department of Speech and Hearing Sciences at the University of Washington using PRAAT, Version 4.2.09 (Boersma and Weenink, 2004) or more recent versions, measuring f_0 in Hz.

The octave equivalence range (OER) was defined as any pitch that fell within one semitone above or below twice the stimulus frequency. As the adult targets were, on average 116 Hz, the average OER was 220–247 Hz. For individual vowel imitations, a child/adult f_0 ratio was calculated. Ratios between 1.887:1 and 2.119:1 represented the range within one semitone of twice the adult target’s f_0 level, following the formula

$$\text{OER} = 2(f_0 2^{-1/12}), 2(f_0 2^{1/12}). \quad (1)$$

Statistical analyses included paired t tests, t tests with adjustments for unequal variances, and the nonparametric Wilcoxon signed-rank test. The acoustic measurements were checked for reliability by remeasuring approximately 15% of the tokens. The average discrepancy was 5.36 Hz for vowels and 5.05 Hz for the conversational samples. This was considered acceptable for the purposes of this study.

III. RESULTS

As a group, the participants with SSD produced an average of 35.6 syllables with correctly placed stress out of the 43 available model syllables (N =368). The participants with TD produced an average of 40.7 syllables with correctly placed stress (N =422). The participants with SSD had an average conversational f_0 of 285 Hz [standard deviation (SD)=42 Hz], and the participants with TD averaged 292 Hz (SD=43 Hz), which closely approximates conversational values expected for the participants’ ages. Table II summarizes participant characteristics and f_0 measures.

107 of 421 (25%) unstressed and 160 of 368 (43%) stressed imitated vowels fell within the OER. Differences were statistically significant (paired t test using per-participant averages: t =9.11, p <0.0001; t test with adjustments for unequal variances using all vowel measurements: t =14.16, p <0.0001).

The SSD group included two participants whose average conversational f_0 levels fell within the defined OER. Of the remaining nine, whose conversational f_0 levels all fell above that range, one participant downshifted the conversational f_0 levels to the OER in the imitations; four participants lowered their conversational f_0 levels by at least 32 Hz without reaching the OER, and four did not adjust their conversational f_0 levels appreciably during the imitation task (difference $\leq$ 8 Hz). The TD group included one participant with both conversational and imitated f_0 levels within the OER, five participants who downshifted their conversational f_0 levels to fall within the OER during the imitations, four who lowered their conversational f_0 levels substantially during the imitations without reaching the OER, and one participant who did not adjust the conversational f_0 levels appreciably. Figure 1 shows conversational and imitated f_0 values for each participant, in relationship to the adult target and the OER.

124 of 368 (34%) vowels in the SSD group and 143 of 422 (34%) in the TD group fell into the OER. Nonparametric

TABLE II. Summary of participant gender, age, mean conversational and imitated f_0 levels, and child/adult f_0 ratios. Child/adult f_0 ratios within one semitone of twice the stimulus f_0 levels (range: 1.888, 2.119) are indicated with a *. [Table adapted with permission from Peter and Stoel-Gammon (2008).]

Code	Gender	Age in months	Conversational f_0	Imitated f_0	Mean C/A f_0 ratio	C/A f_0 ratio (unstressed syllables)	C/A f_0 ratio (stressed syllables)
D01	M	73	222	224	*1.933	*2.103	1.744
D02	M	60	288	248	*2.095	2.246	*1.961
D03	M	72	256	259	2.260	2.531	*1.930
D04	M	67	269	268	2.323	2.681	*1.988
D05	M	78	238	254	2.133	2.266	*1.990
D06	M	61	318	286	2.453	2.712	*2.116
D07	F	67	283	223	*1.929	*2.000	1.842
D08	M	55	313	268	2.224	2.421	*2.062
D09	F	59	355	262	2.251	2.398	*2.081
D10	F	59	331	328	2.806	2.993	2.596
D11	M	76	258	251	2.157	2.340	*1.934
T01	M	74	283	265	2.230	2.244	2.216
T02	M	61	312	236	*2.040	2.171	1.867
T03	M	72	325	249	2.164	2.369	*1.895
T04	M	63	279	223	*1.915	*1.956	1.869
T05	M	81	265	235	*2.054	2.216	1.839
T06	M	61	303	238	*2.018	2.108	*1.918
T07	F	73	229	232	*1.961	*2.064	1.857
T08	M	58	330	282	2.375	2.480	2.245
T09	F	58	353	283	2.391	2.650	2.170
T10	F	59	316	233	*2.002	2.131	*1.877
T11	M	76	214	220	*1.925	*1.988	1.838

