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, *Seattle, 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, and 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 with that in the targets and in the participants' nonimitated 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.

Key Words: Phonetic convergence–Nonword imitation–Direct pitch matching–Octave equivalence–Octave-shifted pitch matching–Perfect fifth–Speech sound disorder.

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, eg, 3:2 ("perfect fifth") and 4:3 ("perfect fourth"), are perceived as more harmonious than intervals with ratios involving higher integers, eg, 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 used 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 because of distorted, substituted, omitted, or inserted speech sounds. For clinical and research purposes, imitated nonwords are evaluated for accuracy by comparing their phoneme sequences with those in the target. Lexical stress errors are evaluated in only very few tests of

Accepted for publication June 13, 2014.

The authors B.P., B.F., and H.H. contributed equally to this work.

From the Department of Speech and Hearing Sciences, University of Washington, Seattle, Washington.

Address correspondence and reprint requests to Beate Peter, Department of Speech and Hearing Sciences, University of Washington, Box 354875, Seattle, WA 98195. E-mail: bvpeter@u.washington.edu

Journal of Voice, Vol. 29, No. 2, pp. 260.e21-260.e30
0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2014.06.011

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 although 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 six different families with an age range of 3 to 80 years (14 men, 21 women, and 19 children defined here as aged 13 years or younger). Thirteen children had a present or past SSD diagnosis and eight women and eight men reported a history of SSD. Before the study, a hearing screening was administered whenever possible and participants passed the screening at 25 dB sound pressure level at 0.5, 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 two syllables (eight items), three syllables (six items), or four syllables (three items), all with trochaic rhythmic patterns. The prerecorded targets were spoken by an adult female with an F0 averaging 194.8 Hz (standard deviation [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 ($t = 10.23$; $P < 0.0001$). The T-TRIP[18] is a nonstandardized 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 two to six 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 of 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 ($t = 11.63$; $P < 0.0001$). 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 eight ST and there were no F0 overlaps. Additional details are shown in Table 1.

Together, the three participant groups (men, women, and children) and two target types (a man's voice and 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, ie, 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 the 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 nonimitated 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."