Music and Lyrics Interactions and their Influence on Recognition of Sung Words: An Investigation of Word Frequency, Rhyme, Metric Stress, Vocal Timbre, Melisma, and Repetition Priming

RANDOLPH JOHNSON
School of Music, Ohio State University

DAVID HURON[1]
School of Music, Ohio State University

LAUREN COLLISTER
Department of Linguistics, University of Pittsburgh

ABSTRACT: This study investigated several factors presumed to influence the intelligibility of song lyrics. Twenty-seven participants listened to recordings of musical passages sung in English; each passage consisted of a brief musical phrase sung by a solo voice. Six vocalists produced the corpus of sung phrases. Eight hypotheses derived from common phonological and prosodic principles were tested. Intelligibility of lyrics was degraded: (i) when archaic language was used; (ii) when words were set in melismatic rather than syllabic contexts; (iii) when the musical rhythm did not match the prosodic speech rhythm; and (iv) when successive target words rhymed. Intelligibility of lyrics was facilitated: (i) when words contained diphthongs rather than monophthongs; (ii) when a word from an immediately previous passage reappeared; (iii) when a syllabic setting of a word was preceded by a melismatic setting of the same word. No difference in word intelligibility was observed between musical-theatre singers and opera singers.

THE most popular forms of music typically involve the human voice. In nearly all cultures, singing is one of the preeminent forms of music making. Although singing often employs nonsense words or vocables, most singing employs lyrics conveying a narrative or poetic text. Nevertheless, concertgoers and music listeners frequently note that it is difficult to recognize or comprehend musical lyrics. In earlier work, we measured the intelligibility of sung English text. For advanced voice student performances, sung passages showed a 75 percent decrease in intelligibility compared with spoken counterparts (Collister & Huron, 2008). The number of errors in lexical item identification for sung words proved to be 7.3 times the number of errors for spoken words. Post hoc analyses of the word-identification errors identified a number of common confusions. Three-quarters of phonetic errors involved the misidentification of consonants. Among vowels, roughly one-third of identification errors arose from *centralization*, as when "beat" is misheard as "bet." Another phonetic error was *diphthongization* in which monophthongal vowels were misheard as diphthongs, such as mishearing "toe" as "toy."

In the current study, we attempted to further isolate some possible factors that may either confound or facilitate vocal intelligibility in sung passages. Specifically, we put forward the following hypotheses:

- H1: High frequency (*vernacular*) words are more intelligible than low frequency (*archaic*) words.
- H2: Words containing diphthongs are less intelligible than words without diphthongs.

- H3: Melismas (passages where a single syllable is sustained across more than one note) reduce intelligibility compared with syllabic settings (where each syllable is assigned to a single note).
- H4: Intelligibility is improved when syllable stresses are aligned with musical stresses.
- H5: Intelligibility is improved when the same word appears in the previous melodic phrase.
- H6: A melismatic setting is more intelligible when preceded by a syllabic setting of the same word.
- H7: Intelligibility is improved when successive words rhyme.
- H8: Musical theatre voices are more intelligible than operatic voices.

With **hypothesis one**, we examined how the lexical frequency of words affects their recognition in sung contexts. We predicted that vernacular words tend to be more intelligible than archaic words for two reasons: first, word frequency studies (e.g., Balota & Chumbley, 1985) have demonstrated that frequency is a salient factor in recognition of words based on phonetic material; second, because vernacular words are more likely to be concept *prototypes*. Prototype words are salient representatives of a particular concept, and prototypes provide a basis to which other concept members (or potential members) are compared (Rosch, 1999). A prototype word is usually the first word that is listed when someone is asked to name a representative (e.g., "apple") of a concept such as "fruit." Given this primed status of prototype words, it seems reasonable to predict that in word-recognition tasks involving degradation (e.g., singing), an archaic lexicon containing fewer prototype words would be less easily recognized than a lexicon of vernacular words containing relatively more prototype words.

Hypothesis two was designed to explore an observation from our previous study (Collister & Huron, 2008): our post hoc analysis showed that vowel diphthongization was a common identification error. Diphthongs are dynamically changing phonemes that begin with one vowel and morph or glide into another. An example occurs in the word "boy" whose vowel begins with a "oh" and ends with an "ee." Diphthongization occurred when participants incorrectly heard a monophthongal word as a diphthongal word. This finding is similar to the effect of *centralization*—when participants mistake a vowel produced by a remote tongue position as a vowel produced by a less remote tongue position (Collister & Huron, 2008; Hollien, Medes-Schwartz, & Nielsen, 2000). On hearing monophthongs, perhaps listeners perceive small phonemic deviations from the prototypical pronunciation as an illusory vowel glide—making a pure vowel seem like a diphthong.

In our 2008 study, we did not manipulate any vowel-type variables. However, in the present study, we directly compared the intelligibility of monophthongal and diphthongal words. We predicted that diphthongal words are less intelligible than words not containing diphthongs. Diphthongs consist of more than one pure vowel component connected by a vowel glide. Since there are several pure vowel components in a diphthong, its possible that deviations from each prototypical vowel can multiply the number of potentially perceived vowels in each part of the diphthong.

When diphthongal words are sung, there is an additional factor that may affect their intelligibility. When producing a diphthong, singers are taught to sustain the initial vowel for the duration of a tone and to glide to the final vowel only at the end of the tone. For example, in singing the word "toy," a singer will sustain the "oh" vowel and delay the glide to the "ee" until the termination of the tone. The effect might be represented as: "toe ... oh ... oh ... ee." Notice that the majority of time is spent singing the initial vowel, suggesting that listener might well mistake the intended diphthong-containing word ("toy") for a related monophthongal word ("toe"). In summary, since diphthongal words contain more opportunities for perceived vowel shifts to occur, and singing emphasizes one vowel component of a diphthong over the other vowel component, we predicted that diphthongal words are less intelligible overall than monophthongal words.

Hypothesis three was a consideration of the effect of text setting type on word intelligibility. When words are set to music, a distinction can be made between *melismatic* and *syllabic* settings. A melisma occurs when more than one note is sung to a single syllable. Melismas are common, for example, in Gregorian chant where a florid sequence of pitches is associated with a relatively small number of syllables. By contrast, a setting is said to be syllabic when each syllable is assigned to one and only one note.

Since melismas cause the vowel portion of a syllable to be sustained for a much longer duration than occurs in normal speech, we predicted that melismatic settings are less intelligible than syllabic settings of the same word.

In **hypothesis four**, we investigated the ramifications of the strong syllable stresses in the English language. In English, stress is occasionally used to distinguish lexical items as in the case of CON-tent and con-TENT. Two-syllable words may exhibit either a trochaic (strong-weak) rhythm, or an iambic (weak-strong) rhythm.

Music also exhibits stress. For example, beats are distinguished as either stronger or weaker. Depending on the degree of prosodic-musical alignment in songs, musical rhythms may or may not reinforce the text's prosodic rhythm (Huron & Royal, 1996). Burleson (1992) noted that influence of musical stress might interfere with prosodic rhythm; the present study examined this intuition empirically.

Figure 1 illustrates two melodies in which the linguistic stresses are aligned with the music (Figure 1a), or misaligned with the music (Figure 1b). We predicted that the stress-matched stimuli are more intelligible than the stress-mismatched stimuli.

Figure 1a. Example of an iambic word ("release") that is aligned with the music's metric stress.

Figure 1b. Example of a trochaic word ("really") that is *misaligned* with the music's metric stress.

