Published in final edited form as:

J Voice. 2015 March ; 29(2): 260.e21–260.e30. doi:10.1016/j.jvoice.2014.06.011.

Direct and octave-shifted pitch matching during nonword imitations in men, women, and children

Beate Peter*, Bronsyn Foster*, Heather Haas*, Kyle Middleton, and Kiersten McKibben

Dpt. of Speech and Hearing Sciences, University of Washington

Summary

Objectives—To evaluate whether children, women, and men match the speaker’s fundamental frequency (F0) during nonword imitation directly when the target F0 is within the responders’ vocal ranges and at octave-shifted levels when the target is outside their vocal ranges. To evaluate the role of a history of speech sound disorder (SSD) in the adult participants.

Study Design—Observational.

Methods—Nonword sets spoken by a man and a woman were imitated by 14 men, 21 women, and 19 children. Approximately half of the adults and two thirds of the children had a history of SSD. F0 in the imitations was compared to that in the targets and in the participants’ non-imitated control word productions.

Results—When the target F0 was within the responders’ vocal ranges, the imitations approximated the target F0. Men imitating a woman’s voice approximated F0 levels one octave below the target F0. Children imitating a man’s voice approximated F0 levels one octave above the target F0. Women imitating a man’s voice approximated the target F0 at a ratio of 1.5, known as the perfect fifth in music. A history of SSD did not influence the results.

Conclusions—This study replicates previous findings showing that target F0 was a salient aspect of the stimuli that was imitated along with the targets’ segmental and prosodic components without explicit prompting. It is the first to show F0 convergence not only directly but also at relevant target/imitation intervals including the octave interval.

Keywords

Phonetic convergence; nonword imitation; direct pitch matching; octave equivalence; octave-shifted pitch matching; perfect fifth; speech sound disorder

© 2014 The Voice Foundation. Published by Mosby, Inc. All rights reserved.

Corresponding author: Beate Peter, Ph.D., CCC-SLP, Department of Speech and Hearing Sciences, Box 354875, University of Washington, Seattle WA 98195, bvpeter@u.washington.edu, Telephone: 206 713 5839, Fax: 206 543 1093.

*These authors contributed equally to this work

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

INTRODUCTION

During verbal interaction, speakers adjust various aspects of their speech to assimilate to the speech of their interlocutors. This phenomenon, as recently reviewed [1–3], has been described regarding a variety of aspects of communication including pragmatic, lexical, and syntactic traits. Examples of phonetic assimilations, also termed convergence, accommodation, entrainment, alignment, or chameleon effect, have been described not only at the segmental level [4] but also regarding suprasegmental features such as speech rate and intonation [5, 6]. In a vowel imitation experiment, speakers imitated the fundamental frequency (F0; also referred to as pitch) in the tokens without explicit instructions to do so [2], showing that F0 was a salient trait of the target that was imitated along with phonetic properties. Here we replicate this type of F0 convergence in samples of men, women, and children imitating nonwords. We expand the focus from direct pitch matching to also include octave-shifted pitch matching.

In the music literature, pitch matching of tones has been the object of several studies, for instance investigating the distinct roles of perception versus production in pitch matching tasks [7], various types of stimuli [8], neural substrates of absolute pitch [9], the role of musical training [10], and extreme deficits in pitch perception and production known as “tone deafness” [11]. In music, two tones separated by a frequency interval corresponding to ratios between low integers, e.g., 3:2 (“perfect fifth”) and 4:3 (“perfect fourth”), are perceived as more harmonious than intervals with ratios involving higher integers, e.g., 7:8 [12, 13]. Two tones related via the simplest ratio, 2:1 (“octave”), are perceived as highly similar (“octave equivalence”) and this perceptual similarity is utilized in the various musical scales around the world [14]. Children’s abilities to perceive and reproduce the octave-based similarity in music have not been studied extensively. In one study evaluating the ability of first grade children to perceive “octave-transfer of pitch” in singing classes when the model was an adult male, children who could directly pitch-match a model within their vocal range could also pitch-match a model below their vocal range indirectly at an octave above the target; conversely, children who were unable to match pitches directly were also unable to match pitches indirectly at octave intervals [15]. These findings suggest that musical pitch matching abilities govern both direct and octave-shifted targets and vary among individuals.

Nonword imitation tasks are routinely used to assess phonological processing skills. Accurate imitation of phoneme sequences that follow the phonotactics of English in the absence of semantic content is interpreted as evidence that the responder successfully perceived the phoneme sequence, stored it in short-term memory, retrieved it from there, and converted it into a spoken phoneme sequence via the speech production system. As extensively reviewed [16, 17], difficulties with this task characterize children with certain types of speech sound disorder (SSD). SSD is defined as a childhood disorder interfering with the ability to produce speech that is easily understood by others due to distorted, substituted, omitted, or inserted speech sounds. For clinical and research purposes, imitated nonwords are evaluated for accuracy by comparing their phoneme sequences to those in the target. Lexical stress errors are evaluated in only very few tests of nonword imitation, for

instance the Tennessee Test of Rhythm and Intonation Patterns (T-TRIP) [18]. F0 is rarely evaluated during clinical nonword imitation testing.

