Comparison of Word Intelligibility in Spoken and Sung Phrases

LAUREN B. COLLISTER
School of Music, Department of Linguistics, Ohio State University

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

ABSTRACT: Twenty listeners were exposed to spoken and sung passages in English produced by three trained vocalists. Passages included representative words extracted from a large database of vocal lyrics, including both popular and classical repertoires. Target words were set within spoken or sung carrier phrases. Sung carrier phrases were selected from classical vocal melodies. Roughly a quarter of all words sung by an unaccompanied soloist were misheard. Sung passages showed a seven-fold decrease in intelligibility compared with their spoken counterparts. The perceptual mistakes occurring with vowels replicate previous studies showing the centralization of vowels. Significant confusions are also evident for consonants, especially voiced stops and nasals.

Submitted 2008 July 15; accepted 2008 August 7.

KEYWORDS: *intelligibility, singing, lyrics, voice*

MOST popular forms of music involve the human voice. In nearly all cultures, singing is one of the preeminent forms of music making. Although singing may involve nonlanguage vocables (such as "fa-la-la"), the vast majority of vocal music takes advantage of the opportunity to employ a narrative, lyrical or poetic text. Despite the common use of text in vocal music, concertgoers and music listeners frequently complain of the difficulty in comprehending the lyrics of the music. This raises two questions: How intelligible are sung lyrics? And what are the causes of the loss of intelligibility?

Existing research has already shown that listeners have significant difficulty in discriminating different sung vowels. This is especially apparent for tones with relatively high fundamental frequencies as might be sung by a soprano. Smith and Scott (1980), for example, studied the intelligibility of vowels produced by a trained soprano in operatic conditions. Ten listeners were asked to discriminate four similar English vowels produced at different pitch levels. Smith and Scott found that the intelligibility of isolated vowels for pitches above F5 was reduced by 50% compared with the same vowels sung at C#5. That is, they demonstrated a dramatic reduction in intelligibility for high sung vowels. When sung with a raised larynx (as might be done in popular music styles) the intelligibility between C#5 and F5 dropped only 10 percent, but then dropped more dramatically as the pitch height increased.

Benolken and Swanson (1990) carried out a similar experiment with a trained operatic soprano student. The soprano produced twelve different vowels (both sung and spoken). Twenty-eight phonetically untrained listeners were asked to judge the isolated vowels by comparing them to target words. The results of Benolken and Swanson replicated the earlier work of Smith and Scott: American English sung vowels become increasingly difficult to discriminate as the fundamental frequency is increased.

Hollien, Mendes-Schwartz and Nielsen (2000) also carried out an intelligibility study of sung vowels. They employed eighteen professionally trained male and female singers. Each singer recorded three isolated vowels at two pitch levels and two loudness levels. Listeners included voice teachers, phoneticians, speech pathology students and untrained undergraduate students. In total, some fifty listeners were asked to identify the vowel and also identify the sex of the singer. Hollien et al. found that few vowels are correctly identified when the fundamental frequency reaches or exceeds the typical first formant. In general, incorrectly identified vowels tend to be confused with central vowels.

Apart from the difficulties involved in discriminating vowels, other aspects of phonology might be expected to contribute to problems in intelligibility. Burleson (1992) speculated that rhythmic aspects of prosody, such as word stress, might also be disrupted by musical settings. However, Burleson did not produce an empirical demonstration of such disruptions.

The work of Smith and Scott (1980), Benolken and Swanson (1990), and Hollien, et al. (2000) has admirably demonstrated the problem of fundamental frequency on the intelligibility of vowels. While vowel discrimination is a very important aspect of language perception, there are many other elements that contribute to language intelligibility. For example, to our knowledge, no research has examined potential differences between sung and spoken consonants, prosodic stress, syllabic and melismatic settings, or tempo. When singing in a real musical context, how difficult is it for listeners to comprehend the words?

In this study we test directly the intelligibility of individual words in a musical context rather than isolated phonemic components. In contrast to previous experimental studies, our study will collect intelligibility data in a more ecologically valid musical context. To anticipate our results, we will show that, even when singers avoid high pitches, substantial intelligibility problems are evident.

Formally, our hypothesis is that lexical items sung in a musical context are significantly less intelligible than their spoken counterparts. We further hypothesize that stopped consonants (like /b/, /t/, /p/, /k/) are more difficult to recognize than liquids or nasals (like /r/, /l/, /m/, /n/) because stopped consonants require a stoppage of the air flow in the mouth and may be difficult for listeners to distinguish because they do not have a continuous vocal quality like liquids and nasals.

METHOD

In brief, the experiment presented listeners with special-purpose recordings of existing vocal melodies whose lyrics were replaced with a carrier phrase and a target English word. An equivalent number of stimuli in spoken form were also presented. Listeners were then asked to transcribe the target word.

Subjects

Twenty-two subjects were recruited for the experiment, thirteen males, eight females, and one subject who declined to answer. The participants were drawn from a convenience population of sophomore university music students participating in an experimental subject pool. Students of voice were explicitly excluded from participation since singers may possess some special knowledge or experience that would affect their ability to understand sung lyrics. 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 at "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 disposed to attend to lyrics than the general population.

As a screen for possible hearing loss, we used the Coren & Hakstian (1992) survey in lieu of an audiometric examination. This survey poses a series of simple questions (e.g. "Can you hear the telephone ring when you are in the room in which it is located?") whose answers have been show to correlate with audiometric data. Before conducting the experiment, an *a priori* hearing score of below 27, deemed "normal hearing", was established so that participants who scored higher than this value were excluded from the experiment. Using this exclusion criterion, two of twenty-two potential subjects were eliminated.

Stimuli

Three advanced level vocal students were recruited to generate the experimental stimuli. This includes two females and one male representing soprano, alto, and tenor vocal ranges. Each vocalist rehearsed a set of melodic phrases chosen from Barlow and Morgenstern's *Dictionary of Opera and Song Themes* (1976). The melodic phrases consisted of one thematic phrase chosen from twenty vocal works by 18 composers, including Adam, Bach, Beethoven, Bizet, Brahe (sic), Debussy, Donizetti, Fauré, Haydn, Handel, Humperdinck, Mendelssohn, Mozart, Poulenc, Puccini, Rossini, Schumann, and Smetana. Phrases ranged in length from seven to twenty-eight notes with an average of 9.9 notes. Appendix I identifies the specific

melodic phrases used in the study. In addition, the Appendix identifies the number of notes in the phrase, as well as the number of syllables and notes assigned to the target word.