testing for paired data, inputting the per-participant mean scores, did not show statistically significant SSD-TD group differences ($z=1.16$, $p=0.2477$). Parametric testing based on t testing for matched data, inputting the child/adult f_0 ratios from all analyzed vowels, showed a significant group difference ($t=6.24$, $p<0.0001$). Both participant groups deviated from the 2:1 ratio in both directions, but the distributions were skewed toward overshoots in the direction of the participants' conversational f_0 levels. The distribution in the SSD group showed a wider range and greater variability compared to the TD group.

Because OSPM was more robust in the stressed than in the unstressed vowels, the two groups were compared using

data from the two stress types separately. For the unstressed vowels, the SSD group produced 38 of the 193 vowels (20%) within the OER, whereas the TD group produced 69 of the 228 vowels (30%) in the OER. For the stressed vowels, the SSD group produced 86 of 174 vowels (49%) within the OER, whereas the TD group produced 74 of the 194 vowels (38%) in the OER. Nonparametric testing failed to detect statistically significant differences between the SSD and TD groups in the unstressed vowels ($z=1.43$, $p=0.1549$) and in the stressed vowels ($z=0.98$, $p=0.3281$). Parametric testing showed a statistically significant difference between the SSD and TD groups for both the unstressed ($t=6.38$, $p<0.0001$)

FIG. 1. Target adult f_0 , double target adult f_0 , upper and lower bounds of the OER, and mean conversational and imitated f_0 for each participant. Data points shown separately for each participant group, sorted by conversational f_0 level.

FIG. 2. Distributions of child/adult f_0 ratios by participant group and stress type. Reference lines bracket the OER.

and stressed ($t=2.01$, $p=0.0186$) vowels. Figure 2 shows the distributions of child/adult f_0 ratios for the two participant groups and stress types.

IV. SUMMARY AND DISCUSSION

This study addresses the role of lexical stress and presence of SSD in OSPM in imitated speech tokens, produced by 11 children with SSD and 11 controls. A shift from conversational f_0 levels to double the adult's f_0 level, as opposed to no pitch adjustment or direct pitch matching toward the actual adult f_0 level, is supported by several observations. First, none of the three participants whose conversational f_0 levels fell within one semitone of the doubled adult's f_0 levels dropped f_0 levels during the imitation task. This indicates that direct pitch matching at the adult's actual f_0 levels did not occur in these participants. Second, seven participants lowered their conversational f_0 levels to fall within one semitone of the doubled adult's f_0 levels. Third, the distributions of child/adult f_0 ratios in both participant groups show peaks near a ratio of 2:1, which would not necessarily be expected where direct pitch matching was attempted or no pitch matching occurred. Whether or not the eight participants who lowered their f_0 levels substantially but failed to reach the OER were aiming for this range or for the direct target pitch cannot be ascertained.

OSPM was preferentially expressed in stressed vowels. The distributions of the child/adult f_0 ratios showed that participants tended to lower their f_0 to a level equivalent to twice the targets' f_0 when the syllable was stressed while speaking closer to their conversational f_0 when the syllables were unstressed. This result is consistent with the literature on the psycholinguistic prominence of lexical stress and introduces a new manifestation of this phenomenon.

The participants with SSD did not adjust their conversational f_0 levels in the direction of twice the stimulus f_0 levels as frequently as their peers with TD. The distributions of the child/adult f_0 ratios show that this group difference resulted from a difference in spread, not average values, with the imitations in the SSD group being more variable. The observation that nonparametric testing did not show the group differences that were evident in parametric testing is consistent with this outcome. An important observation resulted from considering the joint effects of SSD and lexical stress. In the stressed vowels, the SSD group reached the OER slightly more frequently than the TD group, while showing much less evidence of OSPM in the unstressed vowels. This is consistent with the finding that children with SSD tend to omit unstressed syllables more frequently than peers with TD and allows the interpretation that unstressed syllables have even less psycholinguistic prominence in SSD than in TD.

