

A comparison of the acoustic vowel spaces of speech and song*

Evan D. Bradley

(The Pennsylvania State University Brandywine)

Bradley, Evan D. 2018. A comparison of the acoustic vowel spaces of speech and song. *Linguistic Research* 35(2), 381-394. Speaking and singing are two modes of the same system. These modes are subject to similar constraints, but have different goals. This study examined the acoustic vowel spaces, as defined by formant frequencies, used by singers when they were singing or speaking the same linguistic content. Overall, formant values decreased during singing compared to speaking. This resulted in compression of the vowel space, with more overlapping vowel regions during singing. However, this was not consistent for all vowels and all singers. Differences between the modes are partially explained by known articulatory processes used during singing, such as larynx lowering. This may reflect the way that speakers balance communicative versus aesthetic concerns when articulating lyrics. (The Pennsylvania State University Brandywine)

Keywords singing, music, vowels, formants, speech

1. Introduction

Speaking and singing use the same vocal apparatus, but to very different effect. While the articulatory and acoustic properties of speech cause the speaker to balance perceptibility and articulatory effort, these pressures are moderated in singing by additional concerns for consistent resonance, expression, and style, and there is some evidence that vowels undergo articulatory modification in singing (Howard and Collingsworth 1992). In addition, the acoustic properties of singing vary by gender, singer, and singing style (Bloothoof and Plomp 1984, 1986); vocal training (Sundberg, Fahlstedt, and Morell, 2005); and whether the singing is infant-directed (Trainor et al. 1997). This study aimed to directly

* This project would not have been possible without the assistance of the Xi Mu Chapter of Phi Mu Alpha Sinfonia and Jeffrey Heinz. I also appreciate the helpful comments of two anonymous reviewers which improved the paper.

compare the acoustic properties of spoken and sung vowels, and to compare the vowel spaces of spoken and sung registers in order to better determine the factors which influence vowel modification in singing.

Sundberg (1987: 93–133) provides a review of research on the articulation and acoustic properties of singing, especially as they relate to characteristics of the singer. Men and women differ in average phonation frequency and vocal tract length. This difference in vocal tract length, however, is not to scale. The average mouth length of a woman is 85% that of the average male, while the average female pharynx length is only 77% that of the average man, meaning that the pharynx to mouth ratio differs between men and women (Sundberg 1987: 102, citing Nordström 1977). This partially, but not fully, explains differences in formant frequencies between men and women, which may partially be accounted for by 'sexolects', or gender-dependent articulation. Although F_0 is the largest determinant of the gender of a speaker, differences in formant frequencies also contribute to the voice quality of men and women, and consequently to vocal timbre in singing. This is due in large part to differences in the fourth formant resulting from women's narrower larynx tubes; alto singers, for instance, have a higher fourth formant than do tenors, even though the lower formant frequencies are similar. Because the third and fourth formants are more independent of vowels than are the first and second, these formants are more similar across a range of articulations, and their proximity in men contributes to a harsher vocal quality, by increasing the amplitude of the partials between them as they near. Within men, voice category (*e.g.*, tenor or baritone) also contributes to differences in formant frequencies (in addition to phonation frequency), and individual vowels appear to be articulated differently by tenors, baritones, and basses (Sundberg 1987: 110). These differences seem to be similar to the differences between men and women, which suggests that tenors and basses have different pharynx lengths.

Larynx height varies in normal speech, and is associated with vowel identity (Sundberg 1987: 97). The larynx tends to be raised during the pronunciation of vowels produced with spread lips (such as /i/) and lowered during the pronunciation of rounded vowels. Sundberg explains that this is because the acoustic effects of changes in lip rounding and larynx height are similar, so adjusting larynx height reduces the amount of lip rounding needed and makes

articulation easier. Larynx height also increases with phonation frequency in speech, as well as in untrained singing. Trained singers, however, aim to maintain a generally low larynx position, and in fact, larynx height in trained singers decreases slightly as pitch rises (Sundberg 1987: 113). A lowered larynx lengthens the vocal tract and consequently lowers its resonances.

The effect of larynx lowering causes F1 and F2 of most vowels to be lowered, toward a set of values similar to /æ/ (Sundberg 1987: 114). The effect on F1 is greatest for /a/ and /æ/ (the lowest vowels, and hence those with the highest F1), and the effect on F2 is greatest for /i/ and /e/ (front vowels with high F2). Variance in larynx height specifically changes the length of the pharynx tube, and the fact that the formant changes seen are similar to the differences between men and women suggest that pharynx length contributes to this gender difference. The fourth resonance of the vocal tract is determined by its length, but the exact fourth formant frequency is primarily determined by the shape of the larynx tube. The fourth formant is a major factor in determining individual voice timbre. Larynx height and the fourth formant also play a role in producing the *singer's formant* (Sundberg 1987: 101).