Hypothesis five stemmed from the preponderance of repetition in musical form and text setting. It is common in vocal music to repeat lyrics verbatim. For example, individual words may be repeated, lines of text may be duplicated, or entire refrains may reappear periodically. Psychological research has repeatedly shown a strong *repetition priming* effect in which a prior presentation of a stimulus leads to an enhanced processing of that stimulus in following presentations. This enhanced processing could be anything from memory recall to identification of a degraded stimulus; priming has been shown to apply to words and sounds (see Bigand, Tillman, Poulin-Charronnat, & Manderlier, 2005 for background on repetition priming).

Repetition priming in music has not been studied extensively, but several studies suggest that repetition effects in music might be domain specific and even deviate from models of priming in spoken word stimuli. Bigand *et al.* (2005) studied repetition priming of musical harmonies and found a relatively small repetition priming effect compared to harmonic relation priming. Vocal music combines words with music's many elements: repeated melodies, chords, and rhythms, for example. Since word repetition priming is a robust effect, we predicted that word repetition facilitates intelligibility, beyond whatever other effects arise from various forms of musical repetition.

Hypothesis six involved testing the combined effect of repetition (H5) and syllabic setting (H4). We predicted that a melismatic setting is more intelligible when preceded by a syllabic setting of the same word. This hypothesis stems from the predictions of hypotheses 4 and 5.

Hypothesis seven related to a common feature of musical lyrics—the presence of rhymes. Copeland and Radvansky (2001) measured word recall in both word span (isolated words) and reading span (words in sentences) tasks. When words are in the context of a sentence, it appears that phonological similarities, such as rhyme, enhance word recall. In line with these findings, we predicted that successive rhyming words not only are memorized more easily, but also are more easily perceived in the first place.

With **hypothesis eight**, we explored the influence of vocal style on word intelligibility. There are many different styles of singing, such as pop, country & western, reggae, art song, Broadway, light opera, and so on. Given the differences in singers' training and musical goals, one might expect to see differences in overall intelligibility for different vocal styles. If word intelligibility investigations are limited to only one style, then the observed differences between speaking and singing will reveal only part of the story. The importance of style is implied in Smith and Scott's (1980) study, which found a 50% drop of intelligibility for vowels sung above F5; the same singer's use of a raised larynx limited the drop to only 10% (in the near proximity of F5). Larynx position is one factor that affects vocal timbre and ornament characteristics used in different styles of music. Although diction is an important element in the training of operatic voices, a very high premium is also placed on vocal tone color. In musical theatre, by contrast, greater priority is given to conveying the words to the audience. Thus, musical style may affect the clarity of word production. On the basis of these generalizations, we predicted that musical-theatre voices are more intelligible than operatic voices.

METHOD

In brief, the study involved recording purpose-made vocal phrases whose lyrics consisted of a standard carrier phrase followed by a target English word: "I am singing the word _____." These melodic phrases were then used as stimuli in an experiment in which listeners were asked to transcribe the target word for each stimulus. Some of the stimuli were also presented in spoken form, again with a standard carrier phrase: "I am speaking the word _____."

Stimuli

In Collister and Huron (2008), stimuli were generated by three university music students majoring in vocal performance. These singers were all trained in the Western operatic vocal tradition. In the current study, we endeavored to include a contrasting vocal style—namely, musical theatre. Aside from differences in training and sung repertoire, the most salient contrast between operatic and musical-theatre voices is timbre and vibrato. Operatic singers use greater amounts of vibrato, whereas musical-theatre singers use a straighter tone.

In addition to recruiting two operatic singers (one mezzo-soprano and one baritone), four experienced singers of musical theatre were also recruited—two females and four males representing soprano, alto, tenor, and bass vocal ranges.

All six vocalists rehearsed a set of melodic phrases with the aid of a piano coach. Stimuli were recorded using voice alone. The phrases consisted of the main themes from twenty vocal works. Phrases ranged in length from seven to twenty-eight notes with an average of 13.7 notes. Appendix II in Collister and Huron (2008) identifies the specific melodic themes used in the study. In addition, the Appendix identifies the number of notes in each phrase, as well as the number of syllables and notes assigned to the target word.

While recording these materials, it became clear that our musical-theatre vocalists had less formal musical training compared with the operatic vocalists. For example, three of the four musical-theatre vocalists had difficulty reading the musical notation, and required much more prompting from the piano coach in order to sing the required stimuli.

One hundred and twenty target words were selected according to the same criteria used in Collister and Huron (2008). Specifically, target words were sampled from a database containing 97,082 words representative of English song lyrics; the database was a composite of several classical and popular music word lists. Some hypotheses necessitated the addition of words not in the database (e.g., rhyming words). The number of syllables in target words ranged from 1 to 3. We recorded each vocalist singing 20 melodic phrases containing different target words. Then, each vocalist sang the same 20 melodic phrases with different target words. There were a total of 6 singers, who sang a combined total of 240 melodic phrases. Additionally, we recorded each vocalist speaking 20 target words in the context of spoken phrases. As a result, the experimental stimuli consisted of 360 recorded phrases—240 sung phrases and 120 spoken phrases.

Stimuli were recorded in an 800-seat auditorium/recital hall. Vocalists were positioned approximately 4 meters from the front edge of the raised stage, near the center. Recordings were made with stereo overhead microphones placed approximately 10 meters from the vocalist. Both the position of the

vocalists and the microphone placement coincide with normal performance practice for this auditorium. The recordings made in these circumstances were judged by the authors to be similar in reverberation and ambience to standard commercial music recordings.

The phrases were transposed so that the passages spanned the range of each vocalist. The twenty phrases were randomly assigned into three ranges: conceptually high, medium, and low. Each vocalist reported his/her "comfortable" vocal range. This range was then divided into three equal regions. For example, suppose our soprano reported a range from C#4 to A5—representing a span of 20 semitones. Three divisions might be defined, each separated from the others by 5 semitones. For this vocalist, "low" range was operationally defined as passages with a mean pitch of F#4 (i.e., 5 semitones above the lowest comfortable pitch). A "medium" range was operationally defined as passages with a mean pitch of B4 (i.e., 10 semitones above the lowest comfortable pitch). A "high" range was operationally defined as passages with a mean pitch of E5 (i.e., 5 semitones below the highest comfortable pitch). Passages were transposed so that no passage contained notes higher or lower than the comfortable range, and so that the passages could also be classified as high, medium, or low.

Participants

Twenty-seven participants were recruited for the experiment, 10 males and 17 females, aged between 18 and 48 with a mean age of 21.8 years. The participants were drawn from a convenience population of sophomore music students participating in an experimental subject pool in the School of Music at the Ohio State University. Voice majors were explicitly excluded from participation since singers may possess some special knowledge or experience that may make it easier (or possibly more difficult) to understand sung lyrics. ("Participants" refers to listeners during the experimental phase of this study; not to the singers used for creating the experimental stimuli.) As a result, experimental participants consisted of vocally and phonetically untrained undergraduate music students. Of course, these students may not reflect well the general population of music listeners. Because of their greater musical experience it is possible that music students may be more adept in "catching" the lyrics in vocal music. Conversely, music students might be more attentive to the melodic, harmonic and other aspects of music, and so may be less able to attend to lyrics than the general population.

As a screen for possible hearing loss, we used the Coren and Hakstian (1992) survey in lieu of an audiometric examination. Before conducting the experiment, an *a priori* hearing score of 6/10 was established so that participants who scored lower than this value were excluded from the experiment. Using this exclusion criterion, no subjects were eliminated.

