Can Listeners Hear Who Is Singing? The Role of Familiarity

Molly L. Erickson, *Knoxville, Tennessee*

Summary: Objective/Hypothesis. This study sought to determine whether familiarity with voices increases discrimination of voices across pitch intervals.

Study Design. This is a between-group design.

Methods. This study used a forced-choice paradigm where listeners heard two different singers (singer 1 and singer 2) producing /ɑ/ at the identical pitch and an unknown singer (either singer 1 or singer 2) producing /ɑ/ at a different pitch. Listeners had to identify which singer was the unknown singer. Two baritones and two tenors were recorded producing /ɑ/ at the pitches C3, E3, G3, B3, D4, and F4. Two sopranos and two mezzo-sopranos were recorded producing /ɑ/ at the pitches C4, E4, G4, B4, D5, and F5. For each group of stimuli, male and female, all possible pairs of singers were constructed for the lowest pitch (C2 or C3, respectively) and for the highest pitch (F4 or F5, respectively). The unknown singer was varied across the remaining pitches. Participants in group 1 completed a training session where they were familiarized with the voices being tested. Participants in group 2 did not.

Results. Training did not significantly improve the ability to discriminate voices when the voices being compared were of the same voice category. However, training did significantly improve the ability to discriminate voices when the voices being compared were of different voice categories even when training lasted as little as 5 minutes.

Conclusions. Small amount of exposure to human voices results in voice category formation but does not result in the formation of models of individual voices.

Key Words: Voice classification–Perception–Timbre–Pitch–Training–Familiarity.

INTRODUCTION

This article presents one experiment from a series of studies designed to develop a model of how listeners use timbre to identify singers or voice categories. Critical to this line of research is the technical definition of timbre: two tones are of different timbre if they are judged to be dissimilar and yet have the same loudness and pitch.[1] By the technical definition, timbre is simply a comparison of two sounds of equal pitch and loudness. Therefore, using the strict definition, it is impossible to say that two sounds produced at different pitches by the same singer, speaker, or other sound-producing object, have the same timbre. Yet, the term timbre is often used to describe an invariant property of a sound-producing object, as in the timbre of a clarinet. Cleveland[2] states that an individual singer has a characteristic timbre that is a function of the laryngeal source and vocal tract resonances. Singers with similar timbres, then, constitute members of the same voice timbre type or voice category. However, Mellody and Wakefield[3] found little evidence to suggest an acoustic signature that is invariant over a singer's entire range of production and suggest that singers create the impression of a single instrument by smoothly transitioning from one local region of invariance to the next.

Erickson et al have engaged in a series of studies designed to test how listeners use timbre to identify or discriminate sound-producing objects including instruments and voices.[4–6] Handel and Erickson[4] found that trumpets and clarinets can be perceived as having very different timbres on some pitch-loudness combinations and very similar timbres on others, supporting the idea of a set of timbres rather than an invariant acoustic signature in instruments. Using female voices and a three-note oddball task, Erickson et al[5] found that when comparison intervals exceeded one octave, inexperienced and experienced listeners were unable to detect the different singer within or across voice category and most often chose the most dissimilarly pitched stimulus, again suggesting that there is no one invariant acoustic parameter that can be used to discriminate voices. To further test how listeners discriminate human voices, Erickson[6] conducted a more in-depth examination of this phenomenon with male and female voices and found that the ability to discriminate singers of different voice categories diminishes with increasing pitch interval and reaches lower than chance level somewhere around the intervals of the 9th to the 11th. Performance was far worse when singers were asked to discriminate singers of the same voice category. These findings argue strongly against an invariant acoustic signature for human voices. Instead, it is possible that each voice has a timbre transformation, a systematic way in which timbre changes over pitch and loudness that is based on the interaction of the voice source and the vocal tract. The possible timbres for any one voice might be called a singer's timbre template. Voice categories then would comprise those individuals whose timber templates are most similar.