We recently showed that 4- to 6-year-old children adjusted their conversational F0 to produce the vowels in a set of multisyllabic nonwords one octave above the target, which was far below their vocal ranges [19]. To our knowledge, this was the first study of octave equivalence in speech-like rather than music-like stimuli. In contrast to most studies of music pitch matching, the children had not been told that they should match the model's voice, only that they should say the words they heard. The sample included children with and without SSD. Both groups demonstrated octave shifting during stressed syllables. During unstressed syllables, the children with SSD octave-shifted to a lesser extent than their peers without SSD. In general, these findings imply that the children perceived the octave equivalence in the speech-like tokens even though their perception was not directly tested. The study left many questions unanswered. For instance, it was unknown whether children match F0 in speech-like tokens directly when the token is within their vocal ranges. Similarly, it was unknown whether men and women match F0 in speech-like tokens directly when the token is within their vocal ranges and whether they match F0 at an octave-shifted frequency when the token is outside their vocal ranges. These questions form the hypotheses addressed in the present study. In addition, we investigate the influence of present or past SSD on pitch matching.

METHOD

Participants

Data for this study came from a multigenerational family genetics project investigating the molecular genetics of SSD [20–23]. The study was conducted with the approval of the University of Washington's Human Subjects Division. Adults gave written consent to participate, parents consented for their minor children, and school-age children and adolescents gave assent. For this study, data were available for 54 participants from 6 different families with an age range of 3 to 80 years (14 men, 21 women, and 19 children defined here as age 13 years or younger). Thirteen children had a present or past SSD diagnosis and 8 women and 8 men reported a history of SSD. Prior to the study, a hearing screening was administered whenever possible and participants passed the screening at 25 dB SPL at .5, 1, 2, and 4 kHz. All participants were native speakers of English. Aside from SSD, there was no history of any developmental or acquired disorder in any of the participants.

Tasks

Participants were asked to imitate nonword targets as accurately as possible. No mention of vocal pitch was made. Two tests of nonword imitation were selected for this study and stimuli were presented via a laptop computer and external speakers. The Syllable Repetition Test (SRT) [24] is an 18-item standardized imitation task designed to measure phonemic awareness in children with or without SSD. The phonemes used in the SRT nonwords are the early developing voiced consonants /b, d, m, n/ and the vowel /a/. All syllables have a consonant-vowel (CV) structure and word shapes consist of 2 syllables (8 items), 3 syllables

(6 items), or 4 syllables (3 items), all with trochaic rhythmic patterns. The prerecorded targets were spoken by an adult female with an F0 averaging 194.8 Hz (SD = 6.8 Hz) during the 28 stressed syllables and 165.5 Hz (SD = 11.2 Hz) during the 22 unstressed syllables. This indicates that the stressed syllables were approximately 2.75 semitones (ST) above the unstressed syllables, an F0 ratio of 1.18 ($t = 10.23$; $p < .0001$). The Tennessee Test of Rhythm and Intonation Patterns (T-TRIP) [18] is a non-standardized test of prosody. The *Rhythm* subtest requires imitation of 14 prerecorded sequences of the syllable “ma.” Items 13 and 14 were omitted from the analysis because they contain pauses, causing a premature response in many participants. Word shapes in items 1 through 12 range in complexity from 2 to 6 CV syllables with varying rhythmic patterns. Each item was administered twice but only the first imitation was analyzed unless the second imitation provided a more accurate stress pattern in terms number and stress type of the syllables. The targets were spoken by an adult male with an F0 averaging 128.6 Hz (SD = 1.4 Hz) during the 18 stressed syllables and 104.5 Hz (SD = 1.5 Hz) during the 27 unstressed syllables, with a mean F0 in all 45 syllables of 114 Hz. The stressed syllables were approximately 3.5 ST above the unstressed syllables, a ratio of 1.23 ($t = 11.63$; $p < .0001$). F0 levels in the two tests were mutually exclusive in that the stressed vowels in the SRT and the T-TRIP targets differed by 7.2 ST with no overlaps between the two ranges. Similarly, the unstressed vowels in these two tests differed by 8 ST and there were no F0 overlaps. Additional details are shown in Table 1 below.

Together, the three participant groups (men, women, children) and two target types (a man’s voice, a woman’s voice) provided the opportunity for six experiments. Three opportunities to measure direct pitch matching arose from men imitating the man’s voice in the T-TRIP task and women and children imitating the woman’s voice in the SRT task. Conversely, three opportunities to observe indirect, i.e., octave-shifted pitch matching arose from men imitating the woman’s voice at an octave below the targets and from women and children imitating the man’s voice at an octave above the target. Not all participants completed both tasks, so that group sizes were not necessarily equal for the two tasks.

To determine whether imitated F0 levels differed significantly from F0 levels in a task not involving imitation, participants were asked to complete the Goldman-Fristoe Test of Articulation – 2 (GFTA-2) [25]. The GFTA-2 was designed to elicit word productions using picture stimuli, not spoken models, for purposes of analyzing accuracy of speech sound productions in the context of assessing presence and severity of speech sound disorders in children. Here, the GFTA-2 was used to obtain control F0 measurements in non-imitated word productions. Because the stimuli in the SRT and T-TRIP consisted of two or more syllables, only multisyllabic words were selected, a total of 10 words. These words resembled the SRT and T-TRIP nonwords in terms of simple CV syllable shapes, for instance in “banana,” “telephone,” “shovel,” and “fishing.”