In addition to the sung phrases, each target word was also recorded in a spoken context. For the sung and spoken stimuli, the following carrier phrases preceded the target word:

"I am singing the word _____."

"I am saying the word: _____."

One hundred and twenty target words were selected according to criteria described below. The number of syllables in target words ranged from 1 to 4. Each of the three singers recorded two renditions of each of the 20 melodic phrases – each rendition contained a different target word. In addition to the sung phrases, singers also recorded each word once in a spoken context. Figure 1 illustrates a sample sung stimulus with the corresponding textual underlay. As a result, the experimental stimuli consisted of 240 recorded phrases, half sung, half spoken.

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 permanently installed stereo overhead microphones placed approximately 10 meters from the vocalist. Both the position of the vocalists and the microphone placement conform to normal performance practice for this auditorium. No effort was made to measure the reverberation time for the hall. However, the recordings made in these circumstances were judged by the authors to be similar in reverberation and ambience to standard recital recordings. Notice that the stereo microphones were closer to the singer than would be the case for 95% of the audience members for a hall of this size.

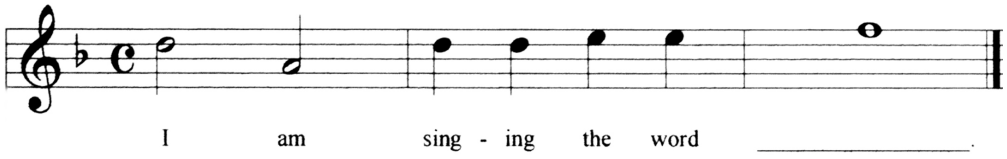

Fig. 1. A sample sung stimulus. The melody is a fragment from Haydn's "Lob der Faulheit" ("In Praise of Idleness") in which the original text, "Faulheit, endlich muss ich dir" has been replaced with our carrier phrase. The blank line indicates where the target word was sung. This carrier phrase was used to eliminate contextual effects on the intelligibility of the target words.

In recording these stimuli, singers received the following instructions:

1. In singing these phrases, sing them as you might normally sing. Do not attempt to enunciate the words more clearly or less clearly than you would in your ordinary singing.
2. In speaking these phrases, speak them as you might normally speak on stage as part of a libretto. Do not attempt to enunciate the words more clearly or less clearly than you would in normal theatrical declamation.

Target Words

Since context is difficult to control experimentally, we decided to avoid contextual information in our study. That is, all lexical items were presented within a uniform carrier phrase. This means that our measures of intelligibility will underestimate the actual intelligibility in normal listening circumstances where auditors may be expected to take advantage of contextual information, such as possible knowledge of a song's title, awareness of the theatrical or ritual context, repetition of lyrical content, or the ability to anticipate words based on previous words.

It is common practice in speech intelligibility measurements to make use of phonetically balanced (PB) word lists. These are lists of monosyllabic words where the frequency of phonemes reflects their relative frequency in English speech. PB word lists are useful for measuring the intelligibility of phonemes, but there are other factors that contribute to speech intelligibility aside from phonetic recognition. One factor is pragmatic context: a listener's awareness of the topic can increase intelligibility by anticipating possible lexical items. Another factor is the presence of multi-syllabic words, whose recognition is facilitated by additional lexical distinctiveness (Francis & Nusbau, 1999). That is, a multi-syllabic word

like "answer" has greater information than a monosyllabic word like "said" and so, on average, may be easier to recognize in isolation. However, in the case of music, multi-syllabic words raise special concerns. In music, a distinction is made between *syllabic* and *melismatic* text setting. In some music (such as Gregorian chant), it is common for a single syllable to be sustained through many pitches. In these types of melismatic situations, the large temporal span used for setting a multi-syllabic word may actually reduce the intelligibility. For these reasons, we decided to include multi-syllabic words in our list of target words. Since PB word lists do not include multi-syllabic words, we decided to create a word list from other sources.

In assembling our word list we were concerned that the words used in vocal lyrics may exhibit systematic differences from ordinary sampled speech. Accordingly, we decided to assemble our own database of vocal lyrics, from which target words would be identified. Our database of vocal lyrics was created from a combination of popular and classical English-language texts. The sample of English-language popular music was based on two sources: one hundred #1 hits from the period 1960-2000 as identified by *Billboard Magazine*, and the Recording Industry Association of America (RIAA) and the National Endowment for the Arts (NEA) "Songs of the Century" list. Combining these two sources resulted in a popular music sample containing 450 English-language popular songs from multiple genres spanning the past century, with primary emphasis on music from the past 40 years.

In addition to the popular music sample, a database of classical music lyrics was assembled based on the *The Lied and Art-Song Text Page* website (www.recmusic.com/leider). From this site, twenty English-language composers were identified. While English language composers represent a minority in the classical vocal music tradition, music by non-English composers are more likely to be sung in the original language, and English translations of these works may raise other unforeseen problems. Only those composers were included for whom electronic corpora of the lyrics were easily available. Appendix II lists the twenty composers sampled, as well as the number of songs included in the sample from each composer.

The resulting database of classical English-language vocal text includes 1,553 poetic lines, containing 6,623 words. The popular music sample includes 17,130 poetic lines of text, containing 90,459 words. The two samples were combined into a larger aggregate database, and this resulting database of English-language vocal texts holds 18,683 poetic lines, containing 97,082 words. In this aggregate database, words from popular songs outnumbered words from classical lyrics in a ratio of roughly 14:1.

In music, lyrics are often repeated; this is especially common in chorus passages. In order to avoid undue influence of repetition in our sample of lyrics words, duplicate text lines were discarded. This reduced the sample from 18,683 lines to 11,554 lines containing 77,744 words, of which 7,059 were unique words. Using this reduced sample, we created an inventory of all words used, and tallied the total number of appearances for each word. Table 1 shows the twenty most frequently used words in our sample. With the exception of the word "love," the remaining words are non-content words. The word "love" is of particular notice. While this word is common in vocal lyrics, it is not a common word in samples of ordinary English speech, and so highlights the value of creating a task-specific corpus of words.

Table. 1. The twenty most frequently occurring words in English lyrics, as found in our database of English song texts.

Word	Frequency
the	3415
you	2296
I	2236
and	2160
to	1778
a	1733
in	1042
my	1000
me	1018
of	895
it	780

that	766
on	739
your	665
be	558
for	557
all	534
is	514
love	482
but	447

Common words are typically one syllable in length, but our database also included many multi-syllabic words. Table 2 identifies the proportions of different syllable lengths used in the sample, and gives some examples.

