Can Inexperienced Listeners Hear Who Is Singing? The Role of Onset Cues

Molly L. Erickson, *Knoxville, Tennessee*

Summary: Objective/Hypothesis. This study sought to determine whether the acoustic information contained in the onset of phonation can improve the ability to discriminate voices across pitch.

Study Design. This study was a repeated-measures factorial design.

Methods. Listeners heard two singers producing /ɑ/ at the same pitch and an unknown singer producing /ɑ/ at a different pitch. Listener's identified 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. The unknown singer was varied across the remaining pitches. All stimulus combinations were presented with and without onset cues.

Results. In general, the inclusion of onset information affected listeners' ability to discriminate singers across pitch. When the pitch interval was small, a 3rd or a 5th, onset information improved the ability to correctly discriminate singers across pitch. This effect was greater for female voices than for male voices. However, when the pitch interval was larger, a 7th, 9th, or 11th, the onset information either had no effect, as was seen in the female voices, or had a negative effect, as was seen in the male voices.

Conclusion. Voice onset information affects voice discrimination in certain circumstances, but cannot be used by inexperienced listeners as a consistent cue to improve singer discrimination across all pitch intervals.

Key Words: Voice classification–Perception–Timbre–Pitch–Onset cues.

INTRODUCTION

The technical definition of timbre is as follows: two tones are of different timbre if they are judged to be dissimilar and yet have the same loudness and pitch.[1] Using the strict definition, it is impossible to say that two sounds produced at different pitches by any sound-producing object, including speakers and singers, have the same timbre. However, the term "timbre" is frequently used as if it is 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. The use of the term "timbre" in this fashion implies that a singer has one timbre and that singers with similar timbres constitute members of the same voice timbre type or voice category. However, Mellody and Wakefield[3] did not find evidence suggesting an acoustic signature that is invariant over a singer's entire range of production. Their research suggests that singers create the impression of a single instrument by smoothly transitioning from one local region of invariance to the next. Given the technical definition of timbre, it may be more appropriate to think of every sound-producing object, including singers, as having a timbre space rather than one unique timbre.

Erickson and colleagues have engaged in a series of studies designed to test how listeners use timbre to identify or discriminate sound-producing objects over pitch. 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. Using stimuli with onset and offset information removed, Erickson[5] 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, at least based on steady-state spectral cues.

What remains unknown is the effect of temporal cues such as onset on singer discrimination across pitch. Onsets have been shown to be critical for the identification of musical instruments.[6–10] However, Iverson and Krumhansl[11] found that when two musical instruments were compared with one another in a similarity task, onsets were neither necessary nor sufficient to discriminate the two stimuli. The differing results based on type of task, identification, or similarity are not necessarily surprising given that identification is a categorization task requiring some type of stored representation of the instrument, while discrimination is a very different type of task based on comparison and does not require any prior knowledge or stored representation of the instrument. In other words, when one stimulus is played to listeners and they are asked to name the instrument, onsets provide a very important cue that may better allow listeners to access the semantic representation of that instrument. However, when two instruments are played and listeners are asked to rate the similarity or difference between the two instruments, the presence of onsets may not improve the ability to hear differences in the stimuli.

Accepted for publication June 23, 2017.

These data were presented in a paper at the 45th Annual Symposium: Care of the Professional Voice, Philadelphia, June 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, 578 South Stadium Hall, University of Tennessee, Knoxville, TN 37996. E-mail: merickso@uthsc.edu

Journal of Voice, Vol. 32, No. 4, pp. 466–473

0892-1997

https://doi.org/10.1016/j.jvoice.2017.06.017

Purpose of the study

The purpose of this study was to determine whether singer discrimination across pitch can be improved with the inclusion of onsets. This study proposes the following questions:

(1) Does the inclusion of onset information improve singer discrimination across pitch?
(2) Are onsets' amplitude slopes consistent within a singer and/or voice category across pitch?
(3) Is spectral change during onset consistent within a singer and/or voice category across pitch?
(4) Can onsets' amplitude slope, spectral change during onset, or a combination of the two be used to differentiate singers and/or voice categories across pitch?

METHOD

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: (a) bilateral hearing within normal limits as determined by a 20-dB HL hearing screening at 500 Hz, 1,000 Hz, 2,000 Hz, and 4,000 Hz[12]; (b) voice study at the master's-degree level or higher; and (c) 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 audiotape recorder (Sony PCMR500, Park Ridge, NJ) and an MD 441-U microphone (Sennheiser, Old Lyme, CT). Participants stood in the center of the booth. Lip-to-microphone distance was 30 cm (12 in). A keyboard was used to present pitches. Prior to 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 an FT6 anti-aliasing filter (Tucker-Davis-Technologies, Gainesville, FL), then digitized at 48 kHz using a 24-bit Transit external sound card (M-Audio, Irwindale, CA). The software program *Audition* (Adobe Systems Incorporated, San Jose, CA) was used to create two types of samples from the sung /ɑ/: (1) without onsets and (2) with onsets. One-second without-onset stimuli were created by measuring the exact duration of the sung stimulus, calculating its midpoint, and then extracting a 1-second segment that originated one-half second before the midpoint and terminated one-half second after the midpoint. Spline curve amplitude shaping functions were applied to each sample to provide ramped onsets and offsets. One-second with-onset stimuli were created by extracting the initial 1-second sample and applying a spline curve amplitude shaping function to the offset. The overall amplitude of each stimulus was adjusted so that all were of approximately equal RMS amplitude.

Listeners

All listeners provided informed consent using a procedure previously approved by the Institutional Review Board of the University of Tennessee, Knoxville. Listeners were recruited from students enrolled in introductory psychology courses at the University of Tennessee, Knoxville. Listeners who met the following criteria were recruited: (a) bilateral hearing within normal limits as determined by a 20-dB HL hearing screening at 500 Hz, 1,000 Hz, 2,000 Hz, and 4,000 Hz[12]; (b) no history of choral singing or vocal training; and (c) no interest in classical vocal music or opera. Seventy-seven listeners were recruited for the experiment; however, due to a loss of computer data, only 61 listeners are included in the current study, 40 female listeners and 21 male listeners. Female listeners ranged in age from 18 to 23 years with a mean age of 18.98 years. Male listeners ranged in age from 18 to 25 years with a mean of 19.38 years.

Experimental design

The current study employed 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. It was the listener's task to identify which singer (singer 1 or singer 2) was the unknown singer. For each group of stimuli, male and female, all possible pairs of singers were constructed for the lowest pitch (C2 or C3, respectively) for both without-onset and with-onset stimuli. The unknown singer was varied across the remaining pitches. This resulted in two types of tasks for each group (male and female): (1) without onsets and (2) with onsets (Table 1). For each group, male and female, there were 20 within-voice-category comparisons and 40 between-voice-category comparisons per task (without onsets and with onsets), resulting in a total of 240 comparisons.

Procedure

All listening experiments took place in a single-walled sound booth (Acoustic Systems RE-144-S). Stimuli were presented binaurally using HD 600 headphones (Sennheiser). The stimuli were presented at a comfortable listening level, approximately 65 dB SPL. Stimuli were presented in two counterbalanced blocks: (1) without onset and (2) with onset. Within each block, two counterbalanced sub-blocks were presented: male singers and female singers.

Listeners were presented with three buttons labeled "Singer 1," "Singer 2," and "?". They were instructed both verbally and on screen (see Figure 1) that the sung sound that played when they clicked the button marked with a "?" was sung by one of two singers, Singer 1 or Singer 2, and that it was their task to