Data Analysis

A team consisting of all authors and students in the University of Washington Department of Speech and Hearing Sciences performed acoustic analyses of the recorded data using PRAAT, version 5.3.02 [26]. For the purposes of this study, all word-final vowels were

dropped from the analysis because their endpoints could not be determined unambiguously in open syllables and also because several participants upshifted F0 on the final syllable, consistent with a questioning intonation pattern. Vowel F0 was measured in a 50 msec window in the center of each vowel. Approximately 15% of the data was re-measured by the first and second author for reliability. Average discrepancies were 1 Hz for the SRT, 2 Hz for the T-TRIP, and 3 Hz for the GFTA-2. Discrepancies of more than 5 Hz were measured a third time and resolved by consensus. This level of accuracy was judged sufficient for the purposes of this study. Because the data within each group and task type were nested by individual participant and, hence, not independent for purposes of statistical analyses, average F0 levels were calculated for each participant.

To determine whether participants imitated the target F0 either directly or at the F0 up- or downshifts predicted by our hypotheses, data were analyzed with respect to individual participants as well as per-group F0 levels. Specifically, we investigated the following research questions:

1. Did the participants' imitations fall within 1 ST (strong evidence) or 2 ST (moderate evidence) of the direct or octave-shifted target F0?
2. Was there a significant difference in F0 levels in the imitations during the T-TRIP and the SRT in each participant group?
3. Did the imitated F0 levels scatter about the target F0 more closely than about non-imitated control F0 levels in each group?
4. Did a history of SSD influence the degree of F0 convergence?

Answers to these questions together provide a comprehensive view of pitch matching in the imitation tasks. If the F0 target was far from a participant's comfortable vocal range, as estimated by the non-imitated control word productions in the GFTA-2, the imitations may not fall within a narrow ST window of the target, yet an attempt to approximate the target may be evident in how the F0 levels in the two imitation tasks differed from each other and how closely the imitated F0 scattered about the target, relative to non-imitated control F0. Conversely, if the target F0 was close to a participant's F0 in the non-imitated control task, evidence that the F0 targets were matched was based on whether or not the imitations fell within 1 or 2 ST of the target and whether there was a difference between the F0 levels during the T-TRIP and the SRT, whereas an adjustment from control F0 to target F0 was not necessarily expected in this case. It should be noted that in a mixture of stressed and unstressed syllables, each of which carry statistically significantly different F0 levels, pitch matching within 1 or 2 ST of the targets is more remarkable than if the task involved only one type of stress. Furthermore, participants whose imitations fell within the target ranges in both nonword imitation tests provide especially strong evidence of pitch matching where the target ranges in the two tasks have minimal or no physical overlaps.

To address Research Question 1, an imitation/target F0 ratio was calculated for each imitated vowel and averaged separately for each participant and nonword task type. A 1:1 ratio indicated a perfect direct pitch match and a 2:1 ratio or a 0.5:1 ratio indicated a perfect octave relationship between the target and the imitation. Ratios between .94 and 1.06 fall

within 1 ST of a direct pitch match; ratios between .89 and 1.12 fall within 2 ST of a direct pitch match. Analogously, ratios between 1.89 and 2.12 fall within 1 ST of a pitch match an octave above the target (between 1.78 and 2.24 for a 2 ST interval), and ratios between .47 and .53 fall within 1 ST of a pitch match an octave below the target (between .45 and .56 for a 2 ST interval). For the purposes of this study, the following five ST bands relative to the target F0 were defined: below 2 ST (< -2 ST), lower band of within 2 ST (> -2 ST, < -1 ST), within 1 ST (± 1 ST), upper band of within 2 ST (> 1 ST, < 2 ST) and above two ST (> 2 ST). Imitations that fell outside of 2 ST of the target were interpreted as unmatched F0. Imitations within 2 ST but outside 1 ST of the target were interpreted as moderate evidence of pitch matching, and imitations that fell within 1 ST of the target provided strong evidence of pitch matching. The wider ± 2 ST range was based on a published experiment in music pitch matching where, in a group of untrained singers, musical tones were matched with an average target-imitation difference of 1.68 ST, where individual performance was widely variable [10]. The narrow ± 1 ST range was defined following methods described in the pitch matching literature in music where highly accurate imitations were defined as within 1 ST of the target [27].

Table 1 summarizes the ± 1 ST and ± 2 ST F0 ranges in Hz values for each nonword imitation task, separately for stressed and unstressed non-final vowels. Regarding the narrow ± 1 ST ranges, there were no physical overlaps between direct ranges in one test and octave-shifted ranges in the other test for stressed vowels. Based on our findings described below, a shift for the T-TRIP targets by a ratio of 1.5 was considered here as well. There was a 21 Hz overlap between stressed SRT vowels and stressed T-TRIP vowels shifted by a ratio of 1.5. Even considering the wider ranges based on ± 2 ST, the physical overlaps were minimal. The direct T-TRIP range did not overlap with the octave-shifted SRT for the stressed or unstressed vowels, and there were no overlaps between the direct SRT and octave-shifted T-TRIP ranges. Stressed SRT vowels overlapped with T-TRIP vowels shifted by a 1.5 ratio by 43.7. Unstressed SRT vowels overlapped with octave-shifted T-TRIP vowels by 13.2 Hz and, with T-TRIP vowels shifted by a 1.5 ratio, by 18.8 Hz. It is therefore extremely unlikely that participants would produce imitated F0 levels within these target ranges by chance except for the T-TRIP targets shifted by a 1.5 ratio.