Table. 2. Syllable length distributions in the aggregate song lyrics database, compiled from lyrics of popular music and classical English song texts with exemplars of content words. "Words" containing seven or more syllables mostly consist of nonsense vocables or multisyllabic words that have been lengthened by repeating one or more syllables.

No. of syllables	No. of words	No. of unique words	Exemplars
1	63,253	2,652	yeah
2	12,103	3,176	stompin'
3	1,853	919	together
4	423	232	everybody
5	61	47	unforgettable
6	17	13	originality
7	24	12	
8	4	4	
9	2	2	
10	0	0	
11	3	1	
22	1	1	
Total:	77,744	7,059	

Target words were randomly selected from our database. In order to ensure that the number of syllables in the target words are proportionately representative of the total sample, we randomly selected a prior number of words of a given syllable length. Specifically, we selected 97 one-syllable words, 18 two-syllable words, 3 three-syllable words, and 2 four-syllable words. The 120 target words used in this study are identified in Appendix III.

Procedure

Subjects were tested individually in an Industrial Acoustics Corporation sound isolation room. Stimuli were presented stereophonically over AKG-K240 Monitor headphones. Each participant heard 120 stimuli containing the 120 target words, with 60 sung and 60 spoken, presented in alternating order. In order to avoid possible priming effects, each participant heard each target word just once. The sung and spoken versions for each target word were presented to an equal number of participants. Following each

presentation, subjects were asked to transcribe only the target word (not the carrier phrase) using a standard American English computer keyboard. The experiment lasted approximately twenty minutes.

RESULTS

In total, 2,539 stimuli – both spoken and sung – were heard by subjects. This breaks down into 1,271 spoken stimuli and 1,268 sung stimuli. The unequal numbers of spoken and sung stimuli are due to technical difficulties, which forced the early conclusion of three sessions. Subjects made 380 target word identification errors, which is 14.9% of the total presented stimuli. An error was defined as an instance in which the phonetic composition of a response was not consistent with the phonetic composition of the word presented. That is, homophones (such as right/write) were not counted as errors. The discrepancies were judged by the experimenters using the dialect of American English spoken in central Ohio.

Of the 380 errors, 334 occurred in conjunction with the sung stimuli while just 46 happened with spoken stimuli. These errors constitute 26.3% of the singing stimuli and just 3.6% of the spoken stimuli. The difference between spoken and sung word intelligibility amounts to a 76.4% decrease in intelligibility for the sung versions. There are 7.3 times as many listening errors with sung words as with spoken words. Figure 2 shows the error rates in graphic format.

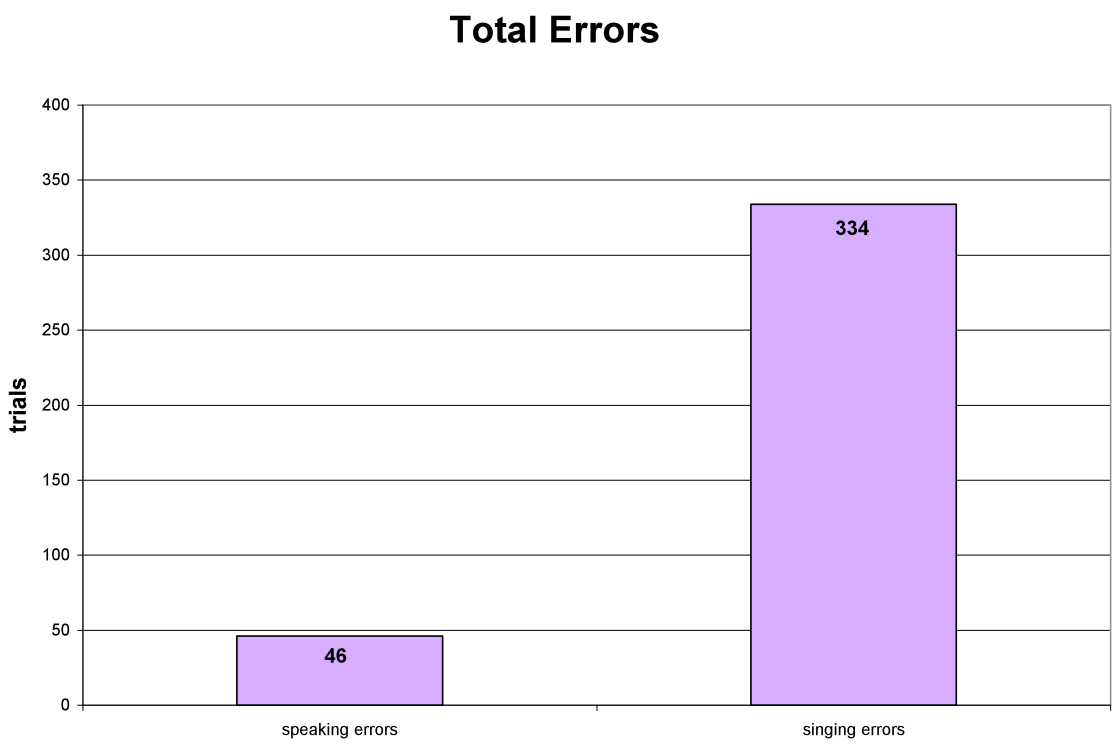

Fig. 2. Total errors – phonetic differences in responses versus presented stimuli – for both spoken and sung contexts. Of a total of 380 errors, 334 occurred for the sung stimuli; this constitutes a 76.4% increase in intelligibility loss.

In post-experiment interviews, subjects were asked to rate how accurate they thought they were on the spoken trials versus the singing trials. Subjects guessed, on average, that they were correct on 93.6% of the spoken trials and 75.1% of the singing trials – both of these are very close to the actual 96.4% speaking and 73.7% singing accuracy rates.

Without resorting to a statistical test, it is obvious that the results of this experiment are consistent with the formal hypothesis (and with common listener intuitions) that there is a considerable loss of intelligibility in sung material as compared to spoken versions.

POST-HOC ANALYSES

Given the nature of our experiment, the data lends itself to a linguistic analysis of phonetic errors. What specific sounds and word contexts are most associated with the loss of intelligibility?

Types of Errors

Responses were phonetically transcribed based a dialect of English spoken in central Ohio. For all erroneous responses, a phonetic comparison was made between the original words and the given incorrect responses. This analysis resulted in approximately 377 individual phoneme errors: 294 consonant errors and 83 vowel errors. There may be more or less depending on how one analyzes the given data set; also, due to possible dialect differences between the experimenters and the subjects, some of the phonetic analysis may not accurately reflect the phonetic values intended by the subjects. Appendix IV shows the complete phonological analysis of all errors.

