Acoustic Properties of the Voice Source and the Vocal Tract: Are They Perceptually Independent?

Molly L. Erickson, *Knoxville, Tennessee*

Summary: Objective/Hypothesis. This study sought to determine whether the properties of the voice source and vocal tract are perceptually independent.

Study Design. Within-subjects design.

Methods. This study employed a paired-comparison paradigm where listeners heard synthetic voices and rated them as same or different using a visual analog scale. Stimuli were synthesized using three different source slopes and two different formant patterns (mezzo-soprano and soprano) on the vowel /a/ at four pitches: A3, C4, B4, and F5.

Results. Whereas formant pattern was the strongest effect, difference in source slope also affected perceived quality difference. Source slope and formant pattern were not independently perceived.

Conclusion. These results suggest that when judging laryngeal adduction using perceptual information, judgments may not be accurate when the stimuli are of differing formant patterns.

Key Words: perception–timbre–independence–source–filter.

INTRODUCTION

Historically, voice production has been viewed as a linear system,[1] with the vibrating vocal folds providing a source excitation that is filtered by the vocal tract. In the linear model, the output acoustic signal is affected by the characteristics of the source and the filtering characteristics of the vocal tract. It is this output signal that is assumed to serve as the input to the human auditory perceptual system. Although researchers have demonstrated that under specific circumstances, source-tract acoustical coupling can and does occur, resulting in a nonlinear system,[2,3] researchers working in the areas of voice and speech perception typically employ linear models to systematically vary acoustic parameters.[4–6]

It has also historically been believed that when presented with the vocal output signal, human beings can extract information concerning the acoustics, and therefore also the physiology, of the voice source separately from the vocal tract. Generally, it is believed that the acoustic characteristics of the vocal tract give rise to perception of (1) vowels[1] and (2) singing voice categories,[4,7,8] whereas the acoustic characteristics of the source give rise to perception of (1) pitch,[9] (2) degree of adduction of the vocal folds,[10] (3) aperiodicity in vibration,[11] and (4) presence of turbulent airflow.[12] Numerous studies have been conducted in an attempt to correlate perceptual findings to acoustic characteristics and also to correlate acoustic characteristics to physiology.[13–29] The success of these studies has varied greatly, depending on the parameter being studied.

Psychoacousticians interested in perception of stimuli (eg, visual, auditory, tactile) have developed statistical procedures designed to mitigate some of the inherent problems in perceptual research.[30] These techniques allow us to answer fundamental questions about perception. Such answers are necessary before attempts to correlate perception to acoustics and physiology can begin. Researchers working on theories of perceptual independence, separability, and integrality[31,32] have developed methods to test the independence of perception dimensions. These techniques have been employed in the study of the perception of musical instruments, vowels,[33] and pathological voice quality.[34,35] The breathy voice has been particularly well studied.[26–29] However, the factors affecting the perception of vocal fold hyper- and hypoadduction remain poorly understood.[36]

From a clinical standpoint, the perceptual voice evaluation is one of the most important parameters used in a voice evaluation.[37] In such an evaluation, clinicians must separate the perception of the vocal tract from the voice source, such that they can assess whether the vocal folds are vibrating in a manner suggestive of either an organic pathology or a behavioral misuse of the mechanism.

Generally, the more tightly adducted the vocal folds, the less steep the slope of the glottal source spectrum.[38,39] Research suggests that stimuli with shallower glottal source slopes are perceived as "pressed" or "brassy," whereas stimuli with steeper glottal source slopes are perceived as "flow" or "fluty."[10] On the other hand, voice category (soprano, mezzo-soprano, contralto, tenor, baritone, and bass) is believed to be related to average vocal tract resonance frequency,[4,7,8] which has in turn been shown to correlate with vocal tract length.[8] It is unknown whether the perception of sound quality associated with vocal tract length is independent from the perception of voice quality associated with glottal spectrum slope. The ability to separately distinguish the properties of the voice source from the properties of the vocal tract is important, as glottal source slope is believed to be a major perceptual cue to underlying vocal fold physiology and as such is important in perceptual voice evaluations.

This research asks the following question: Is the perception of the voice source separable from the perception of the vocal tract? To answer this question, this study utilized synthetic stimuli with specifically controlled parameters of pitch, vibrato, source slope, and formant frequencies.

Accepted for publication November 13, 2015.

Presented at the 39th Annual Symposium: Care of the Professional Voice, Philadelphia, June 2010.

From the Department of Audiology and Speech Pathology, University of Tennessee Health Science 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, Tennessee 37996. E-mail: merickso@uthsc.edu

Journal of Voice, Vol. ■■, No. ■■, pp. ■■-■■
0892-1997

http://dx.doi.org/10.1016/j.jvoice.2015.11.010

METHOD

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 who met the following criteria were recruited: (1) bilateral hearing within normal limits as determined by a 20-dB hearing screening at 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz;[40] (2) no history of choral singing or vocal training; and (3) no interest in classical vocal music or opera. Twenty listeners were recruited for the experiment. There were 12 female and eight male participants with a mean age range of 19.6 years and a range of 18–30 years.