An established literature on the perceptual aspects of OE in speech tokens and on direct pitch matching and OSPM in imitation tasks does not yet exist. Therefore, future studies should be conducted to corroborate these experimental results and to evaluate a number of hypotheses in speech-related tasks. They should address (1) the perceptual salience of octave-related speech stimuli, which are implied in this study but were not explicitly quantified; (2) direct pitch

matching in contexts where the target f_0 is within the participants' conversational f_0 range but offset from their average f_0 values, for instance, in an imitation task where the target speaker is an adult female and the participants are children; (3) a variety of speech stimuli beyond nonwords, for instance, single words of varying phonologic complexity, pseudowords with varied word and syllable structures, and story retelling; and (4) other types of communication deficits seen in children with autism and other psychiatric disorders where prosody is affected, and the speech disfluencies in children who stutter.

ACKNOWLEDGMENTS

This project represents part of the first author's dissertation study, funded by an institutional training grant from NIDCD (Grant No. 05 T32 DC00033-10) and the Department of Speech and Hearing Sciences at the University of Washington.

- Boersma, P., and Weenink, D. (2004). PRAAT, Version 4.2.09, Institute of Phonetic Sciences, Amsterdam, The Netherlands.
- Brazil, D., Coulthard, M., and Johns, C. (1985). *Discourse Intonation and Language Teaching: Applied Linguistics and Language Study* (Longman, London).
- Burns, E. (1999). "Intervals, scales, and tuning," in *The Psychology of Music*, edited by D. Deutsch (Academic, San Diego), pp. 215–264.
- Burns, E. M., and Ward, W. D. (1978). "Categorical perception—Phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals," *J. Acoust. Soc. Am.* **63**, 456–468.
- Echols, C., Crowhurst, M., and Childers, J. (1997). "The perception of rhythmic units in speech discrimination in infants," *J. Mem. Lang.* **36**, 202–225.
- Eguchi, I. S., and Hirsh, I. (1969). "Development of speech sounds in children," *Acta Oto-Laryngol., Suppl.* **2**, 57.
- Hahn, E. (1942). "A study of the relationship between stuttering occurrence and phonetic factors in oral reading," *J. Speech Hear. Disord.* **7**, 143–151.
- Hollien, H., and Shipp, T. (1972). "Speaking fundamental frequency and chronologic age in males," *J. Speech Hear. Res.* **15**, 155–159.
- Ingram, D. (1974). "Phonological rules in young children," *J. Child Lang.* **1**, 49–64.
- Jusczyk, P., Cutler, A., and Redanz, N. (1993). "Infants' preference for the predominant stress patterns of English words," *Child Dev.* **64**, 675–687.
- Kehoe, M. M., and Stoel-Gammon, C. (1997). "Truncation patterns in English-speaking children's word productions," *J. Speech Lang. Hear. Res.* **40**, 526–541.
- Koike, K. J., and Asp, C. W. (1981). "Tennessee test of rhythm and intonation patterns," *J. Speech Hear. Disord.* **46**, 81–86.
- Langner, G. (2007). "Die zeitliche Verarbeitung periodischer Signale im Hörsystem: Neuronale Repräsentation von Tonhöhe, Klang und Harmonizität (Temporal processing of periodic signals in the auditory system: Neuronal representation of pitch, timbre, and harmonicity)," *Z. f. Audiologie, Audiological Acoust.* **46**, 8–21.
- Mysak, E. D., and Hanley, T. D. (1959). "Vocal aging," *Geriatrics* **14**, 652–656.
- Peter, B., and Stoel-Gammon, C. (2008). "Central timing deficits in children with primary speech disorders," *Clin. Linguist. Phon.* **22**, 171–198.
- Peter, B., Stoel-Gammon, C., and Kim, D. (2008). "Octave equivalence as a measure of stimulus-response similarity during nonword and sentence imitations in young children," in *Proceedings of the Fourth Conference on Speech Prosody*, edited S. Maduerira, C. Reis, and P. Barbosa, Luso-Brazilian Association of Speech Sciences, Campinas, pp. 731–734.
- Saxman, J. H., and Burk, K. W. (1967). "Speaking fundamental frequency characteristics of middle-aged females," *Folia Phoniatr. (Basel)* **19**, 167–172.
- Shriberg, L. D., Campbell, T. F., Karlsson, H. B., Brown, R. L., McSweeney, J. L., and Nadler, C. J. (2003). "A diagnostic marker for childhood apraxia of speech: The lexical stress ratio," *Clin. Linguist. Phonetics* **17**, 549–574.
- Wennerstrom, A. K. (2001). *The Music of Everyday Speech: Prosody and Discourse Analysis* (Oxford University Press, Oxford).