Can Listeners Hear Who Is Singing? The Development of Voice Category Perception

Molly L. Erickson, Jacob Burchard, and Payton Phillips, *Knoxville, Tennessee*

Summary: Objective and Hypothesis. This study sought to investigate if a small amount of training in identification of voices elicits the development of prototypical vocal categories.
Study Design. This study used a between-group design.
Methods. This study used an ABX paradigm where listeners heard two different singers singing "ah" at the same pitch. Listeners identified which of the two singers was the producer of a third "ah" at a different pitch. Stimuli were recorded from two baritones, two tenors, two mezzo-sopranos, and two sopranos across a 1.5-octave range. Participants were randomly assigned to two groups, one group that received a training session using voices that were different from those in the experimental session, but of the same voice category, and one group that received no training.
Results. Training listeners with voices that are different from those of singers presented in the experiment did not significantly improve the ability to discriminate individual voices of the same voice category, but did significantly improve that ability to discriminate individual voices when the voices being compared were of different voice categories.
Conclusions. Small amounts of purposeful exposure to human voices appear to result in the beginnings of listener voice category formation, providing listeners with prototypical categories that can aid them in discrimination of novel voices of those same categories.
Key Words: Voice classification–Perception–Timbre–Training–Category formation.

INTRODUCTION

This paper presents one experiment from a series of studies designed to develop a model of how listeners use timbre to identify singers or voice categories. To conceptualize such a model, it is first necessary to understand how timbre is perceived and how such perception might result in the development of perceptual voice categories.

Timbre is technically defined as the attribute that leads to a perception of dissimilarity between two sounds at the same pitch and loudness.[1] Inherent in this definition is the implication that timbre is the property of a sound that emerges only in comparison with another sound and therefore is not a constant attribute across all sounds produced by a sound source. Yet, the term timbre is often used to describe an invariant property of a sound-producing object, as in the timbre of a flute or the timbre of a specific singer. Researchers, however, have found little evidence to suggest an acoustic signature that is invariant over a singer's[2] or an instrument's[3] entire range of production.[3] Erickson and colleagues[3–6] have engaged in a series of studies designed to test how listeners use timbre to identify or discriminate instruments, voices, and voice categories, and have found that this ability decreases as the pitch interval used in the comparison increases, suggesting that there is not one perceptual cue or one set of perceptual cues that listeners use to identify instruments or singers. Handel and Erickson[7] argue that these results strongly suggest that sound-producing objects do not have one timbre but have timbre spaces that may be defined by a "timbre transformation" across pitch.[7] Hypothetically then, voice categories would comprise those individuals whose timbre spaces are most similar.

The timbre space of a sound-producing object provides a rich perceptual basis for categorization. Categorization is the process whereby ideas and objects are identified, differentiated, and understood.[8] Identifying a voice as "soprano" is analogous to identifying a category such as "chair," a superordinate category. Identifying a voice as a "spinto soprano" is analogous to identifying a "dining room chair," a subordinate category. Identifying an individual singer is analogous to identifying one individual chair. It may be that attempting to discriminate unfamiliar voices across pitch is analogous to attempting to differentiate the legs of one chair from the rungs of another without being familiar with the individual chairs, whereas attempting to discriminate voices of unfamiliar voice categories across pitch may be analogous to attempting to differentiate the legs of a table from the rungs of a chair without having knowledge of the categories of "table" and "chair."

There are 2 primary theories of categorization: (1) the exemplar theory and (2) the prototype theory. According to the exemplar theory, individual exemplars of a category are stored in memory,[9,10] and classification decisions are made based on the similarity of the stimuli to the stored exemplars.[11–13] According to the prototype theory, similarities between exemplars lead to the development of a summary representation or prototype that is stored in memory.[9,10] Stimuli are included in the category represented by the prototype based on the weight and the number of the prototype features the stimuli contain.[14–18] Erickson[19] tested whether or not familiarity with a singer's timbre space improves the ability to discriminate the singer across pitch and found that a short training session using the same voices later used in the discrimination experiment improved discrimination of singing voice categories, but not of individual singers, suggesting that, during the training, listeners might be

Accepted for publication June 23, 2017.
These data were presented in a paper at the 45th Annual Symposium: Care of the Professional Voice, Philadelphia, 2016.
From the Department of Audiology and Speech Pathology, University of Tennessee Health Sciences Center, Knoxville, Tennessee.
Address correspondence and reprint requests to Molly Erickson, Department of Audiology and Speech Pathology, University of Tennessee, 578 South Stadium Hall, Knoxville, TN 37996. E-mail: merickso@uthsc.edu
Journal of Voice, Vol. ■■, No. ■■, pp. ■■-■■
0892-1997

https://doi.org/10.1016/j.jvoice.2017.06.016

developing prototypical voice category timbre spaces rather than storing individual timbre spaces in memory.