To address Research Questions 2 and 3 asking whether there were overall statistically significant differences between the imitated F0 levels in the T-TRIP and the SRT and the non-imitated control F0 levels from the GFTA-2, ANOVA testing for repeated measures was conducted. To investigate differences between each pair of tests, differences were assessed with nonparametric two-tailed z tests for matched pairs. Our hypothesis predicted that F0 levels during the SRT and T-TRIP differed significantly because of the pitch matching effect. Differences between the non-imitated control F0 and each of the tests of nonword imitation were predicted to reflect the magnitude of an adjustment of the participants' non-imitated control F0 in the direction of the target F0. Lack of differences between targets and imitations, as assessed with nonparametric two-tailed z tests for matched pairs, was interpreted as additional evidence of pitch matching. For these statistical procedures, only vowels from stressed non-final syllables were used. For the T-TRIP, 14 vowels (target mean Hz = 130.9, SD = 4.02) were available and for the SRT, 28 vowels

(target mean Hz = 194.7, SD = 6.7). The selected GFTA-2 words to generate non-imitated control F0 measures provided 10 vowels from stressed non-final syllables.

To address Research Question 4 regarding the role of SSD, nonparametric rank-sum tests between SSD affected and unaffected participants with respect to imitation/target F0 ratios were conducted. Nonparametric tests were selected instead of parametric *t* tests because of the small sample sizes, as within each group and task, subgroups of individuals with, and without, a history of SSD were compared to each other.

RESULTS

Research Question 1: Imitations Within 1 ST and 2 ST of the Direct and Octave-Shifted Targets

Table 2 lists the number of participants whose mean imitation/target F0 ratios fell into each of the ST bands for each nonword task and participant group. These counts were based on the weighted combination of stressed and unstressed syllables excluding final syllables. Regarding direct pitch matching, the men produced a mean imitation/target ratio = .92 (SD = .08), which included one low outlier at .73. Here, 50% of the men produced imitated F0 levels within 1 ST of the T-TRIP targets and 71%, within 2 ST of these targets. For the women, the mean imitation/target ratio was .94 (SD = .09). Only 19% of the women produced F0 levels within 1 ST of the SRT targets but 81% produced F0 levels within 2 ST of the SRT targets, where 52% of the mean F0 values fell between 2 ST and 1 ST below the target F0. For the children, the mean imitation/target ratio was 1.23 (SD = .19). Only 18% reached the ± 1 ST range and 35% reached the ± 2 ST range, with the remainder falling above this wider range.

Regarding octave-shifted pitch matching, the men imitated the SRT targets at a mean imitation/target ratio of 0.54 (SD = 0.05). Here, 21% of the men produced imitated F0 levels within 1 ST of the octave-shifted SRT target but the ± 2 ST range was reached by 86% of the men, with 50% of the mean F0 values between 2 ST and 1 ST below the target. Of the women, only 12% reached F0 levels within 1 ST and simultaneously 2 ST of the octave-shifted target, with the rest falling below this range. The mean imitation/target ratio was 1.58 (SD = .19) and, excluding the two women whose imitations were within 1 ST of the octave-shifted target, 1.53 (SD = .13). This finding raised the question of whether a target shift by a factor of 1.5 instead of 2.0 was relevant for these women's imitations. This type of shift represents a perfect fifth in music, a tone relationship that is perceived as highly harmonious. When a ratio of 1.5 was considered, 41% of the women produced F0 levels within 1 ST of the target and 76%, within 2 ST of the target. In the children's group, the mean imitation/target F0 ratio was 2.03 (SD = 0.25). Here, 47% produced imitations within 1 ST of the octave-shifted target and 59%, within 2 ST of the target, whereas the remainder fell either below or above that range.

Regarding both SRT and T-TRIP imitations together, none of the men produced imitations that fell within the hypothesized target ranges (direct matching for the T-TRIP, octave-downshifted matching for the SRT) within 1 ST. Seven men produced imitations within 1 ST in one test and 2 ST in the other, and two men produced imitations within 2 ST of both

tests. Three produced imitations within 1 ST of one test only and one, within 2 ST of one test only. Only one man missed the target range in both tests.

Of the 17 women for whom both T-TRIP and SRT data were available, only one matched the octave-upshifted ± 1 ST range for the T-TRIP as well as the direct ± 2 ST range for the SRT. When the shift by a factor of 1.5 was considered, nine women produced imitations that fell within the ± 1 ST range for one test and within the ± 2 ST range of the other test, and two women produced imitations within the ± 2 ST range for both tests. The imitations of two women fell within the ± 1 ST range for one test and, in four cases, imitations fell within the ± 2 ST range for one test. None of the women missed the target ranges in both tests.

Of the 15 children for whom data from both the T-TRIP and the SRT were available, one produced imitations within the ± 1 ST hypothesized target ranges for both tests (direct pitch matching for the SRT, octave-upshifted matching for the T-TRIP), one produced imitations within the ± 1 ST range in one test and within the ± 2 ST range in the other test, and one child produced imitations within the ± 2 ST ranges of both tests. In six cases, the imitations fell within the ± 1 ST range in one test only, and in two cases, the imitations fell within the ± 2 ST range only. In four cases, the target ranges were missed altogether.