One result of this manipulation of larynx height is the production of the so-called *singer's formant* (Sundberg 1987: 118–124). In addition to the effects on F1 and F2, a lowered larynx has differential effects on the higher formants. In male opera singers, the fourth formant becomes much closer to the third than in speech, and the fifth formant lowers to a level at or below that of the fourth formant in speech. This clustering of formants greatly increases the energy in this frequency range (around 3,000 Hz).

One probable reason for the production of the singer's formant is loudness. Although trained singers do produce a higher sound pressure level than novices, they can increase the perceived loudness of their voice by manipulating formants to increase energy within certain parts of the frequency spectrum (Sundberg 1987: 120). The singer's formant, which may be around 3,000 Hz, is much higher than the spectral peak of an orchestra, which partly explained why an unamplified opera singer can still be heard above the accompaniment of the orchestra. This seems to be a more efficient means of increasing loudness than only increasing total sound pressure level. The singer's formant is affected by vowels, as well; this formant in vowels with a high F2 (e.g., /e/, /i/) has a

greater amplitude than that in vowels with a lower F2 (e.g., /u/).

The singer's formant is more prominent in men than in women, and more prominent in altos than in sopranos. This may be due to the high phonation frequencies sung by female singers, which may exceed the formant frequencies of their vocal tract. Instead of 'wasting' these resonances, sopranos appear to use an alternative strategy to increase loudness; instead of using larynx height to produce the singer's formant, they increase the jaw opening to raise F1 to the level of phonation frequency (Sundberg 1987: 124).

1.1. Questions

The present investigation aimed to answer two primary questions:

1. Manipulation of larynx height and other song-specific articulations appear to have differential effects on each vowel. How does this change the overall acoustic vowel space and the relationship between vowel classes as compared to speech?
2. Do gender, vocal training, and other factors influence or determine the acoustic vowel space of a singer?

In order to address these questions, the formant frequencies of spoken and sung vowels were measured from the same speaker/singers with particular attention to F1 and F2, which define the traditional acoustic vowel space. The previous findings discussed above regarding the consequences of larynx lowering lead to several predictions about the vowel space of singing:

1. Overall, formants will be lower in sung vowels, as compared to spoken vowels.
2. As vowels migrate, the variance in formant values will be lower in singing than in speech.
3. Low vowels will show the greatest changes in between speech and song, while front vowels will show the greatest changes in.
4. If the larynx control that causes these changes requires training, then amateur singers should have vowel spaces which are more similar to their

speaking voices than are those of trained singers.

2. Procedure

All procedures were approved by the University human subjects review board.

2.1 Singers

Fifteen singers (8 women) were recruited from the University community. Their ages ranged between 18 and 77 years (mean = 25.8). Singers were classified as 'professional' or 'amateur' based on their previous vocal training. 'Professional' singers had at least five years of solo vocal training. 'Amateur' singers had various levels of singing experience, but no formal, individual vocal training. All singers had choral or solo singing experience, but only singers who had received more than five years of instruction in solo singing and were currently studying voice were classified as 'professional'. All professional singers were studying voice at the college level. Table 1 summarizes the characteristics of the singers studied.

Table 1. Singer characteristics

subject	gender	age	range	experience
1	female	21	soprano	professional
2	female	19	soprano	professional
3	female	18	soprano	professional
4	female	19	soprano	professional
5	female	27	soprano	professional
6	female	52	soprano	professional
7	female	18	alto	professional
8	female	21	alto	professional
9	male	19	tenor	professional
10	male	28	baritone	professional
11	male	20	baritone	amateur
12	male	23	baritone	amateur
13	male	18	baritone	amateur
14	male	19	baritone	amateur
15	male	77	bass	amateur

2.2 Music

The piece selected was an American folksong, *Shenandoah* (Lomax, 1960). The piece was chosen for its simple melody, comfortable range for many voice types, and ballad style, which allows for easy vowel measurement. The lyrics were adapted by the author to create additional tokens of the vowels of interest (Figure 1).