The purpose of the present study was to further explore singing voice category formation by replicating the Erickson[19] study with one small change: the training stimuli used in the current study, although consisting of voices representing the same voice categories as those of the experimental stimuli, are a different set of singers than those presented as the experimental stimuli. The hypothesis is that, if listeners develop prototypical representations of voice categories during training, then they should be able to use those category prototypes to improve between-category singer discrimination of the experimental singing voices, even though the training stimuli consist of a set of voices completely different from that of the experimental stimuli. A corollary to this hypothesis is that, just as in the 2016 Erickson study, training will not improve discrimination in within-category comparisons.

METHODS

Stimuli

Two groups of vocal stimuli were recorded, training stimuli and experimental stimuli. Master's level singers from the Department of Music at the University of Tennessee, Knoxville, provided the stimuli used in both stimulus groups. All participants provided informed consent using a procedure that was previously approved by the Institutional Review Board (IRB) of the University of Tennessee, Knoxville. These participants met the following criteria: (1) bilateral hearing within normal limits as determined by a 20-dB hearing screening at 500, 1000, 2000, and 4000 Hz[20]; (2) voice study at the master's degree level or higher; and (3) no voice problems at the time of taping as determined by a certified speech-language pathologist. Additionally, all participants had been consistently categorized by the voice faculty as soprano, mezzo-soprano, tenor, or baritone for a minimum of 3 years. The singers in the training group ranged in age from 21 to 24 years, with a mean age of 23.375 years. The singers in the experimental group ranged in age from 23 to 31 years, with a mean age of 25.7 years.

For each group, training and experimental, two baritones and two tenors were recorded, producing /ɑ/ at the pitches C3, E3, G3, B3, D4, and F4, whereas two sopranos and two mezzo-sopranos were recorded producing /ɑ/ at the pitches C4, E4, G4, B4, D5, and F5. Each singer produced a sustained /ɑ/ for approximately 4 seconds. Recordings were made in a single-walled sound booth (Acoustic Systems RE-144-S, Austin, TX). Participants were recorded using either a digital audio tape recorder (Sony PCM-R500; Sony, Park Ridge, NJ) or a solid-state recorder (Marantz PMD670, Marantz, Mahwah, NJ, Japan) and a Sennheiser MD 441-U microphone (Sennheiser, Old Lyme, CT). Participants stood in the center of the booth. Lip-to-microphone distance was 30 cm (12 inches). A keyboard was used to present pitches. Before taping, the participants were allowed to vocalize freely and to become comfortable with the recording environment.

One-second digital samples were constructed for each sung stimulus using the software program *Audition* (Adobe Systems, San Jose, CA). One-second stimuli were created by measuring the exact duration of the sung stimulus, calculating its midpoint, and then extracting a 1-second segment that originated 0.5 second before the midpoint and terminated 0.5 second after the midpoint. Spline curve amplitude shaping functions were applied to each sample to provide ramped onsets and offsets. The overall amplitude of each stimulus was adjusted so that all were of approximately equal amplitude.

Listeners

All listeners provided informed consent using a procedure previously approved by the IRB of the University of Tennessee, Knoxville. Although the Department of Audiology and Speech Pathology is officially part of the University of Tennessee Health Sciences Center, it is physically located on the University of Tennessee Knoxville campus and, as such, is allowed by contractual agreement to obtain IRB approval from that organization. Listeners were recruited from students enrolled in introductory psychology courses at the University of Tennessee, Knoxville, and from students enrolled in courses in the University of Tennessee Health Sciences Center's Department of Audiology and Speech Pathology. Listeners who met the following criteria were recruited: (1) bilateral hearing within normal limits as determined by a 20-dB hearing screening at 500, 1000, 2000, and 4000 Hz[20]; (2) no history of choral singing or vocal training; and (3) no interest in classical vocal music or opera. Eighty-two listeners were recruited for the experiment. The listeners were divided into two groups. Subjects placed in the training group (n = 48) received training to familiarize themselves with the eight voices in the training set before the experimental session. Those placed in the untrained group (n = 42) did not receive training before the experimental session. The trained group consisted of 23 students from introductory psychology courses and 25 students from courses in audiology and speech pathology. This group included 34 female and 14 male participants with a mean age of 21.85 years and an age range of 18–38 years. All 42 of the students in the untrained group were recruited from introductory psychology courses and consisted of 18 female and 24 male participants with a mean age range of 19.952 years and an age range of 18–40 years.

Training

There were two separate training sessions, one for the four male training set voices and one for the four female training set voices. During a training session, listeners were presented with four buttons on the screen labeled "singer 1," "singer 2," "singer 3," and "singer 4." Each of these buttons was randomly assigned to one of the 4 singers in the study. When listeners clicked a button, they heard all six of the recorded stimuli for that singer in ascending pitch separated by 0.25 seconds of silence. Listeners were told to listen to all of the singers until they believed that they could identify all four singers. When the listeners believed they could identify the voices, they clicked a button labeled "Test Myself." Listeners were presented with four buttons labeled with "?" on the Test Myself screen. Each of the four singers being tested were randomly assigned to a "?" button. Next to each button were four mutually exclusive radio buttons labeled "singer 1",