Research Questions 2 and 3: F0 Differences Among T-TRIP and SRT Imitations and GFTA-2 Productions

Statistical tests of differences among the three measures, T-TRIP (imitations of nonwords spoken by a man), SRT (imitations of nonwords spoken by a woman), and GFTA-2 (non-imitated word productions), were based on stressed non-final syllables only. The men as a group imitated the stressed non-final T-TRIP vowels at an average of 117 Hz (SD = 10.5 Hz) and the stressed non-final SRT vowels at an average of 104 Hz (SD = 9.8 Hz). Mean imitation/target ratio for the T-TRIP was .90 (SD = .08) and, for the SRT, .55 (SD = .05). Mean F0 for stressed non-final GFTA-2 vowels was 109 Hz (SD = 17.3 Hz). Repeated measures ANOVA testing for the stressed F0 levels in the three tasks showed that the model was overall statistically significant ($F = 7.51, p < .0001$), with statistically significant differences among the participants ($F = 6.97, p < .0001$) as well as the three tasks ($F = 11.04, p = .0003$). Pairwise comparisons showed that the difference in F0 levels between the T-TRIP and the SRT was statistically significant ($z = 3.30, p = .0010$), where all men had lower F0 levels during the SRT, compared to the T-TRIP. Relative to the stressed non-final vowels in the GFTA-2, the men raised F0 levels for the T-TRIP with nominal statistical significance ($z = 2.42, p = .0157$), whereas the general trend toward lowering F0 levels from the GFTA-2 to the SRT task was not statistically significant ($z = -1.287, p = .1981$). The men undershot the T-TRIP targets ($z = -3.17, p = .0015$) and slightly overshot the SRT octave-shifted targets ($z = 2.29, p = .0219$). Figure 1 summarizes the stressed non-final F0 in the three tasks using boxplots. Figure 2 shows the distribution of the per-participant mean F0 values for the three measures, where the men were rank-ordered by age from youngest to oldest. Both figures indicate direct and octave-shifted target F0.

The profiles of three men differed from the rest of the group. The youngest man, age 18, had by far the highest F0 level during the GFTA-2, whereas his T-TRIP and SRT F0 levels showed adjustments toward the targets. A 47-year-old man showed only minimal F0

differences among the three measures, and a 66-year-old man did not differentiate F0 levels between the T-TRIP and SRT, whereas his GFTA-2 F0 levels were substantially lower. As mentioned above, significant differences among individual participants were also observed in the repeated measures ANOVA results.

As a group, the women imitated the stressed non-final T-TRIP targets at 197 Hz (SD = 24.5 Hz) and the stressed non-final SRT targets at 183 Hz (SD = 17.6), with a mean imitation/target F0 ratio of 1.50 (SD = .17) for the T-TRIP and .94 (SD = .09) for the SRT. The women produced the stressed non-final GFTA-2 vowels at 172 Hz (SD = 16.5). Repeated measures ANOVA testing revealed that the model was overall statistically significant ($F = 8.92, p < .0001$), with statistically significant differences among the participants ($F = 7.52, p < .0001$) as well as the three tasks ($F = 9.85, p = .0005$). Nonparametric testing showed that the F0 levels in the T-TRIP and SRT differed statistically significantly ($z = 2.96, p = .0031$). The dissimilarity between the F0 levels from the GFTA-2 and the T-TRIP were much more significant ($z = 2.67, p = .0076$) than that between the F0 levels from the GFTA-2 and the SRT ($z = 1.22, p = .2209$), implying significantly raising F0 from GFTA-2 levels during the T-TRIP. A comparison between the SRT imitated and target F0 levels showed that the targets were undershot by most women ($z = -2.76, p = .0057$). As before, the observation that the women averaged an imitation/target F0 ratio of 1.50, representing a highly harmonious tone interval called perfect fifth in music, raised the question whether the perfect fifth became a possible target when an octave would have been at the upper edge of the women's vocal ranges. A comparison between the imitated and 1.5-shifted target F0 resulted in a highly insignificant difference ($z = -0.21, p = .8313$), consistent with low dissimilarity between the imitations and this type of target shift.

Figure 3 shows boxplots of the stressed non-final F0 in the three measures and Figure 4 shows the mean F0 levels for the three tasks for each participant, rank-ordered by age. Both figures indicate the direct target F0 for the SRT and the octave-shifted target F0 for the T-TRIP. Figure 4 also shows the T-TRIP target at a ratio of 1.5:1.

The children imitated the stressed non-final T-TRIP targets at a mean F0 of 253 Hz (SD = 38.7 Hz) with a mean imitation/target ratio of 1.93 (SD = .29) and the stressed non-final SRT targets at a mean F0 of 238 Hz (SD = 35.6) with a mean imitation/target ratio of 1.22 (SD = .19). Mean GFTA-2 F0 was 236 Hz (SD = 30.5 Hz). Repeated measures ANOVA showed that the model was statistically significant overall ($F = 9.49, p < .0001$), a result driven both by differences among the individual children ($F = 9.94, p < .0001$) and variability among the three tests ($F = 5.60, p = .0084$). F0 levels in the T-TRIP differed from those in the SRT with nominal statistical significance ($z = 2.44, p = .0146$). The F0 difference between the GFTA-2 productions and SRT imitations were far from statistically significant ($z = .3550, p = .7226$), whereas F0 differences between GFTA-2 productions and T-TRIP imitations met nominal statistical significance ($z = 2.53, p = .0113$). SRT targets and imitations were highly dissimilar for the group ($z = 3.62, p = .0003$). T-TRIP targets upshifted by a factor of 1.5 were also highly dissimilar from the imitations ($z = 3.62, p = .0003$), but octave-upshifted targets and imitations were not dissimilar ($z = 1.21, p = .2274$). Figure 5 shows boxplots of the stressed non-final F0 values in the three tasks. Figure 6

shows the distribution of mean F0 values for the three measures in this group. Both figures indicate the direct target F0 for the SRT and the octave-shifted target for the T-TRIP.