Procedure

Participants sat in an acoustical isolation booth equipped with a computer monitor, computer keyboard, and speakers. The experiment began after the experimenter gave spoken instructions and a brief demonstration of the procedure. A written version of the instructions was displayed on the computer screen, and there were two practice trials before data collection began. The computer ran a program (written in Perl) that automatically played the recorded stimuli and prompted the participant to type in their responses after each stimulus. The program would present each stimulus only once and would pause until the participant typed a response and pressed the "enter" key. Participants worked alone in the sound booth with the door closed.

Each participant heard approximately 178 target words; 159 were sung and 19-21 were spoken (this varied depending upon the randomization process of words in each experiment). There were fewer spoken words used in the present study than in Collister and Huron (2008) because the experimental hypotheses mostly pertained to sung phrases. The 178 target words were selected from the total of 360 words via a computer program that randomly selected a representative subset, while maintaining some conditions, e.g., successive words, involved in the rhyming hypothesis. Each participant heard a unique random selection and sequence of stimuli.

Tests of Significance

In conceiving this research project, we generated a number of hypotheses concerning vocal intelligibility. Long established phonological research suggests a number of features that might be expected to influence

phonemic confusions. Some 24 hypotheses were originally identified. With sufficient data, we expected that the majority of the hypotheses would be borne out. However, the number of hypotheses was regarded as excessive.

Ideally, each hypothesis would be tested using independent stimuli, with different singers, different target words, different musical settings, and different listeners. However, this goal was deemed impractical. Although we made 316 separate recordings for this study, many of the hypotheses identified above were tested using just 8 or 16 stimuli. Unless the effect sizes were especially large, we expected that the assembled data would have insufficient statistical power to test some of the hypotheses.

Since the motivation for this research was to identify practical principles in text-setting, we resolved to test as many hypothesis as we could, knowing that this would sacrifice statistical power. Statisticians emphasize that there is nothing sacrosanct about either the 0.01 or 0.05 alpha levels. The choice of alpha depends on the moral risks associated with making a Type I error. For potentially life-threatening phenomena (such as occurs in medical research), alpha levels ought to be set very small. However, in the case of text setting, the moral risks attending making a false claim are less onerous. Accordingly, we resolved a priori to establish an alpha level of 0.1 for all statistical tests. That is, in deeming a hypothesis to be "statistically significant," we will accept a 1 in 10 risk of committing a Type I error, rather than the more common 1 in 20 risk. Weakening the alpha value has the salutary effect of allowing us to test a greater number of hypotheses from the same volume of data.

RESULTS

H1. Vernacular and Archaic Words

In order to test the archaic-words hypothesis (H1), fourteen stimuli were recorded: seven containing high frequency words, and seven containing low frequency words. Each archaic word was matched with a modern synonym, such as the pair "rogue" and "thief." Target words were recorded in both sung and spoken contexts. Table 1 shows the 14 words used.

Table 1. Seven archaic (low frequency) and common (high frequency) word pairs used in hypothesis one.

Low Frequency	High Frequency
thine	your
dale	hill
kith	friend
ire	rage
gall	spite
rogue	thief
gloaming	evening

Figure 2 compares the intelligibility of the vernacular (common) and archaic (uncommon) words. The dark bars show the proportion of correctly identified words for the sung context whereas the open bars show the proportion of correctly identified words for the spoken context.
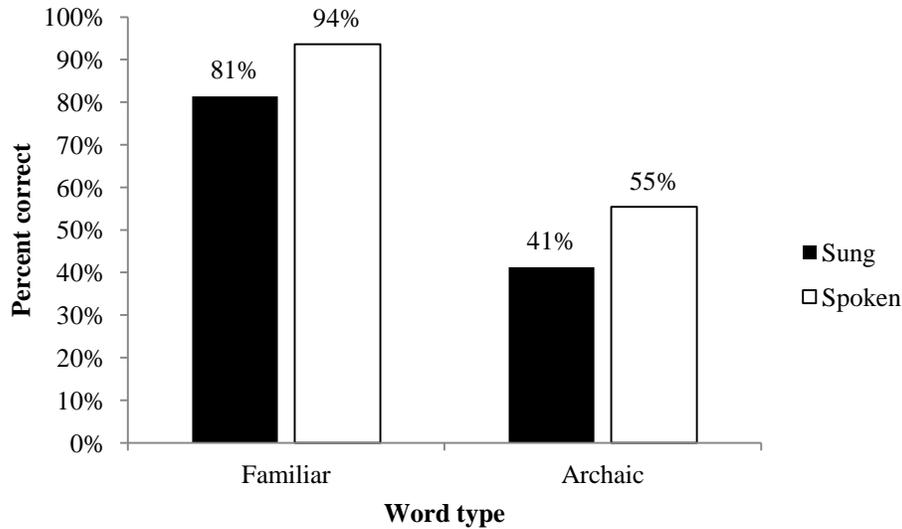
Figure 2. Percentage of correctly identified words for familiar words (428 correct out of 513 total) and archaic words (91 correct out of 189) in both the sung and spoken conditions.

As can be seen, vernacular words were more easily recognized than the archaic words for both the sung and spoken contexts. This difference was statistically significant (χ^2 = 87.4; *df* = 1; *p* < .001; *Phi* = .35; two-tailed test; Yates' correction for continuity applied here and hereafter). The poorest recognition was evident for the sung–archaic condition.

Since the different word types were presented to the participants in two different modes of delivery, it was possible that an interaction effect influenced word intelligibility. For example: is there a greater decline in intelligibility of archaic words from spoken to sung contexts than familiar words? To test the proportional data for interactions, we applied the arcsine transformation to the percent-correct values. This transformation helps prevent misleading results when running parametric tests on proportional data. The results were consistent with main effects observed in the above chi-square test and in Collister and Huron (2008): word type was a strong effect (F (1, 107) = 97.3; *p* < .001), and mode of word delivery was a moderately strong effect (F (1, 107) = 16.0; *p* < .001). The interaction between word type and word delivery fell short of statistical significance (F (1, 107) = 2.1; *p* = .15).

H2. Diphthongs and Monophthongs

Eight target words containing diphthongs were selected. For comparison purposes, each diphthong word was matched with two other words: one employing the initial vowel of the diphthong, and a second employing the final vowel of the diphthong. For example, a diphthong target word "lie" ("ah"—"ee") might be matched with the control words "law" ("ah") and "lee" ("ee"). In some cases, no word existed containing both the same consonants and the precise single vowel from the diphthong; in these cases, a word with the same consonants and a similar vowel was used. Twenty-four target and control words are shown in Table 2. We predicted that the intelligibility of each diphthong word would be lower than the average intelligibility of its two control words. Each minimal triplet (a set of three related words) was recorded using the same melody and sung by a single singer, so the stimuli differed only in the target word. In addition to the sung versions, vocalists also spoke all of the words in a common carrier sentence, "I am speaking the word ___".

Table 2. Minimal triplets consisting of a diphthong and its two vowel components placed into separate pure-vowel words. International phonetic alphabet (IPA) symbols of vowel sounds are listed after each word.