1. VOWEL ERRORS

Past research has shown that, in isolation, sung vowels are difficult for listeners to perceive. Our data replicates this finding within word context. In particular, the previously observed 'centralization' phenomenon is evident, as demonstrated by Hollien et al. (2000). Centralization occurs when sounds produced using more remote tongue positions are perceived as generated using less remote tongue positions -- for example, when /i/ is perceived as /I/. Of the 83 vowel errors, 27 were instances of centralization. The vowel error that was made most often, that being the target word 'steel' heard as 'still', is a prime exemplar of this phenomenon, although there were other cases of such vowel mistakes. Figure 3 is a vowel chart showing some of the most common centralizations encountered in this study. Centralization errors are the most common phonetic errors exhibited in our data; the next most common error is diphthongization, in which a monophthongal vowel is transformed into a diphthong. This type of error occurred 13 times out of 83 errors. Diphthongization may be due to a shifting in the position of the mouth in the middle of a sung vowel in order to prepare for an articulatory attack of the next consonant. In the cases of melismatic diphthongization, of which we had three instances, the singers may be changing the shape of the mouth in order to prepare for the next pitch – this preparation may affect not only the tone, but also the quality of the sung vowel.

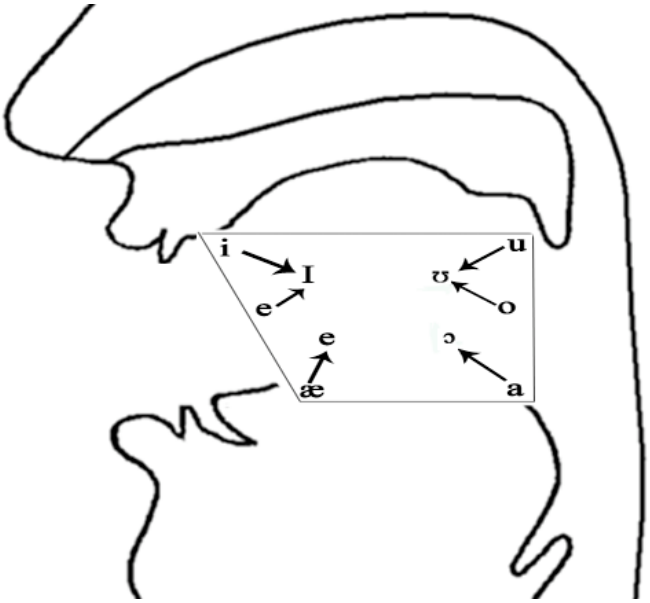

Fig. 3. A vowel chart showing the most common vowel transformations. The figure shows the human mouth (with the lips and teeth at the left and the uvula and throat at the right), and the quadrilateral chart in the center with the vowel symbols showing the locations where the vowels are generally articulated. The phonetic symbols represent English vowels and the arrows indicate the directions of the mistakes that were made most often. For instance, the vowel /u/ was often heard as /¡/, and the vowels /i/ and /e/ were often heard as /Ò/. This chart shows graphically the replication of past studies of sung vowels, which showed that vowels tend to be heard more centrally when sung.

2. CONSONANT ERRORS

In the beginning, the hypothesis was stated that stopped consonants (like /b/, /t/, /p/, /k/) are more difficult to recognize than liquids or nasals (like /r/, /l/, /m/, /n/) in sung stimuli. We anticipated this effect because the continuous vocal quality of the liquids and nasals are generally presumed to be easier to hear in spoken language. In contrast, stopped consonants involve a stoppage of air flow in the mouth and so may be more difficult to hear. It turns out that the data prove to be consistent with this hypothesis; listeners in our study had a larger error rate for stop consonants than for liquids. However, contrary to our hypothesis, nasal consonants pattern with the stops, having twice the error rate of liquids. Furthermore, listeners had much more difficulty with voiced consonants than voiceless ones, which, in the context of singing, might seem counter-intuitive.

In spoken stimuli, the reverse is true. Listeners had more trouble with voiceless stops than any other class of sounds. Even though the total number of errors is much lower for the spoken stimuli, the patterning of errors is much different. There appears to be no correlation between voiced and voiceless consonants with listener confusion – rather, the manner of articulation affects the hearers' perception. Tables 3 and 4 show the percentages of errors for different classes of consonants in first the sung and then the spoken data, and Table 5 shows some of the common confusions in the sung trials.

Table 3. Error frequency for classes of consonants in sung stimuli.

Sung Stimuli			
Sound class	***# of errors***	***# total occurrences***	***error %***
Liquids	47	830	5.70%
Nasals	52	520	10.00%
Fricatives	65	770	8.40%
Voiced Stops	36	330	10.90%
Voiceless Stops	50	490	10.20%
Voiced	152	1840	8.30%
Voiceless	50	1100	4.50%

Table 4. Error frequency for classes of consonants in spoken stimuli.

Spoken Stimuli			
Sound class	***# of errors***	***# total occurrences***	***error %***
Liquids	7	830	0.80%
Nasals	5	520	1.00%
Fricatives	10	770	1.30%

Voiced Stops	1	330	0.30%
Voiceless Stops	10	490	2.00%
Voiced	19	1840	1.00%
Voiceless	14	1100	1.30%

Table 5. Most common consonant confusion errors with examples, indicating the confusion of single phonetic features.

Original sound	Perceived sound	Examples	Occurrences
/l/	/ã/	light-right	13
/ts/	/t/	that's-that	12
/n/	/d/	gone-god	7
/t/	/d/	set-said	6
/z/	/s/	choose-truce	5
/s/	/t/	trance-trant	5

In addition to these confusion errors, there were also insertion and deletion errors. One of the most frequent consonant errors was insertion of consonants in places where consonants were not present in the original words. This occurred in 54 of the 294 consonant errors. Most commonly, this happened at the beginnings and ends of words (almost equal in frequency – 24 instances at the beginning of words and 21 instances at the end of words). By far, the most common insertion was /h/ at the beginning of words that begin with vowel sounds – this happened 11 times of the 24 word-initial insertions. Other frequent word-initial insertions are /b/ and /p/, which occurred four times each; these are both bilabial sounds which might indicate that listeners are hearing the singers physically opening their mouths.

Word-final insertions were most commonly nasal sounds (/m/ and /n/ occurred 6 times each), though fricatives like /f/, /v/, and /È/ were also heard by listeners. The presence of these sounds at the ends of words, particularly sounds made so close to the front of the mouth, indicates that the listeners might be hearing the speakers physically close their mouths before the end of the sung tone. /d/ was heard at the end of words ending in the English /ã/ a total of four times (all of which were the target word 'were' being mistakenly heard as 'word').