O Shenandoah, I long to hear you.
o ʃɛnændo aɪ lɔŋ tu hɪr ju

Away, you rolling river.
əweɪ ju ˌrɔlɪŋ ˌɪvə

O Shenandoah, I long to hear you.
o ʃɛnændo aɪ lɔŋ tu hɪr ju

Away, we're bound away,
əweɪ wiː bəʊnd əweɪ

Across the wide Missouri.
əkɹɔs ðə waɪd mɪzɔɪ

Figure 1. Lyrics and IPA transcription of *Shenandoah*, adapted from Lomax (1960: 53)

2.3 Recording and analysis

Singers were recorded in a sound-attenuated booth or another quiet environment. All singers were recorded with a Labtec AM-22 microphone connected to a laptop computer. Singers were instructed first to read the lyrics of the piece, as if to an audience in the manner of a poem. Then, they sang the piece, as if for an *a cappella*, unamplified performance.

Recordings were transcribed and segmented. Pitch and formant frequency values of each vowel were measured using Praat (Boersma and Weenink, 2010).

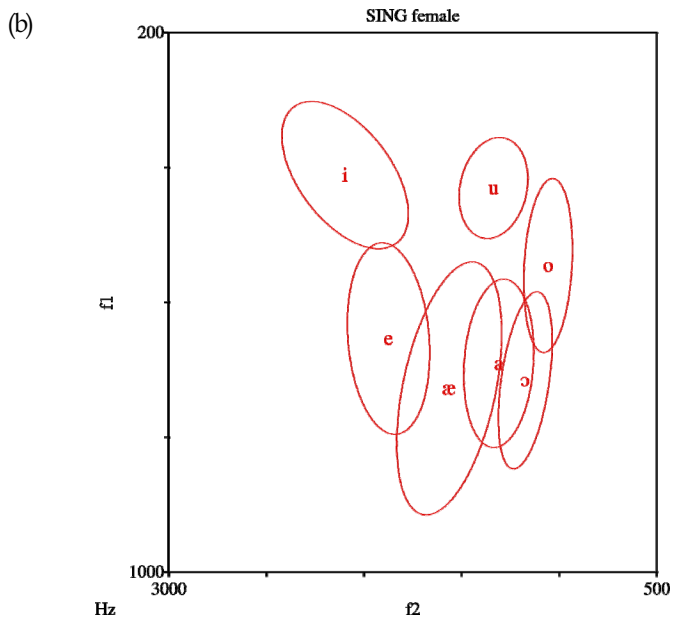
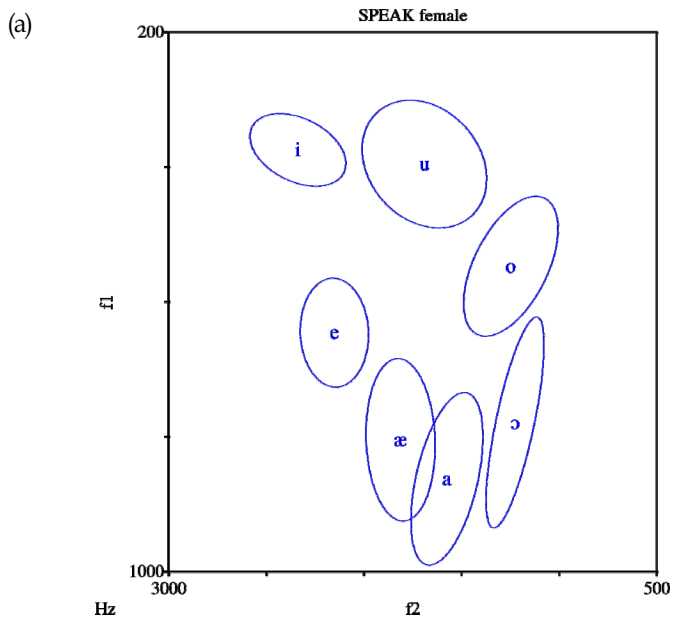
Monophthongs were measured near the midpoint of the vowel. Each segment of a diphthong was measured independently (though the second portion was not included in the analyses discussed here). F0 and the first four formants were recorded. Recording quality was not sufficient to consistently measure the fifth formant.

3. Results

Because F1 and F2 are the most important for determining the identity of a vowel, the analysis will focus on these. I will also focus on a set of cardinal vowels (a, æ, e, i, o, ɔ, u), and ignore reduced vowels, non-point vowels, and diphthongs. The dataset included 23 vowels from each participant in each mode (4 /a/, 2 /æ/, 3 /e/, 3 /i/, 5 /o/, 3 /ɔ/, 3 /u/). All figures were created using Praat (Boersma and Weenick, 2010).