Diphthong	IPA	Initial	IPA	Final	IPA
boy	ɔi	bow	o	bee	i
foiled	ɔi	fold	o	field	i
bout	aʊ	bought	a / ɔ	boot	u
sale	eɪ	sell	ɛ	sill	ɪ
wait	eɪ	wet	ɛ	wheat	i
live	aɪ	love	ʌ	leave	i
might	aɪ	mat	æ	meet	i
gown	aʊ	gone	a / ɔ	goon	u

Table 3 shows the results. The results proved to be inconsistent with the hypothesis. In fact, diphthongs were more easily recognized than monophthongs: 16.4 percent of words containing diphthong vowels were misidentified, whereas 25.2 percent of words containing monophthong vowels were misidentified.

Table 3. Number of correct/incorrect identifications of words containing monophthongs or diphthongs. Sung and spoken words are combined.

	Correct	Incorrect	Total
Monophthongs	485	163	648
Diphthongs	356	70	426
Total	841	233	1074

A chi-square analysis indicated that this difference was statistically significant ($\chi^2 = 11.0$; $df = 1$; $p < .001$; *Phi* = .10; two-tailed test).

Figure 3 shows a comparison of the average word intelligibility for the diphthong and monophthong words. The dark bars show the percentage of correctly identified monophthong words whereas the open bars show the number of correctly identified diphthong words.
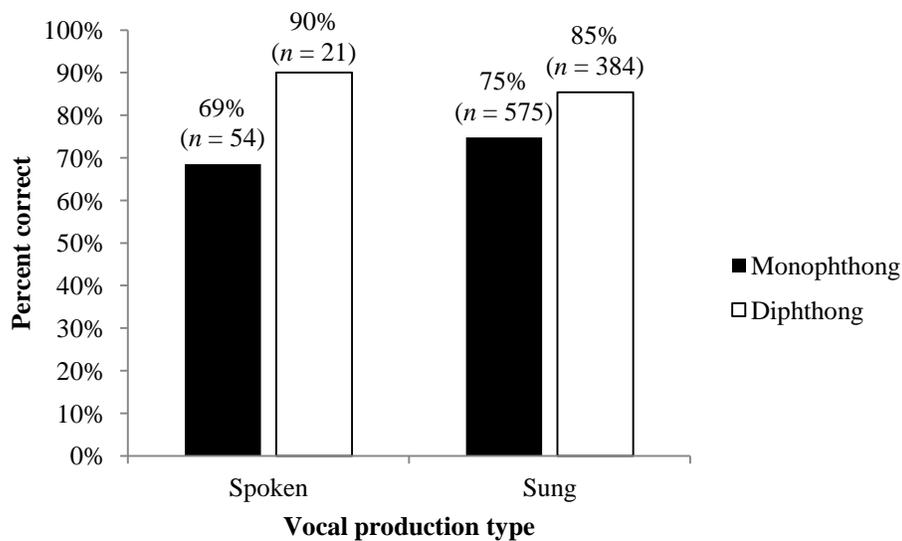

Figure 3. Percent correct word identifications in the spoken and sung conditions for words containing monophthongs or diphthongs.

When each diphthong word was compared with the average of its two control words, 6 of the 7 triplets showed a higher intelligibility for the diphthong word when sung. When spoken, 4 of the 7 triplets showed a higher intelligibility for the diphthong, and 3/7 showed equal intelligibility between the monophthong and diphthong. Overall, the results were inconsistent with the idea that sung diphthongs reduce intelligibility.

H3. Melismatic and Syllabic Settings

The two most experienced operatic singers (baritone and mezzo-soprano) each sang four unique melismatic passages. In addition, they sang the same four passages with the melismas removed. Table 4 displays the eight target words used in the stimuli, four one-syllable and four two-syllable words.

Table 4. One-syllable and two-syllable target words used in melismatic and syllabic contexts.

sing	floor	darling	summer
rain	fade	city	forget

Each listener heard each target word once, in either a melismatic or non-melismatic context (i.e., between-subjects design). Figure 4 compares the word intelligibility for the melismatic and syllabic settings.
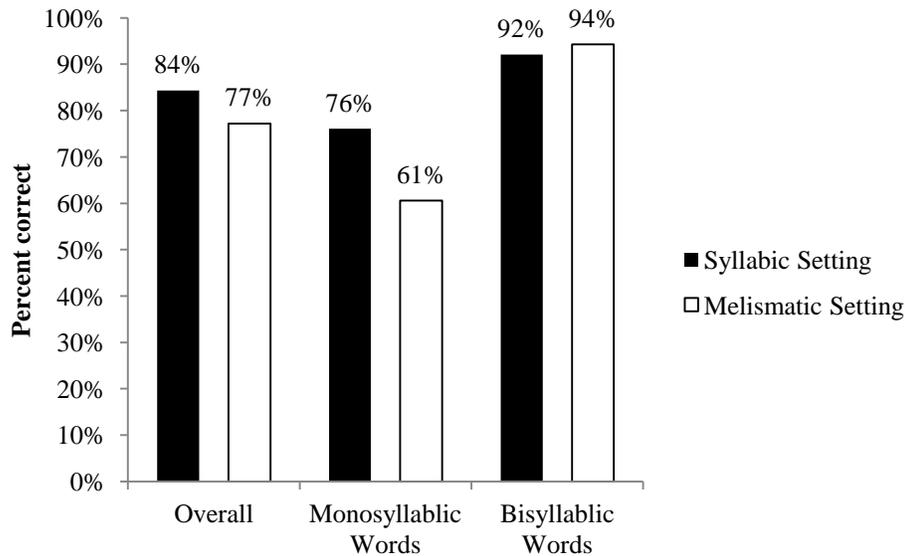

Figure 4. Correct word identification percentages for syllabic and melismatic settings of monosyllabic words and bisyllabic words.

In syllabic settings, 273 of 324 words (84%) were correctly recognized, whereas in melismatic settings just 250 of 324 words (77%) were correctly recognized. Using a chi-square test, this difference was statistically significant ($\chi^2 = 4.80$; $df = 1$; $p = .03$; *Phi* = .09; two-tailed test).

Additionally, there was a difference between the intelligibility of monosyllabic and bisyllabic words; syllabic and melismatic settings of bisyllabic words showed no significant difference, whereas melismatic settings of *monosyllabic* words had a noticeably lower identification frequency than syllabic settings of single syllable words. After collecting all of our data, we realized that the number of notes occurring in a melismatic presentation of a word could have confounded these results: if there were more notes in monosyllabic melismas than in the first syllable of a bisyllabic melismas, then the length of the melismas may have lowered the intelligibility of monosyllabic words set to melismas. Examining our stimuli, we did indeed find that the monosyllabic melismas were longer (mean of 5.0 notes) than the bisyllabic melismas (mean of 3.8 notes). This difference is consistent with the idea that the length of a melisma might account for the evident reduction in intelligibility. Alternatively, the difference in the

number of notes may be regarded as rather small and so the reduction in intelligibility might be due to the different number of syllables. Unfortunately, the data as collected cannot be used to distinguish between these two interpretations.

H4. Stress Match and Stress Mismatch

Participants heard sixteen passages containing bisyllabic target words. Eight of the target words had a trochaic rhythm (strong-weak) and eight of the target words had an iambic rhythm (weak-strong). In half of the settings, the syllable stresses were aligned with the musical stresses. In the remaining settings, the syllable stresses contradicted the musical stresses. Table 5 identifies the words used.

Table 5. Iambic and trochaic words used in stress-matched and stress-mismatched settings.