Consonant deletions proved to be an interesting phenomenon. For an error to be considered "consonant deletion", there had to be a consonant in the original target word that was entirely removed in the subject's response; that is, the consonant was not transformed into another consonant or merged into a consonant cluster, nor was it displaced in the word. In total, in our data there are approximately 42 deletions of consonants – 36 of which are voiced consonants. Of these voiced consonants, 17 deletions were liquids or nasals, while 10 were fricatives and 9 were stops. This is more evidence against our hypothesis that liquids and nasals would be easier for listeners to hear – in 40% of deletion cases, liquids and nasals are not heard at all.

The fact that voiced consonants provide the listeners with the most difficulty may be counter-intuitive. It is well-known that voiced consonants are easier to hear because they include a voiced aspect, while voiceless consonants include only air manipulations; yet, 85% of deletions are voiced consonants, and 75% of consonant confusions happen with originally voiced consonants. However, these confusions may be a consequence of vocal training – students of voice are often instructed to hyperarticulate voiceless consonants in order to facilitate audience perception. The fact that only six deletions were voiceless out of 42 total deletions shows that the listeners are, in fact, hearing these voiceless consonants. Conversely, those consonants that are voiced – particularly liquids and nasals – are often being dropped entirely out of words, especially when they occur juxtaposed with a vowel. Perhaps these voiced consonants are being heard as part of their neighboring vowels, since the voiced aspect may be carried through the consonant itself and the final articulation delayed until the very end of the sound (typically coinciding with a musical beat). This delay of articulation may be what is causing the confusion regarding the nature of the articulations – since they do not happen when the listener anticipates them from experience with spoken language, the sense of

place and manner of articulation are thrown off and may cause confusion about the nature of the consonant. In addition, further confusion may arise because the mouth of the singer is formed in a manner that would best convey the nature of the sung vowel rather than the consonant.

3. ERROR LOCATIONS

Word-initial errors may be cases of hyperarticulation by the singer (as described above), and/or the influence of the melodic beat on the perception of the initial attack. Word-final errors, especially the insertion of voiceless stop consonants, may be a function of the stoppage of air at the end of a sung tone. Consonant clusters involve strings of different articulations, and the mouth shape may affect the shape of the vowel that follows them. This distortion centralizes vowels and deforms diphthongs as the singer attempts to form the proper vowel shape after the consonants. One example of this cluster distortion is in the stimulus 'straight', /stãeÒt/ in the authors' dialect, which was often turned into 'right' (/ãaÒt/). The influence of the English /ã/ on the vowel is evident; as the singer produces the /ã/ sound, the tongue tenses, which alters the sound and forces the diphthong lower in the mouth, hence the initial /a/ in the diphthong instead of the /e/.

4. MELISMATIC-CONTEXT ERRORS

In total, twenty-four target words were presented within the context of four melismatic melodies. Eighteen of these words were one syllable, while the remaining six were two syllables. The four melodies are shown below in Figures 4 through 7, along with the target words for each melody and the error frequency for each individual word (number of errors/total trials). The total trials for melismatic melodies numbered 223.

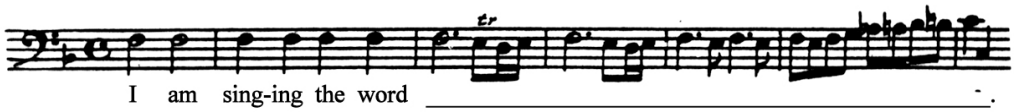

Fig. 4. Melismatic melody #14, target words: *one* (0 errors/7 trials), *if* (1/9), *might* (3/8), *free* (0/9), *named* (4/9), *done* (2/9).

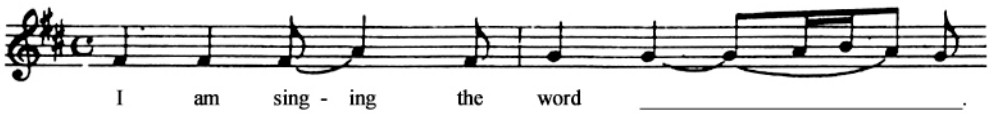

Fig. 5. Melismatic melody #15, target words: *time* (0 errors/12 trials), *this* (0/13), *round* (3/8), *dance* (1/8), *shall* (0/12), *kind* (3/11).

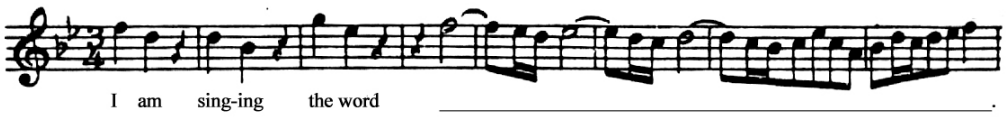

Fig. 6. Melismatic melody #16, target words: *oh* (6 errors/14 trials), *what's* (7/8), *light* (7/7), *tough* (4/10), *shine* (3/6), *knock* (6/9).

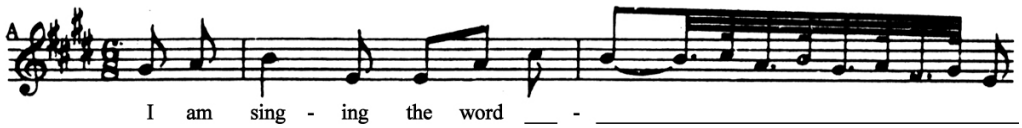

Fig. 7. Melismatic melody #19, target words: *instant* (2 errors/7 trials), *received* (0/7), *behold* (2/7), *army* (7/9), *Diane's* (8/11), *riches* (1/5).

In all, listeners made 70 mistakes on melismatic melodies out of 223 trials, which represents a 31.4% error rate. This is roughly twice the overall intelligibility loss exhibited in all stimuli. In post-experiment interviews, subjects expressed concern regarding their performance on the melismatic melodies, and indicated that these trials were the ones that they thought were most difficult.

Evidently, listeners had the most difficulty with melody #16, which is similar, in terms of length, to melody #14, the melody that listeners found to be the easiest of the four. One difference is that melody #16 is the only one of the melodies that has a rest before the target word. Listeners had great trouble with the initial sounds of the target words associated with this melody, which may be an effect of having the rest before the first attack. Because of the rest before the initial consonant in the word 'light', listeners may have been thrown off and, perhaps, the singer also overenunciated the /l/ consonant. It is not just initial consonants that have this problem, but initial vowels as well. With the example 'oh' in melody #16, the most common error was the insertion of an /h/ at the beginning of the word. In the discussions above, we learned that singers distort the qualities of consonants by hyperarticulating them – melody #16 shows us that, following a rest, singers may hyperarticulate not only initial consonants, but initial vowels as well, causing the insertion of a consonant at the beginning of the word. Alternatively, the beginning of a sung pitch may indicate an articulation to the listener's ear, hence the insertion of the breathy fricative /h/ to begin the sung word.