Figure 2 summarizes the vowel spaces for male and female participants separately during speaking and singing. Figure 3 directly illustrates changes in each vowel between the two modes. Overall, formant values were lower during singing, and their variance also decreased (Table 2).

Formant values for each vowel were entered into separate mixed-effects analyses of variance with mode (singing or speaking) as a fixed effect, and subject as a random effect. F1 was lower during singing compared to speaking, $F(1,15) = 5.1788$, $p = .038$, as was F2 $F(1,15) = 46.518$, $p < .001$. Although the change in F1 was significant at the $p < .05$ level, the effect on F2 was considerably greater.



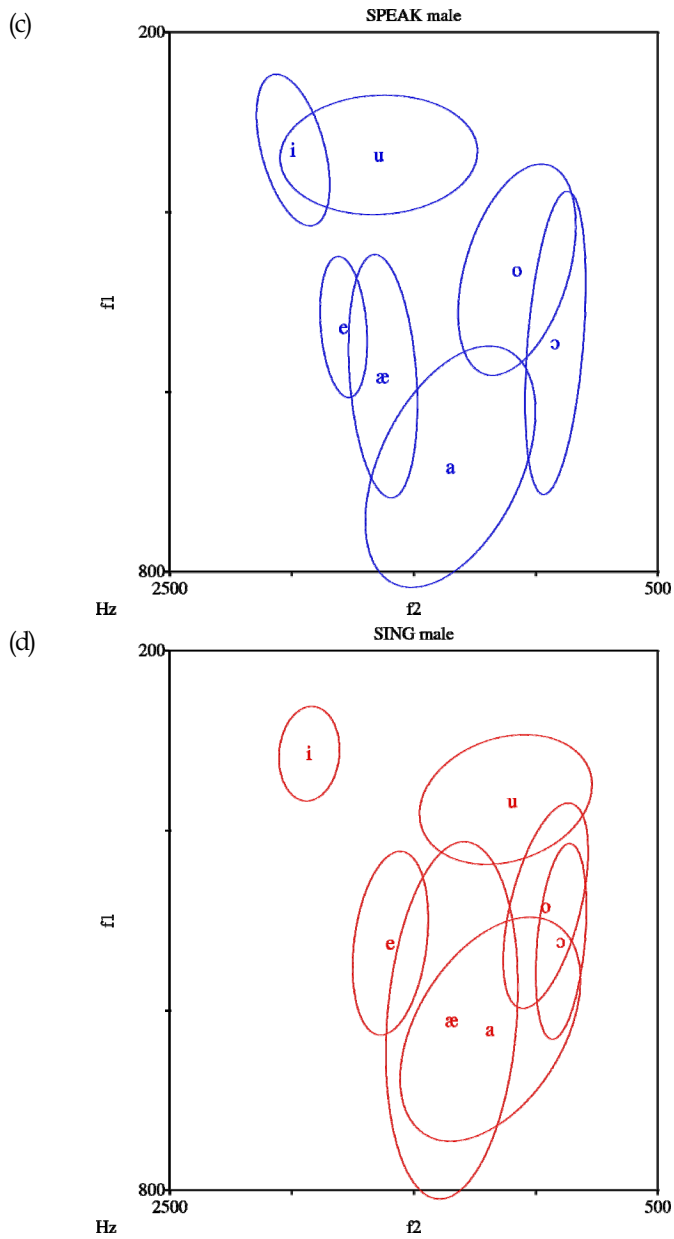


Figure 2. Acoustic vowel spaces of singing and speech for male and female singers