Iambic	*Trochaic*
quartet	quarter
disease	sweeping
release	really
instead	instant
myself	laughter
tonight	quiet
become	rhythm
received	morning

Figure 5 compares the word intelligibility for the stress-matched and stress-mismatched settings. As can be seen, intelligibility was much greater for the stress-matched settings.
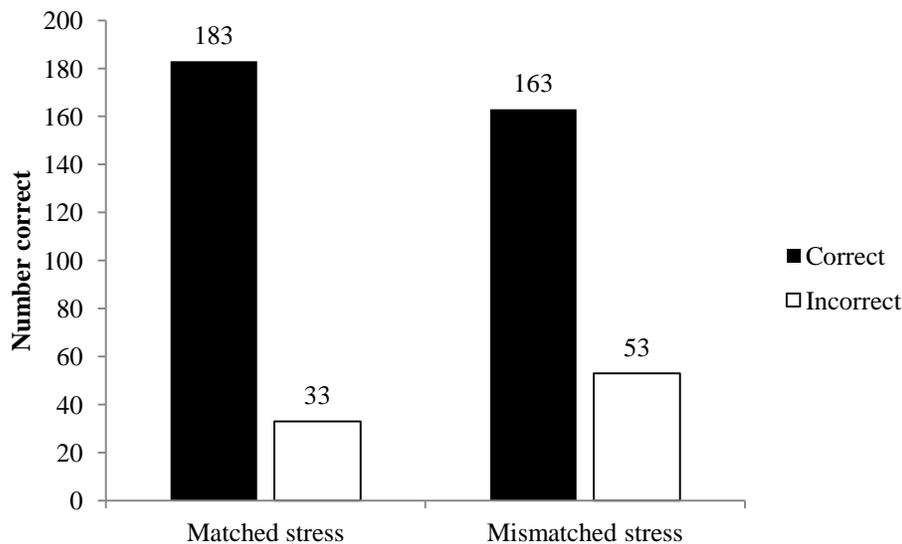

Figure 5. Number of correctly/incorrectly identified words in stress-matched and stress-mismatched settings (*n* = 216 trials for both conditions).

A chi-square analysis of the association showed a significant effect ($\chi^2 = 5.24$; *df* = 1; *p* = .01; *Phi* = .11; one-tailed). Stress-matched settings enhanced intelligibility. When the results were analyzed by stress type (iambic or trochaic), the individual results remained statistically significant at the 0.1 alpha level:

- For the iambic settings: $\chi^2 = 1.93$; *df* = 1; *p* = .08; *Phi* = .10; one-tailed.
- For the trochaic settings: $\chi^2 = 2.32$; *df* = 1; *p* = .06; *Phi* = .10; one-tailed.

For both iambic and trochaic rhythms the results were consistent with the hypothesis that intelligibility is facilitated when word stress is matched with musical stress.

H5. Successive and Delayed Repetition

Regarding the effect of word repetition on intelligibility, we considered two situations: *immediate repetition* where the same word appears in successive trials, and *delayed repetition* where the same word appears with a single intervening trial involving an unrelated word. Apart from these two conditions, the experiment originally was designed to distinguish further the effect of three different forms of repetition:

(1)   Intelligibility may be facilitated by simply repeating the same recorded stimulus twice in succession ("Exact repetition condition").
(2)   Intelligibility may be facilitated by hearing two different recordings of the same passage produced by the same singer ("Different recording condition").
(3)   Intelligibility may be facilitated by hearing two recordings of the same passage produced by different singers ("Different singer condition").

Despite having made these distinctions, lack of statistical power necessitated a simpler analysis in which all three forms of repetition were amalgamated.

IMMEDIATE REPETITION

First, we considered the effect of immediate repetition on intelligibility, including all conditions listed above. Table 6 shows the results for the exact repetition without delay condition:

Table 6. Correct and incorrect identifications of immediately repeated words.

2nd trial / 1st trial	Incorrect	Correct
Incorrect	33 *(both incorrect)*	32 *Facilitation*
Correct	21 *Degradation*	292 *(both correct)*

As can be seen, most of the stimuli were correctly identified, producing a borderline ceiling effect. Of 378 pertinent trials, in 292 instances the listeners identified both the initial and repeated presentations of a word correctly. There were 65 failures to recognize the first presentation of the word. Of these 65 instances, immediate repetition resulted in 32 subsequent facilitated recognitions. However, there were in addition, 21 instances in which the listener correctly recognized the first presentation of the word, and yet failed to recognize the second presentation. That is, of the 86 cases in which a change of recognition occurred, 32 showed improvement, 33 showed no improvement, and 21 showed worse performance (failing to recognize a word that they had previously recognized). The 2 □ 2 format of Table 5 might suggest the use of a chi-square test. However, in order to use this test, any observation should in principle be able to occupy any one of the four conditions; in this case, there are dependencies between the cells that render a chi-square test inappropriate for this task. For example, if a listener initially fails to identify the word, then the subsequent observation can occupy only one of two cells. An appropriate test would employ a *t*-test to account for covariance.

Having chosen an alpha level of 0.10 a priori, a one-tailed test in the case of the immediate repetition proved to be significant (t_{crit} = 1.28; t = 1.51; p < .1). The results were consistent with the hypothesis that immediate repetition facilitates word intelligibility.

DELAYED REPETITION

Presuming that this improvement arises from some sort of priming effect, a pertinent question to consider next is whether the facilitating effect of word repetition might still be observed when intervening words are interposed between the two presentations of the same word. If the effect is predominantly attributable to some sort of short-term (rather than intermediate-term) memory, then the facilitating effect of repetition ought to be weakened significantly when an unrelated word is interpolated. In the experiment, a number of stimuli were arranged so that some of the repeated words were separated by a single unrelated word (delayed repetition condition). Table 7 shows the pertinent results.

Table 7. Correct and incorrect identifications of words in delayed repetition.

2nd trial / 1st trial	Incorrect	Correct
Incorrect	21 *(both incorrect)*	11 *Facilitation*
Correct	12 *Degradation*	145 *(both correct)*

Once again, most of the stimuli were correctly identified. Listeners identified both the initial and repeated presentations of a word correctly in 145 of the 189 trials. There were 32 failures to recognize the first presentation of the word. Of these 32 instances, delayed repetition resulted in 11 subsequent facilitated recognitions. However, there were 12 instances in which the listener correctly recognized the first presentation of the word, and yet failed to recognize the later presentation. That is, of the 44 cases in which recognition occurred, 11 showed improvement, 21 showed no improvement, and 12 showed worse performance (failing to recognize a word that they had previously recognized).

Given that the number of facilitated recognitions is lower than the number of declines in performance, the results were not consistent with the experimental hypothesis (t_{crit} = 1.28; t = -0.2; p > .5). At face value, the results suggested that there appears to be no facilitation of word intelligibility in the delayed repetition condition. We might note, however, that the number of pertinent trials is half of the immediate repetition condition, so this negative result might simply reflect a lack of statistical power.

H6. Syllabic and Melismatic Settings of the Same Word

As we have seen, the results imply that repetition facilitates intelligibility, but apparently only in cases of immediate repetition. Earlier, we saw that melismatic settings are less intelligible than syllabic settings. This raises the question of whether a syllabic setting of a word facilitates the intelligibility of an ensuing melismatic setting of the same word.

In order to test this notion, each listener heard 8 pairs of stimuli; each pair used the same target word in a melismatic and syllabic version. In 4 of the 8 pairs, listeners heard the melismatic version immediately prior to the syllabic version. In the other 4 pairs, listeners heard the syllabic version prior to the melismatic version. Our prediction was that a preceding syllabic version will facilitate intelligibility of an ensuing melismatic version more than the comparable facilitation of hearing the melismatic version prior to the syllabic version. Table 8 displays the results:

Table 8. Order of repetition type (melismatic or syllabic) and its effect upon intelligibility.