5. MULTI-SYLLABIC ERRORS

As noted earlier, multi-syllabic words were deliberately included in our target words database in order to better reflect the corpus of song lyrics. The individual words used in the study can be seen in Appendix III. First, we will discuss the multi-syllabic word errors in the spoken stimuli, and then we will compare the data from the sung stimuli.

Of the 1,271 spoken stimuli, 194 words were multi-syllabic. This includes 153 two-syllable words, 28 three-syllable words, and 13 four-syllable words. Of these presented stimuli, listeners made only 10 errors, all of which were in bisyllabic words. The percentage of incorrect answers per total number of stimuli of the same syllable count is presented in Table 4, along with the numbers from the sung stimuli.

Although there were fewer presented stimuli of spoken polysyllabic words, the number of errors does decrease. However, the error rates expressed in percentage of errors express that listeners do not actually have an easier time hearing 2-syllable words than 1-syllable words. 3% of 1-syllable responses were incorrect, while 6% of 2-syllable responses were incorrect. However, for 3- and 4-syllable words, there were zero errors, though the number of presented stimuli is much less.

The sung stimuli present similar figures. There were 1,268 presented sung stimuli, of which 144 were 2-syllable, 26 were 3-syllable, and 16 were 4-syllable. Of the presented stimuli, there were a total of 67 errors, or 36% of all presented stimuli. Table 6 shows the breakdown of all syllabic errors, both spoken and sung.

Table 6. Percentages of errors separated by syllable distribution.

	spoken stimuli	**sung stimuli**
1-syllable word errors	3%	25%
2-syllable word errors	6%	33%
3-syllable word errors	0%	57%
4-syllable word errors	0%	25%

We can see that listeners had a very difficult time with 3-syllable words compared to the others. The most problematic word was 'domestic', which was not once correctly recognized in the sung examples; listeners most often reduced this word into the word 'mistake'. This was the case for many of the polysyllabic words – they were often reduced in syllable count, or turned into two-word responses (even

though listeners were instructed that they were hearing single words in the stimuli). These data show that polysyllabic words are not necessarily easier to understand in sung context.

CONCLUSIONS

The results of this study affirm the common experience reported by concertgoers and music listeners that sung lyrics are often unintelligible. In our experiment, sung lexical items produced more than seven times the number of mishearings compared with equivalent spoken passages, at least in the case of English. Our experimental passages were sung by unaccompanied soloists and were heard over binaural headphones by young musicians whose hearing had been screened. In a concert setting where the vocalist is accompanied by one or more instrumentalists, with a middle-aged listener seated some distance from the stage, one might expect the intelligibility to be degraded much more. At the same time, listeners may take advantage of pragmatic contextual information and musical repetition to better decipher the lyrical content.

In the case of vowels, our results replicate the centralization phenomenon originally observed by Benolken & Swanson (1990), Hollien et al. (1992), and Smith and Scott (1980). That is, vowels are perceived as more centralized when sung versus when spoken. In previous research, the phonetic analysis focused exclusively on vowels in isolation, but this study shows that the same phenomenon can be observed in actual word context. In our study, consonant errors during singing were also analyzed. While we hypothesized that liquids and nasals would be more easily recognized in singing, our data were not completely consistent with this hypothesis. In fact, we discovered that nasals and voiced stops were more difficult for listeners to recognize. Our data also show that the errors made while listening to sung words are different in nature than the errors made with spoken words.

In the case of consonants, we have identified several candidate mechanisms that might account for the observed confusions. Most consonant confusions appear to result from a disproportionate emphasis on the timbre of the surrounding vowels. Other confusions appear to relate to temporal changes, notably delaying articulation until the very end of a tone. In addition, training-induced hyperarticulation may also be leading to consonant confusions, especially in the case of voiceless consonants. The precise nature of the mechanisms leading to the loss of intelligibility in singing awaits further research.

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

Barlow, H. & Morgenstern, S. (1976). *Dictionary of Opera and Song Themes*. New York: Crown Publishers.

Benolken, M.S. & Swanson, C.E. (1990). The effect of pitch-related changes on the perception of sung vowels. *Journal of the Acoustical Society of America,* Vol. 87, No. 4, pp. 1781-1785.

Burleson, R. (1992). Functional-relationships of language and music -- The 2-profile view of text disposition. *Linguistique,* Vol. 28, No. 2, pp. 49-63.

Coren, S. & Hakstian, A.R. (1992). The development and cross-validation of a self-report inventory to assess pure-tone threshold hearing sensitivity. *Journal of Speech & Hearing Research*, Vol. 35, No. 4, pp. 921-928.

Francis, A.L. & Nusbaum, H.C. (1999). The effect of lexical complexity on intelligibility. *International Journal of Speech Technology*, Vol. 3, No. 1, pp. 15-25.

Hollien, H., Mendes-Schwartz, A.P., & Nielsen, K. (2000). Perceptual confusions of high-pitched sung vowels. *Journal of Voice,* Vol. 14, No. 2, pp. 287-298.

Ladefoged, P. (2001). *A Course in Phonetics.* Boston: Heinle & Heinle.

Smith, L.A. & Scott, B.L. (1980). Increasing the intelligibility of sung vowels. *Journal of the Acoustical Society of America,* Vol. 67, No. 5, pp. 1795-1797.