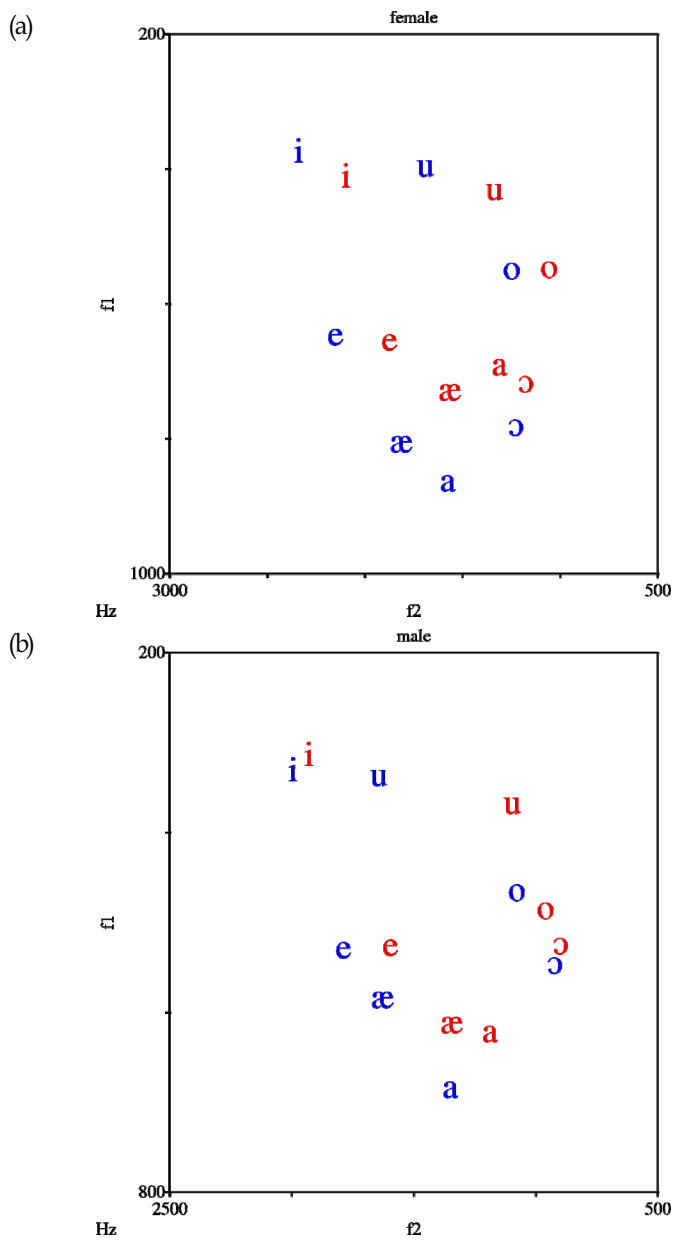


Figure 3. Mean formant values during speech (blue) and singing (red)

Table 2. Means and standard deviations in Hz for F1 and F2 during singing and speaking

	F1			F2		
	female	male	total	female	male	total
speaking	625 (210)	497 (166)	565 (201)	1669 (453)	1438 (437)	1561 (460)
singing	580 (173)	486 (151)	536 (170)	1433 (400)	1244 (404)	1346 (413)

To examine potential differences by vowel class, the data were split into front (i, e, æ) versus back (u, o, ɔ, a) and low (a, æ, ɔ) versus high (i, u). Due to the importance of F2 to frontness, F2 values were compared for front and back vowels by running separate mixed-effects analyses of variance, as above, for the front and back vowel sets. The effect of mode on F2 was significant for both front ($F(1,15) = 35.716$, $p < .001$) and back ($F(1,15) = 56.602$, $p < .001$) vowels. F1 was examined in a similar manner for the high and low vowel sets. Although the effect of mode was significant for both high ($F(1,15) = 6.7741$, $p = .02$) and low vowels ($F(1,15) = 27.004$, $p < .001$), the effect on low vowels was greater.

Differences between male and female singers were examined by further splitting the data by gender. The effect of singing on F2 was significant for both men ($F(1,7) = 17.555$, $p = .004$) and women ($F(1,8) = 29.578$, $p = .004$). The most prominent difference between male and female singers is that the marginally-significant effect of singing on F1 observed across all participants appears to be driven by a small effect for female singers ($F(1,8) = 5.7495$, $p = .04$), while there was no significant difference in F1 between singing and speaking for male singers ($F(1,7) = 0.3562$, $p = .6$).

4. Discussion

Overall, the effect of singing is a shift and compression of the acoustic vowel space, as well as a reduction in variance for most vowels. The shift results from a lowering of formant frequencies, which is consistent with the observation that singers often maintain a lowered larynx during singing. The compression of the vowel space results primarily from a greater degree of change in F1 values for

low vowels; F2 lowers significantly for both front and back vowels, although this may be driven by the fact that /u/ is pronounced fairly centrally in speech, and thus moves back significantly during singing.

It was hypothesized that less-experienced (or less formally-trained) singers would display less difference in their vowel spaces during speaking and singing. Gender and experience differences are difficult to separate in these data, given that the sample female singers had a higher degree of professionalism than the male singers. The overall trend of amateur vs. professional singers is in line with the hypotheses, because the more amateur male singers showed less movement of F1 during singing (Figure 2 (b) vs. (d)), which could result from less explicit control of articulation (such as larynx height) during singing. However, the fact that the male and female singers sing in different registers could affect these articulations as well.