First Presentation	Second Presentation	Improved (Degraded) Instances
syllabic	melismatic	4 (6)
melismatic	syllabic	18 (3)

As can be seen, the results were contrary to the hypothesis. That is, preceding a melismatic presentation by a syllabic presentation of the same word led to less improvement in recognition (more degradation, in fact), than having a melismatic presentation followed by a syllabic presentation. Using a 2 ⊠ 2 contingency table (order of presentation ⊠ amount of improvement/degradation) this interaction was found to be statistically significant ($\chi^2 = 4.83$; $df = 1$; $p = .03$; $Phi = .40$; two-tailed test).

The values in Table 8 give a misleading idea of the number of stimuli used in testing this hypothesis. In total, there were 216 word pairs presented (432 individual stimuli). However, in 185 instances, there was no change in word recognition despite changing the order of syllabic/melismatic presentation. That is, there were only 31 differences. The effect observed in the test of Hypothesis 3—melismas significantly degrade word recognition—seems to operate strongly even when the melisma is preceded by a syllabic setting of the same word.

H7. Rhyme and Non-Rhyme

We recorded 14 stimuli, consisting of 7 pairs or rhyming words. Each rhyming pair was constrained to include the same type of consonant, such as liquids, nasals, fricatives, or plosives. The rhyming pairs used are shown in Table 9. Each of the five singers sang three pairs of rhyming stimuli. Each subject heard seven pairs of rhyming words, e.g., could/good, so/show, and a greater number (which randomly varied for each participant) of non-rhyming pairs that included words from Table 9 (e.g., heel/good, guilt/van).

Table 9. Rhyme pair target words.

could	good
taught	bought
raw	law
so	show
ten	den
van	than
west	rest

If rhyme enhances intelligibility, then we would expect the second stimulus in a pair of successive rhyming stimuli to be more intelligible than the second stimulus in a non-rhyming pair.

The results of the experiment can be seen in Table 10, which compares the intelligibility of the pairs' second words in both rhyming and non-rhyming conditions.

Table 10. Frequency of correct and incorrect identifications of the second word in rhymed and unrhymed pairs.

	Correct	Incorrect	Total
Rhyme	110	60	170
Non-rhyme	268	94	362
Total	378	154	532

Successive rhyming words did not facilitate intelligibility in this experimental setting—in fact, successive rhyming words appeared to *degrade* the intelligibility of the second word relative to the non-rhyming condition. In the rhyme condition, there were more incorrect identifications (60) relative to correct identifications (110), than the non-rhyme condition where there were fewer errors (94) relative to correct identifications (268). A chi-square test showed that this difference is significant at $\alpha = 0.10$ ($\chi^2 = 4.45$; $df =$

1; *p* = .04; *Phi* = .09; two-tailed). The results were consistent with the notion that the second stimulus in a pair of successive rhyming stimuli is less intelligible than the second stimulus in a non-rhyming pair. We will address this unexpected result later in the Discussion.

Informally, we observed that many of the errors were confusions between voiced and voiceless consonants (e.g., "good" and "could"). Also, participants frequently perceived consonants at the ends of words having vowel terminations (e.g., "raw" mistakenly heard as "wrong"). Further discussion of these types of errors can be found in Collister and Huron (2008).

H8. Effects of Vocal Style

Recall that our musical-theatre singers appeared to be less well trained and less experienced than our operatic singers. After recording the stimuli, we therefore eliminated some of the more difficult melodies in which the music-theatre singers had stumbled. This resulted in stimuli employing the 30 target words shown in Table 11. In testing the effect of vocal style on intelligibility, like was matched with like. That is, the music-theatre vocalists sang the identical melodies and words sung by the operatic vocalists.

Table 11. Target words recorded by both operatic and musical-theatre voices.

self	lord	jean	swim	walk
kids	song	new	twice	choose
learn	laugh	yes	why	fog
air	move	would	ain't	rhymes
sweet	hear	man	ill	used
as	are	have	get	boys

Each listener heard a unique random selection of the above words in sung and spoken contexts by both operatic and musical-theatre voices. The results are shown in Table 12.

Table 12. Number of correctly and incorrectly identified sung words.

	Correct	*Incorrect*	Total
Operatic Voice (n=2)	450	84	534
Theatre Voice (n=4)	321	55	376
Total	771	139	910

The sung intelligibility scores for individual singers ranged from 78% to 87%. No significant difference between operatic voices and musical-theatre voices was evident (χ^2 = 0.13; *df* = 1; *p* = .72; *Phi* = .01). The null hypothesis failed to be rejected: there appeared to be no significant difference in intelligibility for the operatic and theatre voices employed in our study.

DISCUSSION

Archaic Words (H1): The first hypothesis test results were consistent with the notion that familiar words are more intelligible than archaic words. When considering lexical frequency, it is necessary to make a distinction between word frequency in *spoken* versus *sung* contexts: archaic words like "thine" and "dale" are not commonly used in spoken contexts, but they do occur frequently in sung lyrics—especially in classical repertoire. This difference might have interfered with our prediction: listeners who are familiar with an archaic lexicon through exposure to classical song lyrics could have a personal lexicon that includes more archaic words. Additionally, we might very well question the notion that prototypes are based merely on statistical frequencies. A number of other factors could mark a particular item as a prototype: "others [factors] appear to be ideals made salient by factors such as physiology (good colors, good forms), social structure (president, teacher), culture (saints), goals (ideal foods to eat on a diet), formal structure (multiples of 10 in the decimal system), causal theories (sequences that 'look' random),

and individual experiences (the first learned or most recently encountered items or items made particularly salient because they are emotionally charged, vivid, concrete, meaningful, or interesting)" (Gabora, Rosch, & Aerts, 2008).

According to prototype theory, we might predict that sung contexts could activate new possibilities for expected words. For example, without any specific context, common examples of "Beverages" include coffee, tea, or soft drinks; but when respondents are primed to name beverages in the context of "Parties," wine might be a more common example (Gabora, Rosch, & Aerts, 2008). Analogously, the mere medium of singing might prime listeners to expect words involving certain themes such as love; or even to expect a lexicon of florid, archaic, and fanciful words.

If we consider more fundamental components of language perception, a phonemic "magnet effect" (Kuhl, Conboy, Padden, Rivera-Gaxiola, & Nelson, 2008) that draws phonemic variations toward particular phonemic prototypes could explain why the use of more common words might not unequivocally lead to facilitated recognition of lyrics. Phonemic prototypes allow listeners to more easily recognize word sounds (phonemes), but prototypes are distortions of perception (Kuhl et al., 2008)—they weaken listeners' ability to discriminate between non-native-language phonemic variations. Presumably, common and archaic English words are built up from the same foundation of phonemes. Some phonemes might be inherently more difficult to perceive, or susceptible to masking. And we could assume that the distributions of difficult/easy-to-perceive phonemes are roughly equivalent between common and archaic lexicons. Therefore, as a rival hypothesis one could posit that the capacity to *perceive* common and archaic words might be the same—confusion rates of lyrics might be the same. Or, if the experimental measurement might fail to locate the source of errors—judgment errors might arise from higher-order comparisons between the perceived lyric (e.g., "gloaming") and a listener's personal lexicon. (E.g., "I heard a word that sounded like 'roaming.'")