Research Question 4: The Role of Speech Sound Disorder

Separately for the men's and women's groups and each imitation task, nonparametric rank-sum tests of group differences for imitation/target F0 ratios between those participants with, and without, a history of SSD was conducted, based on stressed non-final vowels. No statistically significant differences between participants with, and without a history SSD were found. The *p* value for the test regarding the men's imitation/target ratios from the T-TRIP approached statistical significance, but exclusion of one low outlier at .69 resulted in a much higher *p* value. Table 3 summarizes the results from nonparametric testing for group differences.

DISCUSSION

The purpose of this study was to investigate phonetic convergence in F0 measures in nonword imitations. We asked whether pitch matching, whether directly or at octave-shifted levels, is an aspect of imitated responses when men, women, and children imitate nonwords spoken by a man and a woman, a set of six different experiments (3 groups X 2 target types). The first set of analyses (Research Question 1) investigated per-participant average F0 levels in terms of proximity to the hypothesized targets, based on a weighted combination of stressed and unstressed non-final target vowels. The second set of analyses (Research Questions 2 and 3) focused on the relationships of the imitated F0 levels within each group relative to non-imitated control F0 levels and target F0 levels, based on stressed vowels only. The fact that the stressed and unstressed targets varied significantly from each other made pitch matching relatively unlikely by chance. The fact that there was little to no physical overlap between the direct target ranges and the hypothesized interval-shifted target ranges for each participant group resulted led to a low chance probability that participants would reach the target ranges in both tasks.

As a group, the men provided the strongest evidence of direct and octave-shifted pitch matching in that they had a high percentage of participants whose imitations fell into the target ranges, especially in the octave-downshifted SRT imitations of stressed and unstressed vowels. Even though there were no physical overlaps between the direct T-TRIP and octave-shifted SRT target ranges, nine of the 14 men produced imitations that fell into the narrow or wider target ranges of both tests. The within-group F0 comparisons between the T-TRIP and SRT imitations based on stressed vowels supported these dissimilar F0 levels, consistent with the pitch-matching hypothesis. The comparisons of the F0 levels in the two imitation tasks with the F0 levels in the non-imitated GFTA-2 control vowels shows that the upshift towards the T-TRIP targets was more significant than the downshift towards the octave-shifted SRT targets. It was therefore surprising that all but one of the men undershot the T-TRIP targets, even though the speaker was a man. One possible explanation comes from studies investigating discourse dynamics showing that speakers match the pitch register of the preceding speaker to signal agreement [28, 29]. It is possible that the

imitations of initial syllables were influenced by the last syllable of the target, which tended to be unstressed and, hence, lower in F0.

The women provided evidence of direct and interval-shifted pitch matching as well. Only 4 out of 21 did not produce imitations within the wider SRT target range, although a lower percentage produced imitations in the narrow target range, compared to the men as a group. The women produced significantly higher F0 levels during the T-TRIP, compared to the SRT, and the upshift from non-imitated control vowels toward the T-TRIP imitations was significant. Two women showed octave-shifted pitch matching during the T-TRIP, an F0 level that was more than 50 Hz above the women's non-imitated control F0 levels. The remaining women produced F0 levels that scattered closely about a hypothesized target at a perfect fifth interval from the actual stimuli. The harmonicity of this ratio has been established for musical tones. The present results are consistent with the hypothesis that the same harmonicity is relevant in speech-like tokens as well, although this study did not directly address perceptual harmonicity ratings in speech-like tokens with various F0 ratios. The perfect fifth can be hypothesized as a target when an octave shift would be uncomfortably above normal speaking levels.

The comparison between the non-imitated control vowel F0 and the SRT F0 using stressed vowels showed that there were no significant differences, but the comparison with the SRT imitations and the target shows that the women as a group undershot the target, similar to what was observed in the group of men.

Only 35% of the children produced F0 levels within 2 ST of the SRT targets, and a subset of these, 18%, produced F0 levels within 1 ST of the target. Most of the children produced imitations above the ± 2 ST range, consistent with the observation that children's F0 levels are considerably higher than an adult female's F0 level (Figure 3). A higher percentage of the children produced imitations that fell into the octave-shifted T-TRIP target range, which was closer to their non-imitated control F0 levels. Given the complete lack of overlap between the direct SRT and the octave-shifted T-TRIP target ranges, the fact that three children produced imitations within the hypothesized target ranges of both tests supports the pitch matching hypothesis. Further evidence was the high degree of dissimilarity between the SRT and T-TRIP F0 levels.

Together, the results from the two adult groups and the child group show evidence of direct and interval-shifted pitch matching. In all three groups, variability among the participants was considerable, in that a subset of participants in each group matched the target F0 levels in both nonword imitation tasks whereas some participants did not show differential responses to the two tasks. A similar variability had been observed in a pitch-matching experiment of musical tones in untrained singers [10]. The source of this variability should be addressed in future studies of pitch perception and production.