The differences observed in the acoustic vowel space of singing may result from aesthetic choices on the part of the singer. Singing is not simply speaking to a tune, but carries with it other goals, such as maintaining consistent resonance, which may be prioritized over the typical pronunciation of words. Some of the changes that result from singing, such as reduction in the variance of formant values, could actually aid in the intelligibility of the lyrical content; schwa deletion also occurs in speech, dependent on stress, phonological environment, and lexical frequency (Ryu and Hong 2013), but this is unlikely to occur in singing (or to occur differently) due to rhythmic constraints. Other effects of singing observed in this study, such as vowel space shift and compression (possibly resulting in greater overlap of vowel categories) could interfere with word recognition.

The choice to use the same text in each mode allowed precise control over the content, as well as direct comparison of the same set of words/vowels. However, this design may have unintentionally obscured or reduced some differences between singing and speech. Being song lyrics, the words in the text were selected for certain aesthetic qualities, and likely do not exactly resemble natural, spontaneous speech. Also, reading the lyrics of a song out loud may cause speakers to engage in a more performative mode of speech which is different from more typical, conversational speech. The consequences of this are not entirely clear, but could result in speech patterns that are more like singing

(*e.g.*, fewer reduced vowels) than in normal speech.

5. Conclusion

This study directly compared the acoustic properties of vowels in the sung and spoken modes in order to determine what effects the mode (speech or song) has on the distribution of vowel categories within the vowel space, and how characteristics of the singer may affect the articulation of vowels during singing. The acoustic vowel space used in singing shifts from that which occurs in speech, resulting in a vowel space which is both higher and backer compared to that in speech, due to a lowering of the first and second formants; further, the variance of formant values was lower in singing than in speech. Female and male singers differed somewhat, primarily in male singers showing little difference in second formant values between speech and singing. Some of the acoustic effects are predictable given knowledge about singers, and what kinds of articulatory changes singers are likely to produce, especially larynx lowering.

These changes may have consequences for the intelligibility of lyrical content. Future work in this area should examine this perceptual question by measuring the effects of various vocal techniques on the ability of listeners to perceive sung vowels. Further, this research has only examined two points (ballad singing and text reading) along what may be a multi-dimensional continuum of vocal modes, each of which may balance communicative and aesthetic goals differently; future research should explore these issues using a wider range of song styles (*e.g.*, operatic, popular) and speech types (*e.g.*, conversation, public speaking), and genres which may fall in between these (*e.g.*, rap lyrics).

References

- Bloothoof, Gerrit and Reinier Plomp. 1984. Spectral analysis of sung vowels. I. Variation due to differences between vowels, singers, and modes of singing. *Journal of the Acoustical Society of America* 75(4): 1259-1264.
- Bloothoof, Gerrit and Reinier Plomp. 1986. Spectral analysis of sung vowels. III.

- Characteristics of singers and modes of singing. *Journal of the Acoustical Society of America* 79 (3): 852-864.
- Boersma, Paul and David Weenink. 2010. Praat: Doing phonetics by computer. (Version 5.1.25)
- Howard, David, and Jean Collingsworth. 1992. Voice source and acoustic measures in singing. *Acoustics Bulletin* 17(4): 5-12.
- Lomax, Alan. 1960. *The folk songs of North America*. Doubleday.
- Ryu, Na-Young and Sung-Hoon Hong. 2013. Schwa deletion in the conversational speech of English: The role of linguistic factors. *Linguistic Research* 30(2): 313-333.
- Sundberg, Johann. 1987. *The science of the singing voice*. DeKalb, IL: Northern Illinois University Press.
- Sundberg, Johan, Ellinor Fahlstedt, and Anja Morell. 2005. Effects on the glottal voice source of vocal loudness variation in untrained female and male voices. *Journal of the Acoustical Society of America* 117(2): 879-885.
- Trainor, Laurel J., Elissa D. Clark, Anita Huntley, and Beth A. Adams. 1997. The acoustic basis of preferences for infant-directed singing. *Infant Behavior and Development* 20(3): 383-396.

Evan D. Bradley

Department of Psychology
The Pennsylvania State University Brandywine
25 Yearsley Mill Road, Media, Pennsylvania, 19063 USA
E-mail: evan.d.bradley@psu.edu

Received: 2017. 12. 13.

Revised: 2018. 05. 02.

Accepted: 2018. 05. 02.