In the present study, we predicted that archaic words are more difficult to perceive than common words. Although we could challenge this hypothesis based on the distinction between lexical and phonemic prototypes, the rival hypothesis would predict that there is no statistically significant difference between the intelligibility of archaic and common words. This rival hypothesis is not distinct from the null assumption that there is no statistically significant difference between the intelligibility of vernacular and archaic words. (It is difficult to imagine a situation where archaic words would be more intelligible than common words; unless we were dealing with a very knowledgeable set of listeners who are *very* familiar with musical lyrics containing archaic words.)

Diphthongs (H2): We predicted that diphthongal words would be less intelligible than monophthongal words. The motivation for our prediction was the musical practice of delaying the final vowel glide until the end of the sung segment: such delays do not occur in common speech, so we might expect delayed glides to interfere with intelligibility. Contrary to our prediction, however, our experimental results suggested that diphthongal words are *more* intelligible than monophthongal words. Evidently, the presence of a vowel glide enhances intelligibility of sung words rather than interfering with intelligibility. A plausible reason for this is that sung vowels differ from spoken vowels, so sung pure vowels are more difficult to recognize than their spoken counterparts. Potentially, vowel glides provide additional phonetic information that allows listeners to better infer the intended phonemes.

Melismas (H3): As expected, melismatic settings degrade intelligibility compared with syllabic settings. That is, words are more difficult to recognize when a syllable is sustained over several notes. Contrary to expectation, we did not find that melismatic settings of bisyllabic words caused lower intelligibility compared with melismatic settings of monosyllabic words. Instead, our results appear to imply that melismatic monosyllable words are more difficult to recognize than melismatic bisyllabic words. At least two interpretations may be offered. First, our results are broadly consistent with the role of lexical distinctiveness in speech intelligibility: monosyllabic words have less lexical distinctiveness than multisyllabic words (Francis & Nusbaum, 1999), so one might expect that melismas would degrade the intelligibility of monosyllabic words more than bisyllabic words. Notice that this result is also consistent with our earlier finding (H2) showing that diphthongs are better identified than monophthongs. In general, phonological change seems to support word intelligibility. However, the results are equally consistent with a second interpretation. Post hoc examination of our stimuli showed that monosyllable stimuli employed an average of 5.0 notes per syllable, whereas bisyllabic stimuli employed an average of 3.8 notes per syllable. The difference in intelligibility measures may therefore simply reflect the length of the melisma: the greater the number of notes assigned to a syllable, the greater the likelihood of reduced intelligibility.

Stress (H4): The results of this hypothesis test were consistent with Burleson's (1992) prediction that aligned prosodic-musical stress enhances word intelligibility. Musical stress (or "accent") can consist of isolated or combined stress types such as metric accent, dynamic accents, melodic contour accents, and note-length accents. We did not manipulate these different stress types independently; we solely considered metrically stressed or unstressed target words without specifically including or excluding other types of stress. The effect size for musical meter stress alignment was modest—an occasional deviation from aligned stress might not obscure a word completely, and may even enhance the meaning of that word. However, these results suggest that the music would be well-served if factors contributing to intelligibility are balanced with factors such as text painting.

Repetition (H5): Musical elements, such as chords, were not a part of the experimental stimuli— carrier phrases consisted only of melodies set with lyrics. Aspects such as rhythm, melodic contour, melodic repetition could have influenced the results; however, these variables were not manipulated and these questions lie outside the scope of this series of hypothesis tests. The presented results for immediate repetition were consistent with the repetition priming research literature, but delayed repetition results showed no priming for musical lyrics. This runs contrary to the usually observed trend of repetition priming over the course of delays of minutes, hours, or even days. Perhaps other musical variables such as melody can potentially interfere with word recognition; future studies that control for various lengths of delay and melodic repetition could establish the role of repetition priming in musical contexts involving words.

Syllabic Priming (H6): The results contradict the experimental hypothesis—melismatic settings appear to prime the identification of words in subsequent syllabic settings. Here we might only speculate as to the cause of these paradoxical results. One possible interpretation might be as follows:

Suppose that, in the processing of words, if word recognition does not occur within some reasonably short period of time, the auditory system abandons the task and frees attentional resources for other purposes. At the same time, the first syllable in an unrecognized word may be retained in short-term memory. In encountering a melismatic word, therefore, the tendency would be to retain the first syllable and ignore the rest of the sound. If this were the case, then we would expect a small priming effect when a melismatic presentation precedes a syllabic presentation (as observed above). Since the auditory pathway is simply confused by melismas, one might expect to see no priming effect when a syllabic presentation precedes a melisma. However, it bears emphasizing that this account is post hoc speculation.

Rhyme (H7): In light of the frequent occurrence of rhymes in poetry and song lyrics, we expected that an immediately preceding rhyming word would facilitate word intelligibility. However, we were surprised that our data showed a significant reverse effect: the previous occurrence of a rhyming word produces a decrease in intelligibility. A possible explanation for this finding may be found in recent research on phonological similarity.

If we consider the precise nature of this experiment's task, then we see that participants 1) attended a phrase-ending word; 2) stored this "target" word's phonemes as it gradually unfolded in time (sometimes over extended melismas); and 3) attempted to identify the target word. Thus, this task did not rely upon memory alone or processing alone: the experimental design necessitated participants' use of both memory *and* processing skills—*complex span* is one measure of this coupling, in contrast with *span*, which is a measure of memory capacity alone (Daneman & Carpenter, 1980). Participants responded after each presentation of a phrase, so this procedure does not mimic typical psychological tests of span where multiple words are presented before each of the participant's responses. However, the present study's musical delivery of words stretched out their phonemes in time and likely invoked working memory processes rather than short-term auditory storage in isolation. Accordingly, there is an abundance of research, specifically devoted to working memory and language, which bears on the rhyming hypothesis.

Several recent studies have found that phonological similarity degrades performance on both span tasks and complex span tasks; this effect has been replicated often and is known as the "phonological similarity decrement" (Copeland & Radvansky, 2001). In word span tasks that present lists of isolated words for memorization, participants recall phonologically *distinct* words more successfully than phonologically similar words. By contrast, reading span tasks seem not to be affected by this decrement: tasks that use sets of sentences result in enhanced recall for phonologically *similar* words that end each sentence (Copeland & Radvansky, 2001). This reading span task seems much more relevant to rhyming genres like poetry.

A follow-up study (Lobley, Baddeley, & Gathercole, 2005) noted that Copeland and Radvansky (2001) used only rhyming words as representatives of phonologically similar words; rhyme may be an exceptional case of phonological similarity. Lobley, Baddeley, and Gathercole (2005) used words

containing phonological similarities only in their middle vowel location; additionally, the study examined a variety of processing tasks (e.g., veracity judgments) to engage memory and processing in tandem (complex span). They observed the phonological similarity decrement in all three experiments, even when the task changes—different response modalities affect only the degree of the phonological similarity decrement.

Merging these previous studies' results, it appears that rhyme is indeed a special case of phonological similarity: rhyme enhances recall in contextual settings, but decreases recall in isolated word conditions. Baddeley *et al.* (2002) observed that some participants in their pilot study used an initial letter rehearsal strategy to aid in recall. Perhaps contexts that use rhyming words coupled with an initial letter rehearsal strategy could circumvent the decrement effect because the high degree of phonological similarity between rhyming words permits listeners to remember entire words, while actively maintaining only a minimal amount of information, i.e., the first phoneme of each rhyming word (Lobley, Baddeley, & Gathercole, 2005). However, it seems that an intervening context (as in Copeland & Radvansky, 2001) is needed in order for rhyme to actually facilitate recall. If this special condition is required for rhyme-facilitated recall, then the results of the present study may have been due to the use of non-contextual carrier sentences. Recall that the present method did have a context, "I am singing the word _____," but each trial had an *identical* linguistic context and this could be a confounding factor. In short, our negative result for the effectiveness of rhymes is consistent with previous observations of the phonological similarity decrement effect.