APPENDIX I: Melodies Used

Composer	Piece	#notes*	#syllables
Adam, Adolphe	Ah! Vous dirai-je, Maman from "Le Toreador"	7	1
Bach, J.S.	Cantata No. 100 "Was Gott tut, das ist wohlgetan"	7	1
Bach, J.S.	Singet dem Herrn ein neues Lied	27	1
Beethoven, Ludwig van	Gellert Lieder, Op. 48, No. 3	7	1
Beethoven, Ludwig van	Die Liebe des Nächtsen	7	1
Bizet, Georges	Habañera, from "Carmen"	8	2
Brahe, May H.	Bless this House	7	1
Debussy, Claude	La Damoiselle Élue	9	3
Donizetti, Gaetano	Don Pasquale	7	1
Fauré, Gabriel	Pie Jesu, from "Requiem, Op. 48"	7	1
Haydn, Franz Josef	Lob der Faulheit	7	1
Handel, Georg	Atalanta	7	1
Haydn, Franz Josef	Die Beredsamkeit	7	1
Humperdinck, Engelbert	Sandman's Song from "Hänsel und Gretel"	7	1
Mendelssohn, Felix	Die Lorely, Op.98	7	1
Mozart, Wolfgang Amadeus	Die Entführun aus dem Seraglio	28	1
Poulenc, Francis	Ban lite's No.2 "Hotel"	10	4
Puccini, Giacomo	La Fanciulla del West	7	1
Rossini, Gioacchino	Semiramide	16	2
Schumann, Robert	Hochländisches Wiegenlied	12	2
Smetana, Bedrich	The Bartered Bride, Duet	8	2

- The total number of notes does not include grace notes.

APPENDIX II: Classical Songs Composers List

Samuel Barber	10
Benjamin Britten	8
Stephen Albert	6

John Cage	2
David Del Tredici	5
Aaron Copeland	3
Charles Ives	3
Edward MacDowell	6
William Grant Still	1
Amy Marcy Cheney Beach	4
Henry Kimball Hadley	2
Phillip Glass	9
Samuel Adler	10
Sir Arthur Bliss	11
Ralph Vaughan Williams	18
Gustav Holst	13
John Dowland	14
Edward Elgar	11
Percy Grainger	9
Henry Purcell	5

APPENDIX III: Target Words Used in Experiment

One syllable words (97):

ain't	club	good	keep	live	night	shine	think	walk
air	couldn't	have	kids	long	oh	should	this	watch
all	dance	hear	kind	lord	one	smile	through	well
am	done	heard	knock	man	or	song	time	were
are	don't	heels	laugh	means	rhymes	steel	told	what's
as	fog	her	learn	meet	round	straight	tough	why
blues	free	him	leave	might	say	sweet	trance	would
boys	from	hope	let	Mike	see	swim	trees	wrong
care	get	if	lie	move	self	that's	turn	yes
cause	girls	ill	life	named	set	then	twice	
choose	gone	jean	like	new	shall	there	used	

Two syllable words (18):

army	burning	doodle	mighty	received	roxy
behold	Diane's	instant	mornings	record	shaker
belly	dinner	Irish	quartet	riches	sweepin'

Three syllable words (3):

affection	bulletproof	domestic

Four syllable words (2):

alabaster
caterpillar

APPENDIX IV: Phonological Analysis of Errors

Key:
word boundary
_ location of target phoneme
V vowel
C consonant
0 deletion
→ becomes

How to read an example:

/n/ → /m/ / #_C
the sound 'n' becomes the sound 'm' when occurring at the beginning of a word and before a consonant

Speaking Mistakes

/¢/ → /i/ / C_C
/¢ã/ → /ª/ / _#
/£/ → /aÒ/ / _C#
/£l/ → /aÒ/ / _#
/a¡/ → /o/ / _#
/i/ → /ÒÏ/ / _#
/i/ → 0 / _# +1 more instance
/Å/ → /v/ / #_ +1 more instance
/ã/ → /w/ / #_
/ã/ → /y/ / #_
/d/ → 0 / _#
/k/ → /h/ / #_
/k/ → /p/ / #_
/ks/ → /bz/ / V_V
/ks/ → /ltz/ / V_V
/ks/ → /ts/ / V_V
/l/ → /bl/ / #_ +1 more instance
/l/ → /w/ / _#
/l/ → 0 / _#
/l/ → È / #_
/n/ → /d/ / _#
/n/ → /Ï/ / _#
/n/ → /m/ / #_
/n/ → /m/ / _# +1 more instance
/n/ → 0 / _C
/t/ → /d/ / _#
/t/ → /h/ / #_
/t/ → /kt/ / _#
/t/ → /kt/ / _#
/t/ → /l/ / V_C#
/t×/ → /ltz/ / _#
/ts/ → /t/ / _#
/v/ → /d/ / _#
/v/ → /f/ / _#
/z/ → /dz/ / _#
/z/ → /s/ / _#
/z/ → 0 / _# +1 more instance
0 → /d/ / _# +2 more instances
0 → /m/ / #_
0 → /Ò/ / C_C#

Singing Mistakes

/e/ → /i/ / _# +1 more instance
/e/ → /Ò/ / #C_#
/i/ → /¢/ / #C_C
/i/ → /£/ / C_#
/i/ → /Ò/ / C_C + 7 more instances
/a/ → /¡/ / C_C +4 more instances
/a/ → /¥/ / #_
/a/ → /a¡/ / #_
/a/ → /o/ / #_ +2 more instances
/a/ → 0 / #_
/o/ → /¥/ / C_C +3 more instances
/o/ → /a/ / #_ +1 more instance
/o/ → /aÒ/ / #_ +2 more instances
/o/ → /aÒ/ / #C_ã
/o/ → /oÒ/ / C_C
/£/ → /¢/ / C_C +1 more instance
/£/ → /¢/ / CC_CC
/£/ → /§/ / C_C +1 more instance
/£/ → /a/ / C_C
/£/ → /ª/ / C_C
/¥/ → /¡/ / C_C +3 more instances
/¥/ → /§/ / C_C +6 more instances
/§/ → /¥/ / #C_C +2 more instances
/§/ → /a/ / C_C
/¡/ → /¢/ / C_C
/¡/ → /¥/ / C_C +2 more instances
/¡/ → /§/ / C_C
/¡/ → /aÒ/ / C_C