The process whereby ideas and objects are identified, differentiated, and understood is termed categorization.[7] Speech pathologists regularly categorize voices using such labels as "breathy" or "rough," whereas vocal pedagogues apply labels such as "soprano," "lyric," or "spinto." Identifying a voice as "soprano" is analogous to identifying a category such as

Accepted for publication July 14, 2015.

These data were presented in an article at the 44th Annual Symposium: Care of the Professional Voice, Philadelphia, June 2015.

From the Department of Audiology and Speech Pathology, University of Tennessee Health Science Center, Knoxville, Tennessee.

Address correspondence and reprint requests to Molly L. Erickson, Department of Audiology and Speech Pathology, 578 South Stadium Hall, University of Tennessee, Knoxville, TN 37996. E-mail: merickso@uthsc.edu

Journal of Voice, Vol. ■, No. ■, pp. 1-7

0892-1997/$36.00

http://dx.doi.org/10.1016/j.jvoice.2015.07.005

"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 is 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."

The purpose of this study was to examine whether familiarity with a singer's timbre template improves the ability to discriminate the singer across pitch. It is expected that a training session designed to familiarize listeners with the timbre templates of the experimental vocal stimuli will result in improved discrimination of those voices or at least result in the beginnings of category formation that will allow better discrimination when the voices are of two different categories.

METHODS

Stimuli

Master's level singers from the Department of Music at the University of Tennessee, Knoxville, provided the stimuli used in the experiment. All participants provided informed consent using a procedure that was previously approved by the institutional review board 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 Hz, 1000 Hz, 2000 Hz, and 4000 Hz[8]; (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 ranged in age from 23 to 31 years with a mean age of 25.7 years.

Two baritones and two tenors were recorded producing /ɑ/ at the pitches C3, E3, G3, B3, D4, and F4. 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 a digital audio tape recorder (Sony PCMR500, Park Ridge, NJ) and a Sennheiser MD 441-U microphone (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, participants were allowed to vocalize freely and become comfortable with the recording environment.

One-second digital samples were constructed for each sung stimulus. Each stimulus was low-pass filtered at 20 kHz using a Tucker-Davis-Technologies FT6 anti-aliasing filter (TDT, Gainesville, FL), then digitized at 48 kHz using a 24-bit Transit external sound card from M-Audio (Irwindale, CA). The software program Cool Edit Pro (Syntrillium Software Corporation, Phoenix, AZ) was used to extract initial 1 second including the onset of the sung /ɑ/. Spline curve amplitude shaping functions were applied to the end of each stimulus to provide ramped 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 institutional review board of the University of Tennessee Health Sciences Center. Listeners were recruited from students enrolled in introductory psychology courses at the University of Tennessee, Knoxville. Listeners were recruited that met the following criteria: (1) bilateral hearing within normal limits as determined by a 20-dB hearing screening at 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz[8]; (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 familiar group ($n = 40$) would receive training to familiarize themselves with the voices used in the study before the experiment. Those placed in the unfamiliar group ($n = 42$) would not receive training. The trained group consisted of 16 female and 24 male participants with a mean age of 19.475 years and a range of 18–31 years of age. The untrained group consisted of 18 female and 24 male participants with a mean age range of 19.952 years of age and a range of 18–40 years of age.

Training

There were two separate training sessions, one for the four male voices and one for the four female 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 four singers in the study. Listeners were not told that these voices were the voices they would hear in the experiment. When a listener 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 the singers until they believed that they could identify all four singers. When the listener believed they could identify the voices, they clicked a button labeled "Test Myself." On the "Test Myself" screen, listeners were presented with four buttons labeled with "?." Each of the four singers being tested was randomly assigned to a "?" button. Next to each button were four mutually exclusive radio buttons labeled "singer 1," "singer 2," "singer 3," and "singer 4." Listeners played all the "?" singers and attempted to identify which singer in the practice session was associated with which button. When they were sure of their answers, they clicked the "Done" button. If they correctly identified all four singers, they proceeded to the experiment, if they did not correctly identify all four singers, they were returned to the practice screen. This process continued until (1) the listeners successfully completed the practice test or (2) the listener had completed 30 minutes of practice without successfully completing the practice test.