Another aspect of the stimuli's mode of presentation that could have influenced the result was word repetition. Participants heard frequent word repetitions throughout the experiment; this was partly due to the word randomization process and also due to hypothesis 5 concerning word repetition. Words that repeated included ones in hypothesis seven and also words used in other hypotheses. In this "repeated word" context, it could have been more difficult to identify rhyming words because of a presumed difficulty in distinguishing a repeated word from a rhyming word. This consequence of the experimental setting is not entirely artificial: lyrics do repeat frequently in songs. Repetition likely occurs more often than occurrences of rhyming words, however this intuition awaits empirical investigation.

Vocal Style (H8): In the present study, we found no difference in intelligibility between operatic and musical-theatre voices. This negative result could simply have been a consequence of the small number of voices used (n=6). Research on dialects suggests that more detailed studies of vocal style may be warranted. For example, a study by Ambrazevičius and Leskauskaitė highlighted singing differences in two Lithuanian dialects. The vowels in northern Lithuanian dialect, *Aukštaičiai*, are less distinct from each other in speaking and also when sung in that region's style; in comparison, a southern dialect, *Dzūkai*, has greater differences between vowels in both speaking and singing (Ambrazevičius & Leskauskaitė, 2008). In short, the speech patterns of the two Lithuanian dialects lead to distinct styles of singing. It is possible that different stylistic norms in vocal diction might well account for important differences in word intelligibility. In general, cross-stylistic investigations of vocal characteristics have the potential to refine the generalization, "sung words are less intelligible than spoken words"—clarifying the role of anatomical, cultural, and geographic influences upon sung and spoken utterance.

CONCLUSION

Half of the experimental hypotheses set forth in this study did not survive our empirical test; nevertheless, 7 of the 8 hypotheses yielded statistically significant results. Three tests led to significant results that were directly opposite to our predictions. The results are summarized below (✓=stood; ⬚=failed; *=significant reverse trend):

✓   H1: High frequency words are more intelligible than low frequency words.
⬚*   H2: Words containing diphthongs are less intelligible than words without diphthongs.
✓   H3: Melismas reduce intelligibility compared with syllabic settings.
✓   H4: Intelligibility is improved when syllable stresses are aligned with musical stresses.
✓   H5: Intelligibility is improved when the same word appears in the previous melodic phrase.
⬚*   H6: A word's melismatic setting is more intelligible when preceded by its syllabic setting.
⬚*   H7: Intelligibility is improved when successive words rhyme.
⬚   H8: Musical-theatre voices are more intelligible than operatic voices.

With hypothesis 1, we examined the intelligibility differences between archaic words and commonly used words; the results were consistent with the claim that archaic words are less intelligible. This difference was strongest when words were sung rather than spoken.

Contrary to our prediction, our test of hypothesis 2 was consistent with the claim that diphthongs are more easily recognized than monophthongs. Looking at the change that occurred within a single vowel type, we saw that diphthongs may tend to be more easily identified when sung; monophthongs may be more easily identified when spoken—the verity of this ostensible interaction needs to be tested with a larger sample of spoken words.

With regard to hypotheses 3, 4, and 5 the results were all consistent with the research predictions. Words containing syllables that span two or more notes (melismas) were less intelligible than syllabic text settings; however, melismas did not appreciably degrade the intelligibility of bisyllabic words—monosyllabic words were much more affected by melismas. In addition, word intelligibility was facilitated when the stressed syllables coincided with stressed metric moments in musical phrases. Words that appeared in an immediately preceding phrase were also more likely to be correctly recognized. However, when a word appeared earlier than the immediately preceding phrase, no improvement in recognition was observed.

Two further hypotheses (6 and 7) exhibited results contrary to those predicted. Our results implied that recognition of a melismatic word is not facilitated by a preceding syllabic presentation of the same word. Instead the results were consistent with the claim that recognition of a syllabic word is facilitated by a preceding melismatic presentation of the same word. A possible post hoc rationale for this finding was discussed in the results section; however, we have no credible account for this seemingly paradoxical result.

Finally, regarding voice type, we observed no difference in intelligibility between operatic and musical-theatre voices. As noted earlier, however, our experiment employed just six singers.

It bears reminding readers that this study was largely exploratory: the stimuli used in this experiment were designed to serve multiple purposes and so data independence was compromised. Ideally, each hypothesis should be tested using its own independent stimulus and response set—although this approach would require several thousand independent stimuli.

NOTES

[1] Correspondence regarding this study may be sent to Prof. David Huron, School of Music, 1866 College Road, Ohio State University, Columbus, Ohio, U.S.A., 43210.

REFERENCES

Ambrazevičius, R., & Leskauskaitė, A. (2008). The effect of the spoken dialect on the singing dialect: The example of Lithuanian. *Journal of Interdisciplinary Music Studies*, *2*(1&2), 1-17.

Baddeley, A., Chincotta, D., Stafford, L., & Turk, D. (2002). Is the word length effect in STM entirely attributable to output delay? Evidence from serial recognition. *The Quarterly Journal of Experimental Psychology*, *55A*(2), 353-369.

Balota, D. A., & Chumbley, J. I. (1985). The locus of word frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory & Language*, *24*, 89-106.

Bigand, E., Tillman, B., Poulin-Charronnat, B., & Manderlier, D. (2005). Repetition priming: Is music special? *Quarterly Journal of Experimental Psychology*, *58A*(8), 1347-1375.

Burleson, R. (1992). Functional-relationships of language and music—the 2-profile view of text disposition. *Linguistique*, *28*(2), 49-63.

Collister, L., & Huron, D. (2008). Comparison of word intelligibility in spoken and sung phrases. *Empirical Musicology Review*, *3*(3), 109-125.

Copeland, D., & Radvansky, G. (2001). Phonological similarity in working memory. *Memory & Cognition*, *29*(5), 774-776.

Coren, S., & Hakstian, A. (1992). The development and cross-validation of a self-report inventory to asses pure-tone threshold hearing sensitivity. *Journal of Speech & Hearing Research*, *35*(4), 921-928.

Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450-466.

Francis, A., & Nusbaum, H. (1999). The effect of lexical complexity on intelligibility. *International Journal of Speech Technology*, *3*(1), 15-25.

Hollien, H., Mendes-Schwartz, A., & Nielsen, K. (2000). Perceptual confusions of high-pitched sung vowels. *Journal of Voice*, *14*(2), 287-298.

Huron, D., & Royal, M. (1996). What is melodic accent? Converging evidence from musical practice. *Music Perception*, *13*(4), 489-516.

Kuhl, P., Conboy, B., Padden, D., Rivera-Gaxiola, M., Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B, 363*, 979-1000.

Lobley, K., Baddeley, A., & Gathercole, S. (2005). Phonological similarity effects in verbal complex span. *Quarterly Journal of Experimental Psychology*, *58A*(8), 1462-1478.

Smith, L., & Scott, B. (1980). Increasing the intelligibility of sung vowels. *Journal of the Acoustical Society of America*, *67*(5), 1795-1797.