/¡/ → /o/ / C_C
/¢/ → /£/ / C_C
/¢/ → /a¡/ / C_C
/¢/ → /eÒ/ / #_
/¢/ → /i/ / #_
/¢/ → /i/ / C_C
/¢/ → /Ò/ / #_ +2 more instance
/Ò/ → /¢/ / #_
/Ò/ → /¢/ / _CC#
/Ò/ → /¥/ / #_
/Ò/ → /i/ / #_
/Ò/ → /i/ / C_C
/ª/ → /eÒ/ / _#
/ª/ → /o/ / C_C
/ª/ → /¢/ / C_C +1 more instance
/ª/ → /¢ã/ / C_#
/aã/ → 0 / #_
/aãm/ → /Å/ / #_
/¢t/ → 0 / _#
/¦l/ → /a/ / _#
/¦l/ → /wa/ / _#
/oÒ/ → /aÒ/ / C_C
/eÒ/ → /aÒ/ / #CCC_t +3 more instances
/a¡/ → /§/ / / C_C
/a¡/ → /a/ / C_C
/aÒ/ → /¢/ / CCC_C
/aÒ/ → /oÒ/ / C_C
/aÒ/ → 0 / #C_ +1 more instance
/aÒ£/ → /£/ / #C_ +3 more instance
/b/ → /Å/ / V_V
/b/ → /l/ / #_
/b/ → /m/ / #_ + 4 more instances
/b/ → /n/ / #_
/b/ → /p/ / #_
/b/ → /v/ / #_ +3 more instances
/b/ → /w/ / #_ +1 more instance
/b/ → 0 / #_
/d/ → /ã/ / V_V
/d/ → /g/ / #_
/d/ → /k/ / _#
/d/ → /n/ / V_V
/d/ → /t/ / _# +1 more instance
/d/ → /t/ / V_V
/d/ → /v/ / _#
/d/ → 0 / #_
/d/ → 0 / _# +2 more instances
/d/ → 0 / V_C
/dØ/ → /d/ / #_
/dØ/ → /s/ / #_
/do/ → 0 / #_ +2 more instances
/dz/ → /s/ / _#
/Å/ → /m/ / #_ +2 more instances
/Å/ → /n/ / #_
/Å/ → 0 / #_
/È/ → /b/ / #_
/f/ → /È/ / #_
/f/ → /È/ / _#
/f/ → /mp/ / _#
/f/ → /v/ / _# +2 more instances
/g/ → /l/ / #_
/h/ → /f/ / #_
/h/ → /u/ / V_C
/h/ → /v/ / V_V
/Ï/ → /l/ / _# +2 more instances
/Ï/ → /m/ / _#
/Ï/ → /n/ / _# +2 more instances
/Ï/ → /v/ / _# +3 more instance
/k/ → /b/ / V_V
/k/ → /g/ / #_ +1 more instance
/k/ → /h/ / #_
/k/ → /p/ / #_ +1 more instance
/kw/ → /f/ / #_
/kw/ → /k/ / #_
/kw/ → /p/ / #_
/l/ → /ã/ / #_V + 6 more instances
/l/ → /ã/ / #C_ +4 more instances
/l/ → /ã/ / _C#
/l/ → /d/ / #_
/l/ → /h/ / _# +2 more instances
/l/ → /m/ / #_
/l/ → /m/ / _#
/l/ → /n/ / #_
/l/ → /n/ / _# +1 more instance
/l/ → /n/ / V_V
/l/ → /p/ / _#
/l/ → 0 / #_
/ld/ → /g/ / _#
/ld/ → /Ï/ / _#
/ld/ → /m/ / _# +1 more instance
/ld/ → /n/ / _# +1 more instance
/lf/ → /È/ / _#
/m/ → /ã/ / #_
/m/ → /ã/ / #_ +1 more instance
/m/ → /b/ / #_
/m/ → /l/ / #_V +1 more instance
/m/ → /n/ / V_C
/m/ → /nd/ / _#
/m/ → /w/ / #_ +1 more instance
/m/ → /z/ / _#
/m/ → 0 / _#
/m/ → 0 / aÒ_z
/md/ → /nt/ / _#
/md/ → /t/ / _#
/mz/ → /s/ / _#
/mz/ → /z/ / _#
/n/ → /ã/ / _C#
/n/ → /b/ / #V_
/n/ → /d/ / #_
/n/ → /d/ / _# +4 more instances
/n/ → /d/ / ã_ing#
/n/ → /Ï/ / _# +1 more instance
/n/ → /Ï/ / _C# +1 more instance

/n/ → /l/ / #_ +2 more instances
/n/ → /l/ / V_V
/n/ → /m/ / V_V
/n/ → /m/ / _# +1 more instance
/n/ → /p/ / _#
/n/ → /t/ / _#V +1 more instance
/n/ → 0 / #_ +1 more instance
/n/ → 0 / _# + 2 more instances
/nd/ → /l/ / _#
/nd/ → /m/ / _#
/nst/ → /kst/ / #V_
/nt/ → /d/ / _#
/nt/ → /md/ / _# +2 more instances
/nz/ → /t/ / _#
/p/ → /m/ / _#
/rm/ → /n/ / V_V
/ã/ → /lv/ / V_V
/ã/ → /oÒ/ / _# +2 more instances
/ã/ → /v/ / V_V
/ã/ → 0 / #C_V
/ã/ → 0 / _# +3 more instances
/ã/ → 0 / V_C# +1 more instance
/ã/ → /n/ / _C#
/s/ → /f/ / #_
/s/ → /g/ / #_
/s/ → /n/ / _#
/s/ → /t/ / _# +4 more instances
/s/ → /z/ / _# +2 more instances
/st/ → 0 / #_ã +2 more instances
/stã/ → /s/ / #_ +1 more instance
/stã/ → /sw/ / #_
/t/ → /d/ / _# +5 more instances
/t/ → /k/ / #_
/t/ → /nkÈ/ / _#
/t/ → /ts/ / #_
/t/ → 0 / _# +1 more instance
/t×/ → /t/ / #_
/t×/ → /tã/ / #_
/tã/ → /dã/ / #_
/tã/ → /t×/ / #_ +2 more instances
/tã/ → /w/ / #_
/ts/ → /t/ / _# +11 more instances
/u/ → /a/ / #C_
/u/ → /a¡/ / C_C
/u/ → /o/ / C_C
/v/ → /f/ / _#
/v/ → /g/ / _#
/w/ → /ã/ / #_
/w/ → /h/ / #_
/w/ → /l/ / #_
/w/ → /l/ / #C_V
/w/ → 0 / #t_
/z/ → /dz/ / _#
/z/ → /s/ / _# +4 more instance
/z/ → 0 / _# +8 more instances
0 → /§n/ / _#
0 → /ã/ / CC_V
0 → /b/ / #_ +3 more instances
0 → /d/ / #_ +1 more instance
0 → /d/ / ã_# +3 more instances
0 → /È/ / _#
0 → /e/ / _C#
0 → /f/ / #_ã
0 → /f/ / _#
0 → /g/ / #_
0 → /h/ / #_ +10 more instances
0 → /k/ / _#
0 → /kw/ / #_ +1 more instance
0 → /l/ / _# +1 more instance
0 → /l/ / _C#
0 → /ld/ / _#
0 → /m/ / _# +5 more instances
0 → /m/ / V_C#
0 → /n/ / _# +4 more instances
0 → /n/ / _C# +2 more instance
0 → /p/ / #_ +3 more instances
0 → /p/ / _C# +3 more instances
0 → /v/ / _#
0 → /w/ / #_V