No evidence was found that a present or past history of SSD influences F0 levels in nonword imitations. In our previous study [19], young children with SSD produced octave-shifted imitations of targets spoken by a male adult to the same extent as the controls without SSD in stressed syllables but to a lesser extent in unstressed targets. Here, we do not

show separate results for stressed and unstressed syllables in the SSD affectation subgroups. It is possible that group differences would be evident if only unstressed syllables are considered. Alternatively, It is possible that in older children and adults, unstressed syllables have a higher psycholinguistic prominence than in younger children such as those described in our 2009 study. SSD is a heterogeneous disorder with great phenotypic variability and the participant sample may reflect this diversity. It is likely that perceptual or productive F0 skills are not uniform across individuals with SSD. It is further possible that pitch matching ability is a familial trait. However, the fact that many of the participants in this study were biologically related to other participants was not further investigated here.

Conclusions

The results from this study are consistent with the hypothesis that, in imitations of speech-like tokens, F0 is an aspect of the target that is imitated directly when it is comfortably within the responder's vocal range and at harmonically related levels (octave or perfect fifth) when it is outside the responder's vocal range and the harmonically related levels are within the responder's vocal range. These findings replicate previous findings of F0 convergence in a vowel imitation task without explicit instructions regarding F0 [2], and they extend these findings to octave-shifted target ranges. Current or past history of SSD did not influence pitch matching. Future studies should address perceptual and productive F0 abilities separately and investigate these abilities in various types of speech-like stimuli, comparing them to music-like stimuli in the same participants. They should also investigate the role of familial relatedness and the role of age in pitch matching ability.

Acknowledgments

The authors gratefully acknowledge the following funding sources: American Speech–Language–Hearing Foundation New Century Scholars Research Grant (B. Peter), NIDCD T32DC00033 (B. Peter), and NIDCD R03DC010886 (B. Peter). Brett Bankson, Hailey Benesch, Alice Cho, Angela Huang, Kate Sailor, Rachel VanPuymbrouck, and Tiffany Waddington assisted with the acoustic measurements. Sincere thanks to the individuals who participated in this study.

References

1. Pardo JS. Measuring phonetic convergence in speech production. *Front Psychol.* 2013; 4:559. [PubMed: 23986738]
2. Garnier M, Lamalle L, Sato M. Neural correlates of phonetic convergence and speech imitation. *Front Psychol.* 2013; 4:600. [PubMed: 24062704]
3. Pardo JS. On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America.* 2006; 119(4):2382–93. [PubMed: 16642851]
4. Gentilucci M, Bernardis P. Imitation during phoneme production. *Neuropsychologia.* 2007; 45(3): 608–15. [PubMed: 16698051]
5. Sato M, et al. Converging toward a common speech code: imitative and perceptuo-motor recalibration processes in speech production. *Front Psychol.* 2013; 4:422. [PubMed: 23874316]
6. Babel M, Bulatov D. The role of fundamental frequency in phonetic accommodation. *Language and Speech.* 2012; 55(Pt 2):231–48. [PubMed: 22783633]
7. Hutchins SM, Peretz I. A frog in your throat or in your ear? Searching for the causes of poor singing. *Journal of Experimental Psychology General.* 2012; 141(1):76–97. [PubMed: 21875245]
8. Moore RE, et al. Pitch discrimination and pitch matching abilities with vocal and nonvocal stimuli. *Journal of Voice.* 2008; 22(4):399–407. [PubMed: 17509827]

9. Jancke L, Langer N, Hanggi J. Diminished whole-brain but enhanced peri-sylvian connectivity in absolute pitch musicians. *Journal of Cognitive Neuroscience*. 2012; 24(6):1447–61. [PubMed: 22524277]
10. Estis JM, et al. Pitch-matching accuracy in trained singers and untrained individuals: the impact of musical interference and noise. *Journal of Voice*. 2011; 25 (2):173–80. [PubMed: 20456914]
11. Bella SD, Berkowska M, Sowinski J. Disorders of pitch production in tone deafness. *Front Psychol*. 2011; 2:164. [PubMed: 21811479]
12. Schellenberg EG, Trehub SE. Frequency ratios and the perception of tone patterns. *Psychonomic Bulletin & Review*. 1994; 1(2):191–201. [PubMed: 24203470]
13. Foss AH, Altschuler EL, James KH. Neural correlates of the Pythagorean ratio rules. *Neuroreport*. 2007; 18(15):1521–5. [PubMed: 17885594]
14. Burns EM, Ward WD. Categorical perception--phenomenon or epiphenomenon: evidence from experiments in the perception of melodic musical intervals. *J Acoust Soc Am*. 1978; 63(2):456–68. [PubMed: 670543]
15. Small AR, McCachern FL. The Effect of Male and Female Vocal Modeling on Pitch-Matching Accuracy of First-Grade Children. *Journal of Research in Music Education*. 1983; 31(3):227–233.
16. Peterson RL, et al. Neuropsychology and genetics of speech, language, and literacy disorders. *Pediatr Clin North Am*. 2007; 54(3):543–61. vii. [PubMed: 17543909]
17. Pennington BF, Bishop DV. Relations among speech, language, and reading disorders. *Annu Rev Psychol*. 2009; 60:283–306. [PubMed: 18652545]
18. Koike KJ, Asp CW. Tennessee Test of Rhythm and Intonation Patterns. *J Speech Hear Disord*. 1981; 46(1):81–7. [PubMed: 7206683]
19. Peter B, Larkin T, Stoel-Gammon C. Octave-shifted pitch matching in nonword imitations: the effects of lexical stress and speech sound disorder. *Journal of the Acoustical Society of America*. 2009; 126(4):1663–6. [PubMed: 19813781]
20. Button L, et al. Associations among measures of sequential processing in motor and linguistics tasks in adults with and without a family history of childhood apraxia of speech: a replication study. *Clinical Linguistics and Phonetics*. 2013; 27(3):192–212. [PubMed: 23339292]
21. Peter B, et al. Deficits in sequential processing manifest in motor and linguistic tasks in a multigenerational family with childhood apraxia of speech. *Clinical Linguistics and Phonetics*. 2013; 27(3):163–91. [PubMed: 23339324]
22. Peter B, Matsushita M, Raskind WH. Motor sequencing deficit as an endophenotype of speech sound disorder: a genome-wide linkage analysis in a multigenerational family. *Psychiatric Genetics*. 2012; 22(5):226–34. [PubMed: 22517379]
23. Peter B, Raskind WH. A multigenerational family study of oral and hand motor sequencing ability provides evidence for a familial speech sound disorder subtype. *Top Lang Disord*. 2011; 31(2):145–167. [PubMed: 21909176]
24. Shriberg LD, et al. A nonword repetition task for speakers with misarticulations: the Syllable Repetition Task (SRT). *J Speech Lang Hear Res*. 2009; 52(5):1189–212. [PubMed: 19635944]
25. Goldman, RFM. Goldman-Fristoe Test of Articulation 2. Circle Pines: American Guidance Service; 2000.
26. Boersma P. Praat, a system for doing phonetics by computer. *Glott International*. 2001; 5(9/10):341–345.
27. Bradshaw E, McHenry MA. Pitch discrimination and pitch matching abilities of adults who sing inaccurately. *Journal of Voice*. 2005; 19(3):431–9. [PubMed: 16102669]
28. Brazil, D.; Coulthart, M.; Johns, C. Discourse intonation and language teaching: Applied linguistics and language study. London: Longman; 1985.
29. Wennerstrom, A. The music of everyday speech: Prosody and discourse analysis. Oxford: Oxford University Press; 2001.