Stimuli

Stimuli were generated synthetically using a digital synthesizer. The synthesis model was built using *Aladdin Interactive DSP Workbench* (Hitech Development, Stockholm, Sweden). For the pitches A3 (220 Hz), C4 (261.6 Hz), B4 (493.9 Hz), and F5 (698.5 Hz), two sets of source stimuli were created, one with vibrato and one without (Figure 1). For each set, vibrato and no vibrato, three different source signals were synthesized with varying slopes of −9 dB/octave, −12 dB/octave, and −15 dB/octave (Figure 2), resulting in six source stimuli at each pitch. Vibrato source signals used a frequency vibrato rate of 5.6 Hz and a frequency vibrato extent of ±50 cents (0.5 semitone). The vibrato rate and extent are values typical of classical western singing.[41] The vibrato and the no vibrato source signals were filtered using two formant patterns (pattern M and pattern S) for the vowel /a/. Pattern M is representative of that typically seen in a mezzo-soprano. Pattern S is representative of a formant pattern typically seen in a soprano (Figure 3). These formant patterns have been demonstrated to be consistently correlated with their respective voice category.[4] Formant patterns M and S are displayed in Table 1. This resulted in six stimuli for each of the following pitch-vibrato combinations: A3 vibrato, A3 no vibrato, C4 vibrato, C4 no vibrato, B4 vibrato, B4 no vibrato, F5 vibrato,

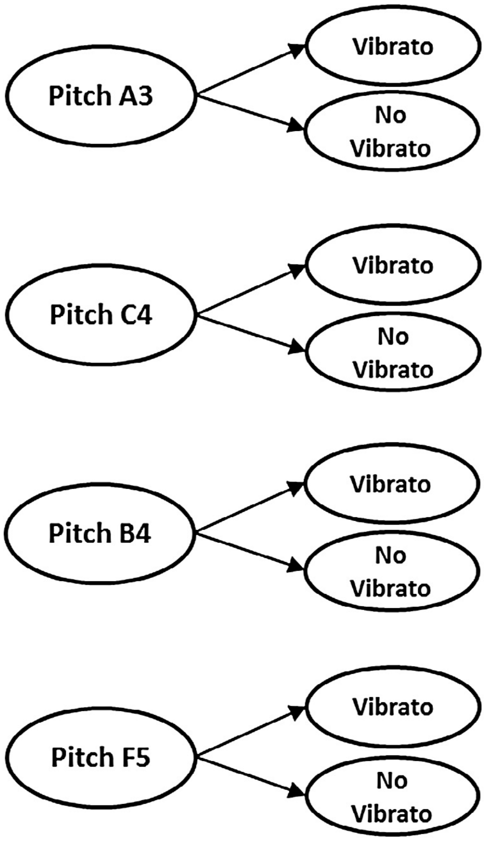

FIGURE 1. Sets of stimuli created for each pitch: A3, C4, B4, and F5.

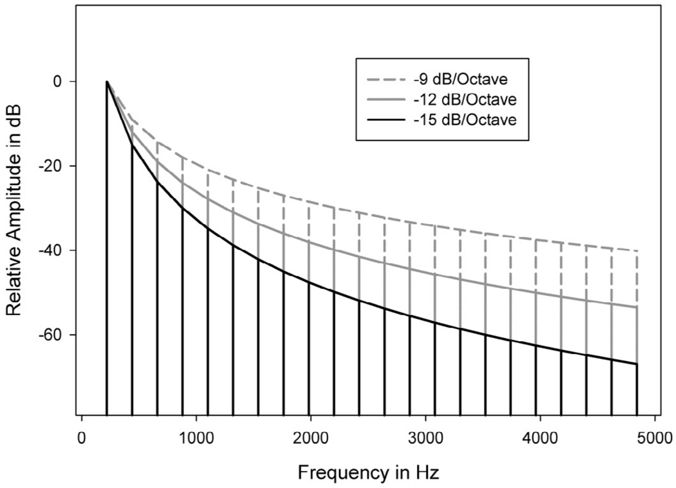

FIGURE 2. Source slopes of −9, −12, and −15–dB/octave.

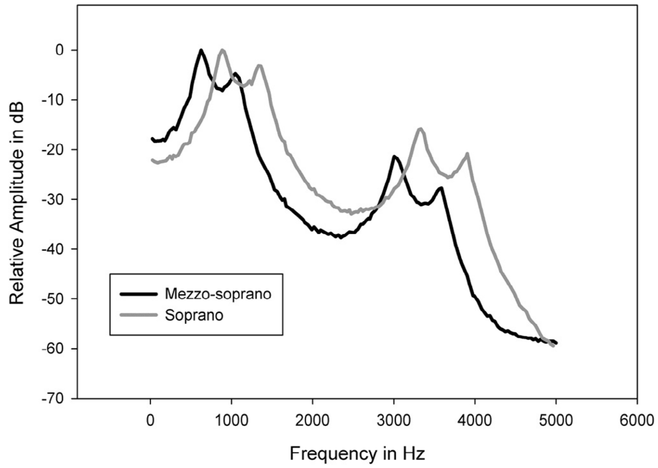

FIGURE 3. Transfer functions produced by formant patterns M (mezzo-soprano) and S (soprano) in response to white noise input.

TABLE 1.
Formant Frequencies for Mezzo-soprano and Soprano Stimuli

Pattern	F1	F2	F3	F4
M	625	1074	3027	3600
S	878	1367	3320	3906