Figure 1. Boxplots of F0 levels and target F0 for stressed non-final T-TRIP and SRT vowels and GFTA-2 F0 in the men. Solid black line = T-TRIP target; dashed gray line = octave-shifted SRT target.

Figure 2.

F0 levels and target F0 for stressed non-final T-TRIP and SRT vowels and GFTA-2 F0 in the men, rank-ordered from youngest to oldest. O-S = octave-shifted.

Figure 3.

Boxplots of F0 levels and target F0 for stressed non-final T-TRIP and SRT vowels and GFTA-2 F0 in the women. Solid gray line = SRT target; dashed black line = T-TRIP target shifted by a ratio of 1.5.

Figure 4. F0 levels and target F0 for stressed non-final T-TRIP and SRT vowels and GFTA-2 F0 in the women, rank-ordered from youngest to oldest. O-S = octave-shifted.

Figure 5. Boxplots of F0 levels and target F0 for stressed non-final T-TRIP and SRT vowels and GFTA-2 F0 in the children. Solid gray line = SRT target; dashed black line = octave-shifted T-TRIP target.

Figure 6. F0 levels and target F0 for stressed non-final T-TRIP and SRT vowels and GFTA-2 F0 in the children, rank-ordered from youngest to oldest. O-S = octave-shifted.

Table 1

Target F0 ranges for the T-TRIP and SRT in Hz based on stressed and unstressed vowels in non-final position

		Stressed Nonfinal Vowels				Unstressed Nonfinal Vowels			
		Within 1 ST		Within 2 ST		Within 1 ST		Within 2 ST	
		Lower Bound	Upper Bound	Lower Bound	Upper Bound	Lower Bound	Upper Bound	Lower Bound	Upper Bound
TTRIP	Direct	123.5	138.6	116.6	146.9	102.4	115.0	96.7	121.8
	O-S Up	247.0	277.3	233.2	293.8	204.8	229.9	193.3	243.6
	1.5 Up	185.3	208.0	174.9	220.3	153.6	172.4	145.0	182.7
SRT	Direct	183.8	206.3	173.5	218.6	173.7	194.9	163.9	206.5
	O-S Down	91.8	103.1	87.7	109.3	86.8	97.5	82.0	103.3

Note: O-S = octave-shifted, ST = semitone

Table 2

Distribution of mean imitation/target F0 ratios across ST bands by participant group and task (non-final stressed and unstressed syllables) and number and percent of participants with mean imitation values within 1 and 2 ST of the target, respectively.

Type	Sample	< -2 ST	> -2 ST	< -1 ST	± 1 ST	> 1 ST	< 1 ST	> 2 ST	Total N	% ± 1 ST	% ± 2 ST
D	Men TTRIP	4	3	7	14	50	71				
	Women SRT	3	11	4	21	19	81				
	Children SRT			3	17	18	35				
O-S	Men SRT .5:1		2	3	14	21	86				
	Women TTRIP 2:1	15		2	17	12	12				
	Women T-TRIP 1.5:1		3	7	17	41	76				
	Children TTRIP 2:1	3	1	8	17	47	59				

Note: D = direct, O-S = octave-shifted, ST = semitone

Table 3
Results from rank-sum testing for group differences between SSD affected and unaffected individuals in each group and task

Group	T-TRIP		SRT			
	N	z	p	N	z	p
Men	8 A, 6 U	-1.29	0.1967	8 A, 6 U	-1.94	0.0528
Women	5 A, 11 U	.40	0.6917	8 A, 13 U	0.65	0.5145
Children	12 A, 4 U	-.24	0.8084	11 A, 5 U	-0.85	0.3955

Note: A = affected, U = unaffected